Advances in Science, Technology & Innovation
IEREK Interdisciplinary Series for Sustainable Development

AbdulLateef Olanrewaju · Silvana Bruno
Editors

Advancements in Architectural, Engineering, and Construction Research and Practice

Integrating Disruptive Technologies and Innovation for Future Excellence

IEREK RESEARCH & KNOWLEDGE ENRICHMENT

Springer

Advances in Science, Technology & Innovation

IEREK Interdisciplinary Series for Sustainable Development

Advances in Science, Technology & Innovation (ASTI) is a series of peer-reviewed books based on important emerging research that redefines the current disciplinary boundaries in science, technology and innovation (STI) in order to develop integrated concepts for sustainable development. It not only discusses the progress made towards securing more resources, allocating smarter solutions, and rebalancing the relationship between nature and people, but also provides in-depth insights from comprehensive research that addresses the **17 sustainable development goals (SDGs)** as set out by the UN for 2030.

The series draws on the best research papers from various IEREK and other international conferences to promote the creation and development of viable solutions for a **sustainable future and a positive societal** transformation with the help of integrated and innovative science-based approaches. Including interdisciplinary contributions, it presents innovative approaches and highlights how they can best support both economic and sustainable development, through better use of data, more effective institutions, and global, local and individual action, for the welfare of all societies.

The series particularly features conceptual and empirical contributions from various interrelated fields of science, technology and innovation, with an emphasis on digital transformation, that focus on providing practical solutions to **ensure food, water and energy security to achieve the SDGs**. It also presents new case studies offering concrete examples of how to resolve sustainable urbanization and environmental issues in different regions of the world.

The series is intended for professionals in research and teaching, consultancies and industry, and government and international organizations. Published in collaboration with IEREK, the Springer ASTI series will acquaint readers with essential new studies in STI for sustainable development.

ASTI series has now been accepted for Scopus (September 2020). All content published in this series will start appearing on the Scopus site in early 2021.

AbdulLateef Olanrewaju · Silvana Bruno
Editors

Advancements in Architectural, Engineering, and Construction Research and Practice

Integrating Disruptive Technologies and Innovation for Future Excellence

A culmination of selected research papers from the International Conference on Parallelism in Architecture, Engineering and Computing Techniques 2022 (4th edition), in collaboration with University of East London. And from the International Conference on Disruptive Technologies: Innovations and Interdisciplinary Considerations 2023 (1st edition), in collaboration with University of East London and British University in Dubai.

Springer

Editors
AbdulLateef Olanrewaju
Department of Construction Management
Universiti Tunku Abdul Rahman
Kampar, Perak, Malaysia

Silvana Bruno
Department of Civil, Environmental, Land,
Building Engineering and Chemistry
(DICATECh)
Polytechnic University of Bari
Bari, Italy

ISSN 2522-8714 ISSN 2522-8722 (electronic)
Advances in Science, Technology & Innovation
IEREK Interdisciplinary Series for Sustainable Development
ISBN 978-3-031-59328-4 ISBN 978-3-031-59329-1 (eBook)
https://doi.org/10.1007/978-3-031-59329-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Scientific Committee

Dr. Beh Jing Han, Universiti Tunku Abdul Rahman, Malaysia
Dr. Moyano Campos, Universidad de Sevilla, Spain
Dr. Muhammad Hasnolhadi bin Samsudin, Universiti Tunku Abdul Rahman, Malaysia
Prof. Brahim El Bhiri, Director SMARTiLab-EMSI
Dr. Zeenat Begam Yusof, International Islamic University Malaysia
Ir. Dr. Nor Ashikin Binti Muhamad Khairussaleh, Universiti Malaysia Pahang, Malaysia
Dr. Khor Soo Cheen, Universiti Tunku Abdul Rahman, Malaysia
Dr. Faridah Muhamed Halil, Universiti Teknologi MARA, Malaysia
Mr. Oni Olatoyese Zaccheus, Universiti Tunku Abdul Rahman, Malaysia
Dr. Dmitry Korzun, Petrozavodsk State University (PetrSU), Petrozavodsk, Russia
Dr. Wong Yee Choong, Universiti Tunku Abdul Rahman, Malaysia
Dr. Lim Chong Heng, Universiti Tunku Abdul Rahman, Malaysia
Dr. Wei Chooi Yi, Universiti Tunku Abdul Rahman, Malaysia
Dr. Ng Lee Peng, Universiti Tunku Abdul Rahman, Malaysia
Dr. Azlin Nordin, International Islamic University Malaysia
Dr. Chan Lee Kwun, Universiti Tunku Abdul Rahman, Malaysia
Dr. Haslinda Husaini, Universiti Teknologi MARA, Malaysia
Prof. M. J. M. Razi, University of Kelaniya, Sri Lanka
Dr. Hasnah Hashim, Universiti Teknologi MARA (UiTM) Malaysia
Ts. Dr. Wong Yee Choong, Universiti Tunku Abdul Rahman, Malaysia
Nur Athirah Nabila binti Mohd Idros, Universiti Tunku Abdul Rahman, Malaysia
Ts. Dr. Muhamad Sadry Abu Seman, International Islamic University Malaysia
Ts. Dr. Ku Chin Soon, Universiti Tunku Abdul Rahman, Malaysia
Dr. Ramesh Kumar Ayyasamy, Universiti Tunku Abdul Rahman, Malaysia
Dr. Nur Aqlima Binti Ramli, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shuhaili Talib, Universiti Tunku Abdul Rahman, Malaysia
Sr. Li Zi Qian, Universiti Tunku Abdul Rahman, Malaysia
Ahsiah Binti Ismail, International Islamic University Malaysia
Ir. Dr. Lee Wah Peng, Universiti Tunku Abdul Rahman, Malaysia
Dong Ling Tong, Universiti Tunku Abdul Rahman, Malaysia

The Editors warmly thank all the Reviewers who have contributed their authority to the double-blind review process, to ensure the quality of this publication.

Preface

The PACT (Parallelism in Architecture, Engineering and Computing Techniques) and DTIIC (Disruptive Technologies: Innovations and Interdisciplinary Considerations) conference series continues to be a dynamic source of innovation, exploration and cross-disciplinary collaboration in the ever-evolving landscape of architecture, engineering, construction and computing improving efficiency, performance and scalability. PACT and DTIIC gathers bringing together researchers and professionals from across the globe to delve into the transformative power of artificial intelligence, big data analytic, disruptive technologies, digital marketing, distributed computing and their far-reaching implications for the built environment.

As editors present this book, they are honoured to provide a written record of a selection of dynamic conversations, pioneering research and forward-thinking solutions that were presented during the PACT and DTIIC conferences.

The authors contributed to an exploration that traverses the confluence of Artificial Intelligence (AI) and distributed computing within the realms of architectural, engineering and construction practice. The book also addresses the nature and impact on disruptive technologies in today's marketplace. The pivotal discussions in the book "Parallelism in Built Environment Management" and Technologies reveal how digitalization reshapes urban landscapes into smart cities, revitalises ancient environments, and enhances historic site preservation. The support of high-dimensional data streams offers fresh insights into pattern recognition and anomaly detection.

The part about "Parallelism in Architectural Design" explores the role of robotics and artificial intelligence in redefining the architectural design process. From client briefs that culminate in 3D-printed constructions to the recognition and vectorization of floor plan sketches, these parts uncover the revolutionary ways that technology is reshaping architectural practice.

In "Parallelism in Performance Assessment", the book presents studies about analysis and optimization, in the calculation of cost-optimal energy efficiency and environmental performance assessment, highlighting the impact of urban design on environmental sustainability. An exploration of computing techniques for detecting mechanical failure and its critical applications concludes the part.

The impact of social media in digital marketing influences everyone's buying plan, whether in the construction market or otherwise. Digital marketing affects customers at different points in the buying process. Effective digital marketing techniques help businesses increase brand awareness, profit margins, foster customer confidence, and eventually increase conversions. We also see the impact of the complexities involved in purchaser retention for non-fungible tokens. In the non-fungible token (NFT) market, enticing purchases require a blend of effective marketing, awareness, education, and developing an appealing value proposition and case study. The problems facing the Small and Medium-sized Enterprises (SMEs) are poor profit margins, which are due to a lack of funding to execute projects, invest in advanced technologies, train, and retain competent staff. To address such problems, a combination of strategic planning, innovation, access to resources, adaptability, and collaboration with industry peers is critical. In a chapter detailing the nature of innovation in the public sector in Dubai, it is noted that the organizational culture and leadership dimension are some of the major constraints. Cryptocurrency has emerged as a form of digital or virtual currency lately. A chapter examined the democratic or decentralized nature of decentralized finance

(DeFi) for facilitating cryptocurrency. There has been a dramatic shift in innovation, especially in the last 20 years. With the coming of AI, there has been more concern and interest in what the future holds in terms of innovation. The next generation, Innovation 3.0, would be about systems innovation, quantum computing, circular economy, and biomimicry. Research involving the application of machine learning techniques to predict the inhibitory effects of compounds on PIM kinases with the ultimate goal of contributing to in silico drug design for targeting neoplastic (tumour-related) processes was presented. The research uses machine learning algorithms to categorize the chemical structure of compounds and their biological activities. However, the topic does not directly relate to building architecture or disruptive technologies in the construction sector; it relates to the application of machine learning in the fields of computational biology and pharmaceutical sciences.

This book stands as a tribute to the convergence of ideas, innovation, and interdisciplinary collaboration that defines the PACT and DTIIC conferences. It is a testament to the remarkable community that gathers annually to explore and push the boundaries of knowledge. We hope that this compilation of knowledge and ideas will inspire readers to embrace the potential of parallelism in architecture, engineering, technological innovation and computational techniques within their respective domains.

To all the authors, presenters, organizers, attendees and peer reviewers who have contributed to the conference series, we extend our gratitude and acknowledgement for being a part of this transformative journey. Thank you for joining us on this intellectual adventure, and we look forward to many more years of exploration and discovery within the community.

Bari, Italy — Silvana Bruno
Kampar, Malaysia — AbdulLateef Olanrewaju

Acknowledgments

The editors would like to thank everyone who helped to create this edited book from the bottom of their hearts. This book would not have been finished without the help, commitment and knowledge of numerous people.

First and foremost, we would like to express our profound gratitude to all of the book's contributors for their thoughtful chapters and insightful viewpoints. The richness and diversity of this volume have been considerably enhanced by your dedication to share your knowledge and experience.

We extend our sincere gratitude to the editorial team for their extraordinary work in arranging and moulding the contributions into a coherent and significant compilation. Your editorial abilities, attention to detail and commitment to upholding high standards have really improved the calibre of this work.

We are indeed grateful to IEREK team members for their great support in coordinating and publishing this book. We also appreciate in large measure the organizers of the conferences, from whom the chapters that make up this book came.

Lastly, we would like to thank the publishers and everyone who helped with the creation and publication of this book. We greatly value your expertise and dedication to academic success.

We would like to express our profound gratitude to everyone who contributed, no matter how minor, to the creation of this edited collection.

In some instances, we have not been able to trace the owners of copyright materials, we would like to appreciate any information that would enable us to do so.

Bari, Italy — Silvana Bruno
Kampar, Malaysia — AbdulLateef Olanrewaju

Introduction

This book addresses the pivotal discussions within the advancements of Architecture and Engineering and Construction (AEC) sector in refining problem-solving and decision-making during design, execution, management of urban landscapes, buildings, components and infrastructure, due to the rapid evolution of digital techniques and technologies aided by big data analytic, machine learning, artificial intelligence, robotics and computing.

Tailored for urban planners, architects, construction managers, developers, sustainability experts, engineers and researchers, this book integrates state-of-the-art research and innovative applications from the 7th edition of the international conference on Parallelism in Architecture, Engineering and Computing Techniques-4th Edition, a collaborative effort between IEREK and University of East London, UK and from selected articles from a conference on Disruptive Technologies: Innovations and Interdisciplinary Considerations. The book provides new perspectives in built environment management, architectural design and performance assessment through case studies and new workflows presented in the selected double-blind peer-reviewed papers.

The part "Parallelism in Built Environment" reveals how digitalization is reshaping the urban landscape into smart cities and is breathing new life into ancient environments enhancing preservation of historic sites. The support of high-dimensional data streams offers fresh insights into pattern recognition and anomaly detection.

The part about "Parallelism in Architectural Design" explores the role of robotics and artificial intelligence in redefining the architectural design process. From client briefs that culminate in 3D-printed constructions to the recognition and vectorization of floor plan sketches, these parts uncover the revolutionary ways that technology is reshaping architectural and engineering practice.

In "Parallelism in Performance Assessment", the book presents studies about analysis and optimization for cost-optimal energy efficiency and environmental performance assessment, highlighting the impact of urban design on environmental sustainability. An exploration of computing techniques for detecting mechanical failure and its critical applications concludes the part.

Digital marketing and social innovations draw attention to the expanding significance of digital media in the connected world of today; talk about how social media sites help to promote social advancements; some instances of creative digital marketing initiatives that have aided with societal innovations; talk about topics including disinformation, privacy problems, and responsible data use; and investigate how technology like virtual reality and artificial intelligence may further social causes. The book also introduces research in machine learning applied to quantitative structure-activity relationship (QSAR) analysis. It involves the quantitative assessment of the relationship between the chemical structure of compounds and their biological activity. It builds an inhibition activity-targeted machine learning classification to model the structure–activity relationships for protein kinase inhibitors. Different machine learning algorithms were used.

Contents

Parallelism in Built Environment: Managing Cities and Buildings with Artificial Intelligence and Distributed Computing

Smart Cities and Technology: The Role of Digital Technology in the Urban Fabric

Samruddhi Phalak

Abstract

Technology has become an integral part of our daily routines and has permeated every aspect of our lives. The advancements in digital technologies have brought about significant changes to people's lifestyles and work practices. Urban planners face unique challenges in managing cities, including population growth, climate change, infrastructure, and urban expansion. Understanding the impact of digitalization on urban environments is crucial for developing effective urban planning strategies. In today's rapidly moving society, we often face an excessive inundation of information. The implementation of technology in cities across the globe is becoming increasingly common. This advancement aims to increase the eminence of life for people. These innovative solutions range from energy-efficient buildings and transportation systems to advanced waste management and public safety measures. At the same time, industries are undergoing significant digital transformations that are revolutionizing operating, communicating, and interacting with their customers. Digital tech improves business effectiveness, efficiency, and innovation for quick market adaptation. Digital tech can enhance business efficiency, productivity, and innovation, enabling them to adapt quickly to market changes. The construction industry has embraced digitization, showing resilience during tough times. Digital platforms are in demand to thrive in competitive environments. AR/VR tools have revolutionized urban planning by improving communication and finding more effective solutions for city challenges. A challenge for architects is convincing clients that virtual views mirror the final design. The digital revolution has had an impact, especially in the construction industry which is rapidly becoming more digital. This research aims to understand and investigate the potential applications of interactive digital platforms and software for architecture and urban planning.

Keywords

Digital technologies · Urban fabric · Augmented reality · Virtual reality · Urban planning

1 Introduction

Across all continents, the technological and environmental developments that are taking place now have a profound impact on urban sustainability and resilience. For instance, the world is seeing unprecedented urbanization, especially in necessitous countries, and extensive environmental and climatological changes. Cities or, more broadly, urban agglomerations tend to develop innovative dominant sites of essential modifications in the individual, monetary, and physical topography of our planet while simultaneously being distinguished by a wide range of serious negative externalities. This is brought on by an essential increase in the extent and size of immense cities around the world. To address these challenges, residents, stakeholders, and governments need to work together to protect our urban planet (Nijkamp et al., 2022). The architecture and urban planning industries are experiencing a significant shift toward digitalization, resulting in a notable impact. The usage of Information and Communication Technologies (ICTs) by Digital Technologies (DTs) to generate, store, and regulate data. The functionality facilitates the exchange of data between individuals and electrical devices, as well as among digital computing systems that operate using binary code. Examining the impact of new technologies on user

S. Phalak (✉)
Vivekanand Education Society's College of Architecture (VESCOA), Mumbai, India
e-mail: samruddhi.phalak@ves.ac.in

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_1

behavior in the architecture is crucial, especially as DTs continue to revolutionize how the industry functions. The growth of digitization and automation ideas in the industry has led to increased attention toward DTs (Manzoor et al., 2021). In recent decades, the smart city movement has been aided by the digitization of the architecture and urban planning sectors brought about by cutting-edge information organization systems, such as the vast information, Internet of Things (IoT), and cloud computing technologies. In this context, novel data sources have provided never-before-seen insights into complex urban dynamics, which can alter governance paradigms and inform more effective urban administration and policy (Furtado et al., 2023). The relationship between technology and architecture is crucial. Technology has revolutionized the way architects work. Their approaches have undergone significant modifications thanks to advancements in the field. The rapid advancement of digital media technology, which began around the turn of the century, is expected to have an enormous effect on the industry. As urban living and communication are transformed by digital media, architects must embrace a new viewpoint. Architectural practices are adopting the most recent technological developments all around the world, with many already utilizing social media, augmented reality, and virtual reality to improve their work. For the benefit of society, architects must work together with a variety of professionals and groups. The digital era brings about new issues and discussions (Song, 2018). The utilization of digital technology and technological progress and a global push to promote smart city concepts and applications are some of the main causes of the past decade's developments. The popularity of city management and development has increased. Some of the underlying causes include technological innovation and a global push to promote smart city ideas and their use on a political, social, and economic level. Urban analytics, urban informatics, and city science methodologies are also gaining popularity. Higher education institutions have introduced new programs to teach the following generation of urban planners in urban analytics, urban informatics, and city science in response to the growing demand for novel skill sets. Urban planning and the organization of developing technologies like IoT, AI, Digital Cities, and Digital Twins are therefore becoming more and more important (Sabri & Witte, 2023). Data gathering and analysis, design selection, visualization, and evaluation are all part of the process of urban design and architecture. The application of augmented reality in all these stages, which aids in analysis and speeds up the production of the optimal design outcome, also allows for extra processes such as collaboration and immersion (Jishtu & Yadav, 2021). An innovative and effective reaction mechanism is required to deal with the numerous issues and challenges in our urbanized society, in which contemporary technology invariably plays a crucial role. In the New Urban World, ICT has become a key driver of resilient and sustainable growth. This has helped the initiative of smart cities or intelligent cities gain traction. The essence of the notion is a smart city's ability to increase its achievements in societal, monetary, technological, environmental, or cultural terms by using cutting-edge knowledge, information, and digital data strategies and technologies (Nijkamp et al., 2022). Examining the diverse prospects that digital platforms and software can provide in architecture and urban planning is a crucial investigation. By investigating their potential uses, we can gain a deeper insight into how technology can aid in the creation and progress of our physical surroundings (see Fig. 1).

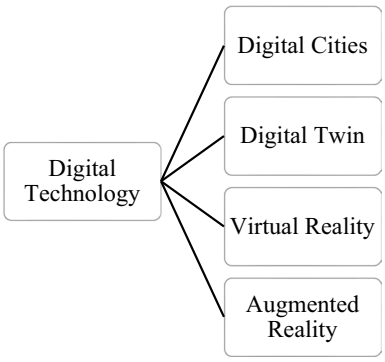


Fig. 1 Digital technology (*Source* Sabri & Witte, 2023)

2 Methodology

A keen interest in defining digital city and digital twin concepts that contribute to the architecture and planning of a city that is complex and challenging is revealed by the literature review. This research paper also focuses on VR (virtual reality) and AR (augmented reality), and its benefits in urban design and architecture to impart an all-inclusive understanding of the technology's potential and challenges. The primary goal of the study is to document the growth of these tools from a design perception and to comprehend their different uses. The data was gathered by reading 29 research publications on the theme cited by diverse researchers, published in conferences, and easily accessible on research sites. The subject-related papers were found using the keywords, and they were then carefully analyzed for greater understanding. The purpose is to understand how technological developments may help solve issues to produce better design and urban planning. The objective is to show the generation and exchange of knowledge via digital technologies may pave the way for more inclusive and data-driven planning.

3 Theoretical Framework

3.1 Digitization in Urban Planning

Urban planning has existed since cities first appeared. Early city design concentrated on critical aspects such as buildings and defenses and has evolved dramatically over the years. Several bold concepts significantly altered the path of traditional urban planning over the twentieth century (Das, 2020). With the use of Digital Technologies (DTs), the Architecture, Engineering, and Construction (AEC) industry is going through a huge transition from traditional labor-intensive methods to automation. This industry has been a major contributor to this revolution. Since the AEC sector places a lot of focus on these technologies, it is crucial to evaluate how they will affect user behavior as DTs change the dynamic between construction and behavior. DTs have attracted increasing interest because of the development of the notions of digitization and automation in Industry 4.0. Using digital technology, particularly in labor-intensive activities, they could be employed for logistics operations, near-real-time information flows, end-to-end supply chain consistency, and greater human contact. "The following technologies are viewed as promising DTs in the AEC sector: Building Information Modeling (BIM), Augmented Reality (AR), Virtual Reality (VR), photogrammetry, Radio Frequency Identification Devices (RFID), Geographic Information Systems (GIS), Global Positioning Systems (GPS), wearable safety devices, Quick Response code (QR), Artificial Intelligence (AI), robotics, blockchain, onsite mobile devices, and laser scanning devices. In general, it is well known that adopting DTs in the AEC sector has many benefits (Manzoor et al., 2021)" (see Fig. 2). How can people make quick, reliable, and effective judgements regarding future operations using information about all the parts of a physical asset that is accessible in real time? DT, or Digital Twin, is the remedy (Botín-Sanabria et al., 2022). The study emphasizes the value of digital twins and cities in urban development. Examine the use, issues, and innovations in the subject of urban planning.

3.2 Digital Cities

Citizens, businesses, and governments rely on data networks more and more for day-to-day operations in today's rapidly developing digital societies. A digital city is made up of Information and Communications Technology (ICT) applications that streamline government processes, cut down on communication costs, and provide extensive end-user services to cater to the demands of city residents. To address the requirements of the government and its staff, residents, and workers, a Digital City is a networked group that associations broadband infrastructures with a flexible, service-oriented computer architecture on open industry standards, as well as novel services (Yovanof & Hazapis, 2009). One of the earliest definitions of the digital city denotes a type of city that is substantially an open,

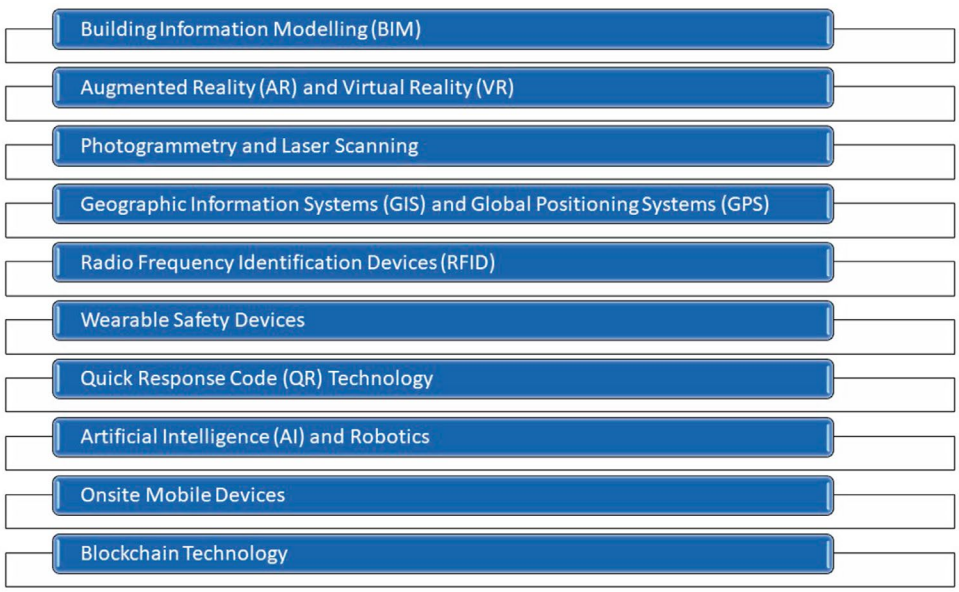


Fig. 2 Digital technologies (*Source* Manzoor et al., 2021)

intricate, and adaptive system based on computer networks and urban data resources, which shapes a virtual space for a city (D'Auria et al., 2014). A digital city framework, rather than attempting to replace real city life with virtual activities, aims to improve selected developments in city life by participating in actual and well-organized digital data and amenities. A determining factor of the worldwide expansion of digital cities is the desire to raise the standard of amenities provided by governments to their residents (Yovanof & Hazapis, 2009). It creates an information service market and a hub for the use of information resources. This sort of urban context's development is directly related to the widespread adoption of technology infrastructures and ICTs that started in the 2000s and gave rise to the term digital city (D'Auria et al., 2014). The concept of a digital city was developed in the 1990s, and it gained popularity, especially around the turn of the millennium, during the so-called Internet era. The application of the Internet in both the public and private sectors, other forms of communication and social networks, e-service delivery, and the accessibility of ample and the newest online information are driving forces behind the implementation of a digital city capable of utilizing all ICT instruments and designs to create a virtual city area (Dameri, 1970). According to the more popular definitions, the digital city is an urban plan that attempts to enhance people's quality of life using new technologies that play an important part in the city's growth by linking many stakeholders and delivering better service. The use of ICTs in an urban setting, which results in a multi-stakeholder approach, improves residents' engagement in urban decision-making as producers of their city (D'Auria et al., 2014). ICT is the foundation of the digital city, so its primary features include the distribution of comprehensive and recent information online; the use of social media and other communication channels to associate people and foster communication between the public and the government; the provision of e-services by both public and private organizations; and the widespread availability of information, communication, and services thanks to mobile. Like the smart city, the digital city can be defined as customizing the essential elements of the city to its features. The digital city can be designated as tailoring the fundamental factors of the city to its features, much like the smart city (Dameri, 1970). Establishing a space for data replacement, association, interoperability, and a seamless experience for all city residents is the purpose of a digital city, which forms the basis of a smart/intelligent city experience. A digital city is based on an information and communication technology architecture that provides broadband access and is linked by a service-oriented infrastructure to enable a diversified range of interoperable multimedia-rich apps. The community is linked and integrated by the technologies that make up the Digital City Fabric. A digital city's three main components: infrastructure: a communications network constructed by industry standards that is expanding and advancing the Internet and IP-based networks to previously unheard-of heights of development. A platform-based architecture for ambient intelligence and computing that supports ubiquitous and pervasive computing. Utility grids, other ICT-enabled physical urban infrastructures, the formation of intellectual urban structure blocks, and other factors all work together to create a smart city environment that is adaptable and flexible to the requirements of the information. Services are made available to everyone at any time and from any location because of this integrated environment's ability to mobilize data, apps, and users across the digital nation. Policy: A comprehensive set of rules and laws that encourages innovation and removes barriers to modernization (Yovanof & Hazapis, 2009) (see Fig. 3). To name a few well-known examples, Kyoto, Seattle, Helsinki, Shanghai, and Amsterdam are constructing information and communication spaces based on the city concept. The Internet can connect digital cities in the same way as land and air transport systems connect actual cities. Despite having a common name, digital cities have established organizational structures and business models, fulfill several functions, offer a range of services, and make use of a variety of system architectures. Due to the many communal backgrounds in which digital cities have advanced, this distinction has been made. The cultural and old capital hub of Kyoto, Japan, saw one of the earliest advances. To develop schemes for digital communities that may act as a communal and data structure for daily life, including commerce, transportation, education, and social welfare, NTT DOCOMO founded Digital City Kyoto in 1998 (Yovanof & Hazapis, 2009). To examine this issue, an intriguing example of a Kyoto commercial street neighborhood. In the area, there are 3,000 Kyoto stores. They co-founded a website where users can pay their bills online with debit and credit cards. Electronic systems are used to process purchase orders from both within and beyond the local community, and logistical companies provide the goods. As a result, a

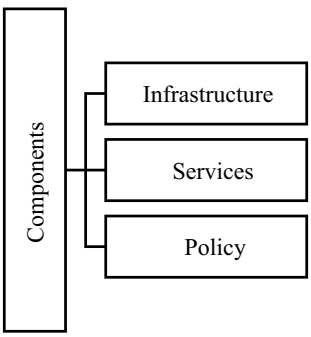


Fig. 3 Three components of digital city (*Source* Yovanof & Hazapis, 2009)

lot of Kyoto's shopping districts are now reachable online. Thousands of companies currently offer services under this heading. This initially appears to be another form of international cybernetic, like Yahoo. But it takes a different strategy in terms of business (Ishida, 2005). Digital city Amsterdam was created as Digital Helsinki concentrated on creating a new generation of metropolitan networks. It is a virtual area that reflects a city metaphor and offers a public communication place for city people. Urban planning that allows locals to actively participate in the design process is another argument in favor of digital cities (Yovanof & Hazapis, 2009).

3.3 Digital Twin

This concept first debuted in 2002 as a model for conception growth organization at the University of Michigan. During the presentation, a real place was duplicated as a virtual space in a digital environment, with linkages transmitting information from the actual to the simulated space and data in the opposite manner. The term digital twin was not applied to Grieves's proposal until 2010 when it appeared in a NASA-integrated technology plan. Around the year 2018, digital twins for urban management began to emerge. Recently, the urban management industry has expressed attention to incorporating digital twins into cities to develop urban planning, enhance asset administration, and generate safe, sustainable cities. Urban digital twins have grown rapidly in recent years because of the information obtained from smart city growth and the use of digital twins in other sectors of the business (Ferré-Bigorra et al., 2022). A digital twin example is a form of a digital twin that represents its physical counterpart throughout its existence. This means that the physical twin's condition is constantly monitored, and any variations or progression faced by the physical twin will influence the digital twin. In this context, this term refers to the management, operations, monitoring, and forecasting of a product or process from conception to completion. It is beneficial for validating the intended behaviors and presentation of a product or object (Botín-Sanabria et al., 2022). Digital twins have the potential to improve city management. However, it seems to vary depending on the objectives of the current urban management. For example, when the urban manager wants to demonstrate technological expertise, digital twins are made avant-garde and creative; when social engineering is taking place, they contribute to the generation of useful interventionist strategies; when the urban manager needs to discover people's attitudes, they emphasize exploratory and collaborative mechanisms; and when societal reform is the primary goal, they incline to emphasis on citizen engagement and mechanisms for social interdisciplinary collaboration (Al-Sehrawy et al., 2023). The usage of digital twins for design purposes is advantageous since it allows designers to better understand client demands and requirements, which leads to an improved understanding of the product's appearance and functionalities. Using the digital twin at an early point of the product life cycle reduces production costs (Alnowaiser & Ahmed, 2023). Urban digital twins could transform this, propelling smart cities and urban models to new heights. By utilizing data collected in smart cities and automatically integrating them into the city model and its processes, a realistic digital duplicate proficiency of independently interacting with the city is maintained. City digital twins might concentrate on the city's social and economic aspects in addition to duplicating, replicating, and relating to the physical characteristics of the city (Ferré-Bigorra et al., 2022). The goal is to form a digital illustration of the city that can be used to study problems like urban planning and climate change. As a result, various components of the digital twin must be updated at various intervals and, when needed, reinforced with real-time data. The digital twin increases urban planning's possibilities. The new apps offer better comprehension, more focused questions, and thus clear responses to concerns that will be important in the future (Schrotter & Hürzeler, 2020). The creation of smart campuses and digital twins aimed at replicating building behavior in the real situations of both the neighboring and internal situations and giving rapid decisions through contextualized data seems promising when using BIM-GIS integration (i.e., GeoBIM). When used in conjunction with Internet of Things (IoT) networks and Artificial Intelligence (AI) systems to create Cognitive Digital Twins (CDTs), or structures with cognitive functions capable of autonomously and dynamically responding to environmental changes, the GeoBIM approach offers several benefits (Meschini et al., 2022). To analyze a 3D model from various angles, two distinct domains—Geographic Information Systems (GIS) and Building Information Modeling (BIM)—virtually employ two separate most common data interchange protocols. GIS focusses on the geographic data of buildings and their parts from a geographical perspective. However, from an architectural and construction standpoint, BIM emphasizes further the specific building modules and project data, such as cost and time. It is an arena that is extensively acknowledged by Architecture, Engineering, and Construction (AEC) because of technological advancements that have steadily improved the construction industry over time and allowed shareholders to capture, manipulate, improve, and substitute information through the life cycle of building construction projects. While GIS integrates various geographical and attribute statistics and derives information through various spatial scrutiny tools and modeling methods to visualize and analyze locality issues in geospatial science and natural resource

management, BIM acts as a mediator of integration among industry players, increasing its acceptance among clients (Sani & Rahman, 2018). A complex and vibrant virtual city in three dimensions is called The Virtual Singapore (VSdt). It is a technological platform that combines tools, toolsets, and ad hoc simulative/predictive algorithms to influence decision-makers. This platform, which was created using the proprietary package, is envisioned for use by the private, public, and research areas to start up new digital technology-based firms. Users will have access to interactive tools and software that are primarily designed for analyzing and choosing effective urban development strategies thanks to funding initiatives managed by several institutional and governmental authorities. Numerous academic studies have examined the varied persons involved, realistic goals, and challenges addressed in DT. The individuals selected enable us to see the VSdt as a true multidisciplinary technology center that can support the subject's cutting-edge research on a global scale. Even though the digital/virtual growth of the system is well recognized in European experiences, some investigations carried out by the project partnership demonstrate extremely timely and cutting-edge workflows for (i) urban microclimate analysis, (ii) 3D representation of open-space components with augmented reality, and (iii) The Level of Detail (LOD) achieved by a computerized model created using partially automated geo-AI informatics techniques (Caprari, 2022) (see Fig. 4). The expertise and standards of conduct for planning professionals will change due to the new business models brought about by advancing technology. Describe the process used to predict Melbourne's demand for public transportation using AI-based neural network models. This modeling practice necessitates that transport planners boost their capabilities for scenario-based advances by utilizing AI and Deep Learning (DL) analytics. Their conclusions reveal that using AI surpasses prevailing prediction techniques and results in over 90% accurate passenger demand projections. The affordability and effectiveness of public transit, as well as consumer satisfaction and sustainability, can all be enhanced in terms of urban management. Technology is advancing more quickly, as the use of AI and DL in Melbourne shows. The planning community is being introduced to innovative technologies that hold great promise for enhancing public services, such as Digital Twins. More study is needed, nevertheless, to fully understand the range of applications they have for urban management. They contend that no matter how well-executed Digital Twins are, they still fall short of illuminating the fundamental beliefs and principles that guide urban planners who create or use Digital Twins in their day-to-day work. Due to this, the function of the developer and the urban environment in which digital technologies are deployed must receive more attention (Sabri & Witte, 2023).

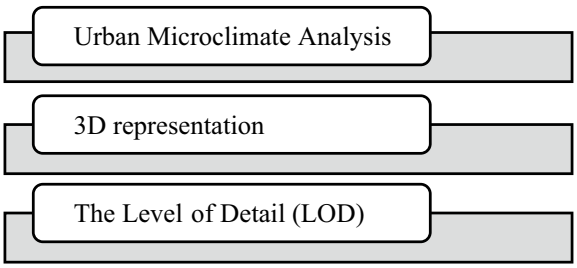


Fig. 4 Digital development workflow (*Source* Caprari, 2022)

4 Digitization in Architecture

4.1 Virtual Reality (VR) and Augmented Reality (AR)

The advantages and uses of Virtual Reality (VR) and Augmented Reality (AR) technologies in the architecture, and construction business have attracted the attention of many scholars. Although effective AR and VR use in the industry is essential, several parts of the sector have already successfully implemented these technologies. Researchers have put a lot of effort into compiling research on the applications of AR and VR in the construction sector (Albahbah et al., 2021). Jaron Lanier of VPL research first used the term Virtual Reality (VR) in 1989 to distinguish his attempt to construct immersive digital worlds from more conventional computer simulations (Paranandi & Sarawgi, 2014). Virtual Reality (VR), which enables a three-dimensional, computer-generated world to be viewed and related by a person, is one of the technological tools utilized by the construction industry (Mutis, 2020). Architectural tour systems were the earliest attempts to use VR as a visualization tool. After 1986, the University of North Carolina began conducting ground-breaking research in this subject, constantly developing new system generations. The visualization of St. Peter's Basilica at the Vatican exhibited at the Virtual Reality World'95 meeting in Stuttgart and the commercial Virtual Kitchen design tool are just two examples among many more research groups' remarkable uses. What makes VR so amazing that it outperforms traditional computer graphics? Even the most life-like still images or animations cannot capture the sensation of presence and space in a virtual building. It may be observed and understood in various lighting situations, just like in actual facilities. The Frauenkirche in Dresden, for example, was destroyed, but one may even stroll through houses that have not been built yet (Mazuryk & Gervautz, 1999). The world witnessed significant changes in construction forms and characteristics in

the twentieth century. The construction industry has seen enormous transformation and improvement. This sector has revolutionized approaches, methods, strategies, and policies by forming immense and improved ideas. The construction industry makes use of Virtual Reality (VR), which allows users to explore and interact with a three-dimensional, computer-generated environment (Ahmed, 2019). According to a specialist, Virtual Reality (VR) technologies are useful for project schedule management, site layout optimization, and construction project safety training. VR technology can also create settings that foster better participant participation and help spot design flaws in complex projects. Describe the building geometry to help people understand a project, make a better design choice, and facilitate group decision-making. The goal of Virtual Reality (VR) simulation is to create immersive settings that allow users to gain novel insights into how the real-world functions. The first immersive Human–Computer Interaction (HCI) prototype known as the "Man–Machine Graphical Communication System" introduced the idea of VR more than 50 years ago (K et al., 2021). "The use of Virtual Reality (VR) in the following areas is on the rise: (i) walkthroughs for visualization, analytical simulation (e.g., energy, circulation, facilities management), and virtual reconstruction; (ii) design decision-making; (iii) collaboration; (iii) marketing; and (iv) building" (see Fig. 5). The most common productive uses of virtual reality, which are still restricted to passive and exploratory VR, are architectural walkthroughs and remote collaboration. The majority of the Immersive VR (IVR) projects were completed in the middle of the 1990s, and they appear to have either come to an end or been diverted into another application field. We were unable to locate any IVR systems that were in regular use in architecture or effective enough for practical usage, except for a few prototype demonstration systems in a few research institutes. Right now, VR systems' degree of immersion and photorealism closely relate to their price. For usage in architecture, a high-end VR system with high levels of immersion and photorealism is still too expensive. Present collaboration, which we claim to be more advantageous, becomes possible if we do away with these elements, which we say are not necessary for making most design decisions (Paranandi & Sarawgi, 2014). The same idea underlies Augmented Reality (AR), but rather than interacting in a non-existent environment (digital fact), AR takes advantage of the current environment while integrating virtual features to make it look as though both are there at the same time (Ahmed, 2019). Data gathering and analysis, design decisions, visualization, and evaluation are all part of the same process in architecture and urban design. Augmented Reality (AR) has been developed for immersive and group usage in each stage to assess the ideal design solution. These systems are designed for the public, design students, and policymakers as professionals (Jishtu & Yadav, 2020). Effective communication and information retrieval from the construction site are crucial prerequisites for a successful construction project. When differentiated from traditional information sources, the ratification of various AR programs dramatically improves project information on the job site and effective communication. In the construction sector, augmented reality technologies are employed for field data collecting and circulation to users. It also reflects the many modes of communication between project participants and project information (Ahmed, 2019). AR technology is anticipated to enhance current engineering management systems, design processes, building construction processes, and architectural visualization (Wang, 2009). A human can interact with a three-dimensional, computer-generated image or environment in a way that seems actual or physical by using a specific digital technology, such as a helmet with an internal display screen or gloves fitted with sensors. The augmented reality technology during walkthroughs should aid staff in obtaining thorough in-person situational awareness. This awareness affects personnel's knowledge of design visualization and other types of design information representation, such as text-based information. Sensors, mobile computers, and wireless capabilities are recent breakthroughs in AR technology development that enable technology construction. Users can examine (1) position tracking, which uses an interior location system for place awareness and navigation supervision; (2) spatially located strategy mechanisms, which use virtual 3D objects; and (3) text-based strategy illustrations (see Fig. 6). Augmented Reality (AR) is the combination of digital and physical observations. Users can interact with virtual objects and perceive them as a part of their environment. The architecture and design industries generate enormous expanses of statistics and information necessary to be accessible by several stakeholders, sometimes in multiple places. The necessity for access to huge volumes of project, engineering, and

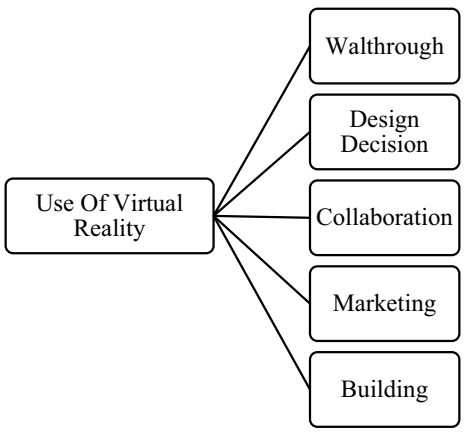


Fig. 5 Use of virtual reality (*Source* Paranandi & Sarawgi, 2014)

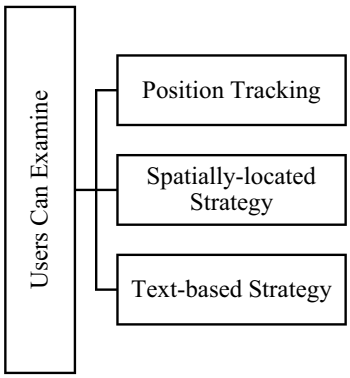


Fig. 6 VR & AR users can examine (*Source* Wang, 2009)

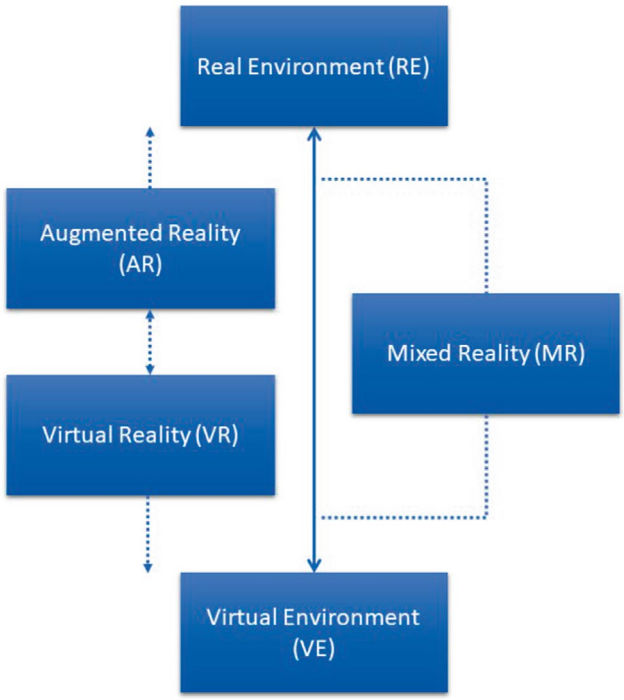


Fig. 7 Mixed reality (*Source* Mutis, 2020)

administration data in the architectural and design sectors sets conditions for the most promising use of AR techniques by including associated workers in the augmented workplace (Wang, 2009). After COVID-19, many executives and managers in the construction industry were unable to continue with their regular jobs because so much of the world had become distant. The distance from the project team that COVID-19 brought with it made it harder to collaborate. As a response, businesses like Zoom and others introduced innovative solutions for users to effectively connect with their workforce. Zoom and other conferencing programs can be used in the construction industry to some extent. Although Augmented Reality (AR) technologies are advantageous to users on their own, the construction industry made significant strides toward integrating AR technology into many of their procedures after the epidemic. By enabling teams to communicate 3D photos and videos with team members who are not present, augmented reality can streamline cooperation in remote situations (Ghoddoucy, n.d.). As technically described by Paul Milgram and Fumio Kishino, Mixed Reality (MR) is a subset of Virtual Reality (VR)-related skills for generating settings in which real and virtual world elements coexist on a single display. The terms Augmented Virtuality (AV) and Augmented Reality (AR) are the major subsections of the Reality Virtuality (RV) continuum. Virtual reality, a technique or setting in which additional information produced by a computer is added to the user's vision of a real-world event, is often used interchangeably with augmented reality (Wang, 2009).

Mixed Reality (MX), which denotes a multi-axis range of fields that includes Augmented Reality (AR), Virtual Reality (VR), and telepresence, is frequently mistaken for Augmented Reality (AR). Virtual Reality (VR) is the term used to describe a simulation of a wholly artificial environment in which the participant becomes completely immersed (see Fig. 7). Telepresence, for example, aims to broaden the operator's sensory-motor and problem-solving skills in an isolated situation. The setting at the remote location creates a natural area for the operator's presence and interactions (Mutis, 2020). Keeping track of all the odd logistical issues during construction is challenging. A single computer screen can now display all the moving parts on the job site thanks to augmented modeling software. It becomes possible to provide specific remedies. Additionally, all stakeholders have access to these models, facilitating improved oversight. More employees having access to a thorough digital model leads to fewer errors and less rework, which reduces construction costs. The decrease in safety concerns is another noteworthy benefit. Employees with access to a digital overlay to identify potential risks feel more protected (Ghoddoucy, n.d.).

5 Conclusion

Advancements in digital technology have revolutionized the fields of architecture and urban planning, presenting professionals with new opportunities to progress in their careers and keep up with the latest technical developments. This technology enables buildings to be constructed and operated in innovative ways, ultimately improving the quality of life for everyone. By embracing new technology, architects can stay ahead of the curve and prepare for the future. At the same time, planners can use Digital Twins and Digital Cities to research and predict changes in the city without disrupting the public. Architects and designers must

stay up to date with cutting-edge technologies such as 3D printers, drones, and computer gaming goggles, as digitalization continues transforming the infrastructure and building industries. Human–computer interaction is still in its infancy, but it has the potential to revolutionize our understanding of environments and their ergonomics. While researching and sketching are crucial in understanding and designing for people, augmented and virtual reality offer a fourth dimension to our comprehension. Incorporating the Internet of Things (IoT) in construction is critical for providing relevant data and ensuring the safety of all stakeholders. To meet architects' and designers' demands, top software providers like Autodesk, Adobe, and Graphisoft offer free access to their services. In a way that voice conferencing cannot, ubiquitous video conferencing brings people together and fosters a sense of community. Video conferencing has taken the role of voice conferencing for internal communication within the organization. Ultimately, embracing technology will enhance architects' creativity, efficiency, and productivity, and contribute to the betterment of society, one structure at a time. Technology is profoundly impacting architecture today and in the future. Digital media and the upcoming technologies have already started to impact life.

References

Ahmed S (2019) A review on using opportunities of augmented reality and virtual reality in construction project management. https://www.researchgate.net/publication/330218685_A_Review_on_Using_Opportunities_of_Augmented_Reality_and_Virtual_Reality_in_Construction_Project_Management

Albahbah M, Kıvrak S, Arslan G (2021) Application areas of augmented reality and virtual reality in construction project management: a scoping review. J Constr Eng, Manag & Innov. https://doi.org/10.31462/jcemi.2021.03151172

Alnowaiser KK, Ahmed MA (2023) Digital twin: current research trends and future directions. Arab J Sci Eng 48(2):1075–1095. https://doi.org/10.1007/s13369-022-07459-0

Al-Sehrawy R, Kumar B, Watson R (2023) The pluralism of digital twins for urban management: bridging theory and practice. J Urban Manag. https://www.sciencedirect.com/science/article/pii/S222658562300002X

Botín-Sanabria DM, Mihaita A-S, Peimbert-García RE, Ramírez-Moreno MA, Ramírez-Mendoza RA, de Lozoya-Santos JJ (2022) Digital twin technology challenges and applications: a comprehensive review. Remote Sens 14(6):1335. https://doi.org/10.3390/rs14061335

Caprari G (2022) Digital twin for urban planning in the green deal era: a state of the art and future perspectives. https://www.researchgate.net/publication/367577515_Digital_Twin_for_Urban_Planning_in_the_Green_Deal_Era_A_State_of_the_Art_and_Future_Perspectives

D'Auria A, Tregua M, Bifulco F (2014) Digital city towards smart city: a theoretical overview. ResearchGate. https://www.researchgate.net/publication/298201516_Digital_City_towards_Smart_City_a_theoretical_overview

Dameri RP (1970) Smart city and digital city implementation: two best practices in Europe. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-319-45766-6_6

Das A (2020) Urban planning, information technology, and artificial intelligence: the theory of evolution. https://www.researchgate.net/publication/341652319_Urban_Planning_Information_Technology_and_Artificial_Intelligence_The_Theory_of_Evolution

Ferré-Bigorra J, Casals M, Gangolells M (2022) The adoption of urban digital twins. Cities. Elsevier BV. https://doi.org/10.1016/j.cities.2022.103905

Furtado LS, Da Silva TLC, Ferreira MGF, De Macêdo JAF, De Melo Lima Cavalcanti Moreira JK (2023) A framework for digital transformation towards smart governance: using big data tools to target SDGs in Ceará, Brazil. J Urban Manag. Elsevier BV. https://doi.org/10.1016/j.jum.2023.01.003

Ghoddoucy C (n.d.) Future of augmented and virtual reality in construction. Calpoly.edu. Accessed 8 Oct 2023. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1769&context=cmsp#:~:text=AR%20and%20VR%20systems%20display,virtual%20tours%20before%20construction%20starts.

Ishida T (2005) Activities and technologies in the digital city Kyoto. In: Digital cities III. Information technologies for social capital: cross-cultural perspectives. Springer Berlin Heidelberg, pp 166–187

Jishtu P, Yadav M (2020) Application of augmented reality & virtual reality in architecture and planning: an overview. https://www.researchgate.net/publication/348020352_Application_of_Augmented_Reality_Virtual_Reality_in_Architecture_and_Planning_an_Overview

Jishtu P, Yadav M (2021) Futuristic technology in architecture & planning-augmented and virtual reality: an overview. https://www.researchgate.net/publication/350009266_Futuristic_Technology_in_Architecture_Planning_-_Augmented_and_Virtual_Reality_an_Overview

K, P, K, C, J, S, Naga C (2021) A Review on virtual reality and augmented reality in architecture, engineering and construction industry. https://www.researchgate.net/publication/354238765_A_Review_on_Virtual_Reality_and_Augmented_Reality_in_Architecture_Engineering_and_Construction_Industry

Manzoor B, Othman I, Pomares JC (2021) Digital technologies in the architecture, engineering, and construction (AEC) industry—a bibliometric—qualitative literature review of research activities. https://www.researchgate.net/publication/352189856_Digital_Technologies_in_the_Architecture_Engineering_and_Construction_AEC_Industry-A_Bibliometric-Qualitative_Literature_Review_of_Research_Activities

Mazuryk T, Gervautz M (1999) (PDF) Virtual reality-history, applications, technology, and future. https://www.researchgate.net/publication/2617390_Virtual_Reality_-_History_Applications_Technology_and_Future

Meschini S, Pellegrini L, Locatelli M, Accardo D, Tagliabue LC, Di Giuda GM, Avena M (2022) Toward cognitive digital twins using a BIM-GIS asset management system for a diffused university. https://www.researchgate.net/publication/365942275_Toward_cognitive_digital_twins_using_a_BIM-GIS_asset_management_system_for_a_diffused_university

Mutis I (2020) Challenges and enablers of augmented reality technology for in situ walkthrough applications. Researchgate.net. https://www.researchgate.net/publication/338899809_Challenges_and_enablers_of_augmented_reality_technology_for_in_situ_walkthrough_applications

Nijkamp P, Kourtit K, Türk U (2022) Special issue on 'the city 2.0–smart people, places and planning.' https://www.researchgate.net/publication/360882789_Special_issue_on_%27The_city_20_-_Smart_People_Places_and_Planning%27

Paranandi M, Sarawgi T (2014) Virtual reality in architecture: enabling possibilities. In: Proceedings of the 7th international conference on computer-aided architectural design research in Asia (CAADRIA). https://www.academia.edu/6313562/VIRTUAL_REALITY_IN_ARCHITECTURE_ENABLING_POSSIBILITIES

Sabri S, Witte P (2023) Digital technologies in urban planning and urban management. J Urban Manag. https://www.sciencedirect.com/science/article/pii/S2226585623000079

Sani MJ, Rahman AA (2018) GIS and bim integration at data level: a review-researchgate. https://www.researchgate.net/publication/328683270_GIS_AND_BIM_INTEGRATION_AT_DATA_LEVEL_A_REVIEW

Schrotter G, Hurzeler C (2020) The digital twin of the city of Zurich for urban planning-PFG. J Photogramm, Remote Sens Geoinformation Sci. SpringerLink. https://link.springer.com/article/10.1007/s41064-020-00092-2

Song YC (2018) Architecture practice in the digital age: how technology will change the profession. Academia.edu. https://www.academia.edu/35795648/ARCHITECTURE_PRACTICE_IN_THE_DIGITAL_AGE_HOW_TECHNOLOGY_WILL_CHANGE_THE_PROFESSION

Wang X (2009) Augmented reality in architecture and design: potentials and challenges for application. Researchgate.net. https://www.researchgate.net/publication/30876075_Augmented_Reality_in_Architecture_and_Design_Potentials_and_Challenges_for_Application

Yovanof GS, Hazapis GN (2009) An architectural framework and enabling wireless technologies for digital cities and intelligent urban environments. Wirel Pers Commun. SpringerLink. https://link.springer.com/article/10.1007/s11277-009-9693-4

Use of Advanced Digital Technologies in Re-Presentation of an Ottoman Caravanserai and Its Surrounding Historic Built Environment in Bursa, Turkey

Sermin Çakıcı Alp

Abstract

Conservation issues in architectural heritage require negotiating complex data encompassing geographical, architectural, social, and economic aspects of a historic urban area. Various types of advanced digital technologies can be used for such detailed work. The purpose of this study is to demonstrate the utility of these techniques in the re-discovery of the original architectural character of a mostly collapsed sixteenth-century Ottoman caravanserai, Ali Paşa Caravanserai, which was constructed under the supervision of Mimar Sinan (the major Ottomans' Major Architect) within the Historic Trade Center of Bursa as one of the World Heritage Sites (WHS) in Türkiye. Meanwhile, it also aims to represent the changes in its form and function with the urban transformation activities that appeared in the city center since the end of the nineteenth century. For this purpose, the prevalent use of digital techniques to preserve a hardly damaged architectural heritage is described initially. Following brief information about the spatial character of the study area, including this historic monumental building, the method prepared in different phases of its rehabilitation project is designated by using digital tools to document its original structure and evaluate the physical changes in parallel with manmade deformations. At the end of the study, the necessity of digital technologies to express the authenticity of architectural heritage is discussed, while using them in finding solutions for the preservation and perception of its surrounding traditional urban texture.

Keywords

Digital technologies · Urban conservation · Architectural heritage · Ottoman identity · Bursa

1 Introduction

The authenticity and integrity of a historic built environment should be interpreted in broader spatial and temporal contexts. There should be logical, flexible, and practical principles defined to achieve accurate and unbiased results using the documentation analysis of demolished historical monuments and areas. Hence, advanced digital technologies can contribute to finding sustainable solutions for conservation through database development, spatial analysis, and visualization. They help to handle a complex structured geo-database with vast spatial, temporal, and heterogeneous information about a historic built environment. Visualization can be accomplished with Computer-Aided Design (CAD). At the same time, some researchers prefer to use Remote Sensing Modeling Technologies (RSMT) for animation and virtual tours to unify the graphic and attribute data of related historical areas (Piwowar 2005; Droj, 2010; Del Pozo et al., 2017). Moreover, Geographic Information Systems (GIS) contribute to managing the complex configuration of collected data and help researchers compare current conditions with original forms to discuss the physical deformation of architectural heritage. It also provides the ability to integrate multiple layers of information that reference the existing diverse datasets during the preparation of conservation management plans for sustainable tourism development of historic towns (Fletcher et al., 2007; Ruoss et al., 2013; Ciski et al., 2019).

According to the literature, integrity among these digital techniques with Information Management Systems (IMS) is essential for conducting a geo-environmental evaluation of urban land use and conservation management planning. Using information systems in entering collected datasets

Some of the visual and written sources used in this article are taken and rev-designed by the author from the master's thesis that she completed in 2008.

S. Ç. Alp (✉)
Faculty of Architecture, Hacettepe University, Ankara, Türkiye
e-mail: serminalp@hacettepe.edu.tr

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_2

makes it easy to record and evaluate the physical and social characteristics of cultural properties while undertaking the whole process of a conservation project, from the documentation to the decision phase. For instance, scholars, who study the advantages of cooperation in the use of photogrammetry together with GIS, mention the importance of digital information systems in the documentation of a historic area, having quality data input into these systems (Duran & Toz, 2002; Sanjuan et al., 1999; Karslı et al., 2003; Pavlidis et al., 2007; Vileikis et al., 2012; Heras et al., 2013). As one of them, Value-Based Monitoring Systems (VBMS) are used to analyze and make multi-criteria decisions while developing conservation plans and rehabilitation projects concerning historic built environments. In addition, 3D inventory and virtual reconstruction can be beneficial in sustaining historical sites while helping to find the most sustainable solutions to the deterioration of the authenticity of fragile cultural heritage. The interest in the digital 3D reconstruction of cultural heritage has increased in the past decades. The basic motivations are mentioned by the scholars (Ludwig et al., 2013; Gomes et al., 2014; Ragia et al., 2015; Heuberger et al., 2015; Lewińska & Zagórski, 2018):

- Providing the appearance of a cultural property, which was badly damaged by natural or accidental causes;
- Contributing to collecting specific geometric and texture information that was invested during the documentation and restitution process in preparation for the conservation project of mostly collapsed architectural heritage; and
- Consequently can display the original form and function of related immovable heritage.

This type of implementation is used to improve the understanding of a historic site without resorting to often traumatic interventions for the original artifact. By this, the objects, that no longer exist or are mostly damaged, can be reconstructed digitally and the 3D model can then be viewed in its correct historical context. More importantly, these digital technologies also provide an opportunity to connect with municipalities, universities, public or private institutions, and stakeholders. That collaborative working environment also results in successful and cost-efficient operations in cooperation with associations interested in cultural heritage sustainability.

Therefore, it is essential to study the utility of advanced digital technologies used for the sustainability of architectural heritage, as a component of traditional texture defining the historic urban identity of Bursa, which is a World Heritage Site in Türkiye as being the first capital of Ottoman Empire. In this respect, this research aims to discover the authentic architectural character of a mostly collapsed sixteenth-century Ottoman caravanserai, named Ali Pasha Caravanserai, which was constructed under the supervision of Mimar Sinan (employed as the major architect of the Ottomans), while analyzing its rehabilitation process together with its surrounding built environment, as a part of the Historic Trade Center of Bursa. In addition, the changes in the spatial form and function of its surrounding built environment are displayed chronologically by the result of this study, to understand the impacts of urban transformation activities that have continuously applied in the historic city core since the end of the nineteenth century.

Within the concept of this article, brief information about the past and present condition of the study area, including this hardly damaged architectural heritage, is given according to the archival documents in addition to the literature review concerning the prevalent use of digital techniques for conservation of cultural heritage. Afterward, the whole process of a rehabilitation project for the study area is described by exemplifying the use of different digital tools in the methodology part of this study while focusing on the importance of photogrammetric documentation and systematic evaluation with the help of geographic information systems. At the end of the study, the contribution of digital technologies to find out the authenticity of Ali Pasha Caravanserai is discussed, while mentioning the importance of a holistic approach to the preservation and perception of a historic urban area in Bursa. By this, it would be possible to estimate how much an architectural heritage that has not physically conserved its integrity may be kept alive against urban change activities, which could help create a rehabilitation strategy for its historic surroundings.

2 Study Area

This paper describes and illustrates the contribution of related digital technologies for assessing the general principles and interventions required to sustain a historic urban site. The study area is located in the eastern part of the historic commercial center of Bursa, which has five World Heritage Sites in Turkey containing various types of architectural heritage dating from ancient times to the twentieth century. By the way, it includes a sixteenth-century Ottoman Caravanserai together with one- or two-storey historic shops and dwellings. There also exist remains of an anonymous monument, constructed in the early twentieth century after the application of *Tanzimat* Reforms (Beneficial Reforms) in the city center of Bursa. However, the endless demand for development has endangered its historic urban identity, while destroying the unity and authenticity of many commercial monumental buildings, one of

Fig. 1 Past and present views of the *Yeni Galle Pazarı Hanı* **a** The original form of the YGPH (*Source* map of Bursa-1862); **b** The current condition of the study area (*Source* the cadastral map of Bursa, 1999) (photographed by the Author, 2007) (the figure was re-designed by the Author, 2022)

which is *Yeni Galle Pazarı Hanı*[1] (YGPH). This monumental building was originally constructed as sixteenth-century Ottoman Caravanserai—Ali Pasha Caravanserai and then reused as an Ottoman Han building in grain trading under the name of *Yeni Falle Pazarı Hanı*. The new roads, named *Cumhuriyet (Hamidiye*) and *İnönü (Gemlik)* Streets, were applied after Westernization Movements in the historic city core of Ottomans Bursa, at the beginning of the twentieth century, which resulted in the loss of authenticity in form and function as well. Since this pressure of change and development has increased to this day in this part of the historical trade center, this architectural heritage within the study area was subjected to additions and demolition and has never regained its original form.

This Ottoman Caravanserai has a structure composed of stone masonry supported with mudbrick. The north wall is still standing, and it includes one of the gates of the Han, which helps us understand its original construction technique and material. There are eight units ordered along its southeast corner, seven of which are composed of one- and two-storey shop buildings. However, it is impossible to read its original construction material, building mass, and structural system since the walls were plastered with cement on the façade (Fig. 1). There is a terrace along the inner

[1] In English named as: New Grain Bazaar Han.

Fig. 2 The existing (and separated) remains of the YGPH **a** the vaults on the east side of the YGPH; **b** small rooms on the upper storey and west side of the YGPH; **c** partially collapsed vault at the southeast corner of the YGPH (photographed by the Author, 2007)

façade of the Han, looking to the courtyard and serving as a semi-closed corridor/balcony. However, the original covering above that corridor/balcony was severely damaged and mostly collapsed. Only the corner of the upper floor is still standing, and it shows the covering structure of a dome. Its masonry structure also connects the columns with arches, combining brick and stone (Fig. 2).

After this fragmentation, six new shops were built along the street and five shops were separated from the spatial organization of the Han, which resulted in the deterioration of its integrity. Contrary to the units of the YGPH, the entrances of these small two-storey shops are from the road instead of the courtyard (Çakıcı, 2008). Meanwhile, a new Ottoman Han was constructed along Cumhuriyet Street and was attached to the destroyed walls of the YGPH, to be used as a grain depot. However, there exist just a few units of this late-period Ottoman Han separated within the study area, which makes it impossible to read the holistic architectural character of this monumental building (Fig. 3). Since it has historical value conserved, this late-period Ottoman Han is included in the field studies and analyses related to the study area. In addition to these monumental historic buildings, there also exist two-storey traditional dwellings being used for both commercial and residential demands. Despite being abandoned and collapsed in history, a few of them are still standing with their gardens. In contrast, there have also been new multi-storey apartment blocks within this historical urban site since the 1970s.

3 Method and Materials

Digital technologies can facilitate the complex structure of the collected data and share them with the related stakeholders and local authorities. For instance, new studies have been initiated by the Turkish Ministry of Culture and Tourism under the National Immovable Cultural Heritage Inventory System (TUES in Turkish), which reveals the impact of collective participation of national and international institutes in a management plan to conserve historic areas, including immovable cultural heritage in different regions of Turkey (Boz et al., 2014). It is also necessary to use the correct tools to prepare a model for revealing the different stages of a conservation project, by using digital tools for a systematic assessment of the changes in this historic built environment, due to its multi-layered urban character. As being the prevalent one, GIS is selected as an effective management tool capable of displaying and analyzing all forms of geographically referenced information, for this case, since it is possible to propose the optimal intervention approaches that will best sustain these cultural heritage buildings and the overall site. Moreover, it is able to vectorize and rasterize the required digital maps to assess the systematic process of rehabilitating a historic urban environment. Hence, the geographical characteristics of the study area can unify with the collected archival information concerning related architectural heritage.

The practice for the perception of such kind of mostly collapsed historic monument and its hardly damaged historic environs is processed under the four following phases (Fig. 4), which also vary according to the types of spatial objects defining the historic built environment within the study area (Fig. 5).

3.1 Documentation Phase

Both visual and written sources[2] gathered from the archives and literature were used to prepare survey sheets and base maps for the site survey and the following phases of the study (Fig. 6). The inclusion of primary visual documents,

[2] For detailed information about the primary and secondary sources of the study, please look at (Çakıcı, 2008: 5–8).

Fig. 3 The traces of historic buildings outside of the boundary of the YGPH: **a** Five small units of the YGPH built along *Çancılar* Street **b** photographs of the *twentieth-century* Ottoman Han newly attached to YGPH from its courtyard (photographed by the Author, 2007)

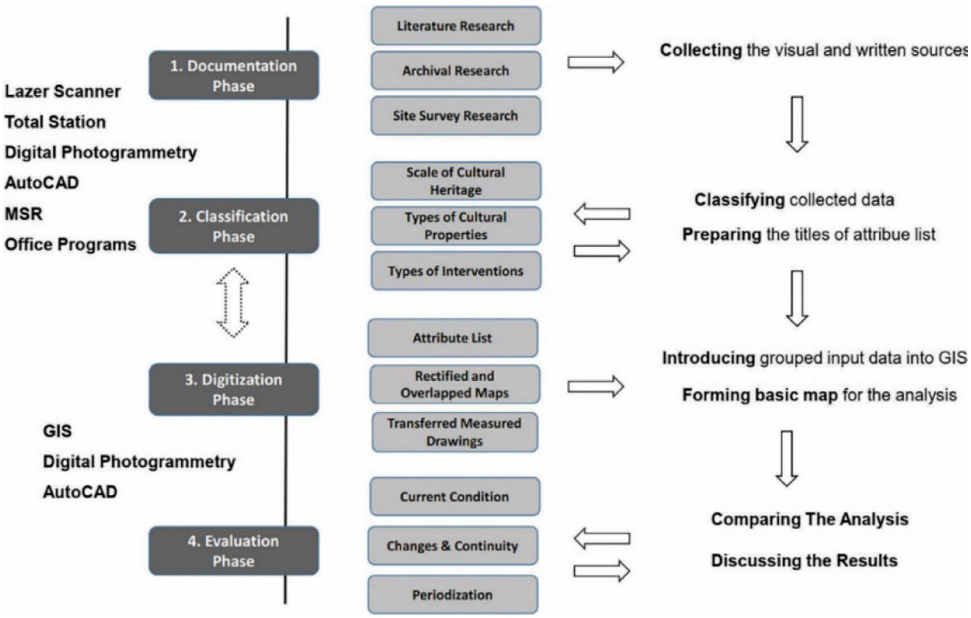


Fig. 4 The phases of the rehabilitation process of the YGPH (prepared by the Author, 2022)

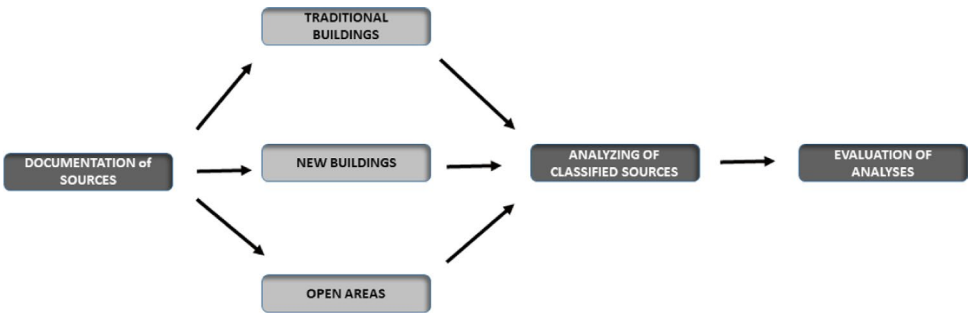


Fig. 5 Phases of research method used in assessment of different types of spatial objects identifying the study area (prepared by the Author, 2022)

such as the old maps, current master plans, and aerial photographs, was essential to achieve reliable and accurate results when comparing the past and present conditions of the study area. In addition to the literature and archival documents, the visual information gathered from cadastral maps (from 1933 to 2001) and aerial photographs (from 1938 to 1984) primarily defines the breakpoints in the transformation of urban history while revealing the changes in the boundary of building lots and buildings. Digital photogrammetry was also used, as it is the most accurate and precise way to get systematic, detailed visual information about the cultural heritage for GIS data entry

SOURCES	Archival Research	Literature Survey	Site Survey
Written Documents  Conservation Decisions approved by the Council	- Inventory and Waqf receipts **Registration List** and **Conservation Decisions** approved by the Council The Decisions and Principles of the Plans	- **Contemporary writings** - **Ottoman Court Registers** - **The Memories** (*Seyahatname*) of the travellers	- **Notes** taken during site survey - **Survey sheets** filled with survey documentary
Visual Documents  Skecth of Resititution Drawing prepared for YGHPH by the reporter of the Council	- Old and New **Photographs of cultural properties** taken by the experts / the reporters / the staff from Municipality **Aerial photographs** **Historic maps** Old and New **Cadastral Maps** Contemporary **Master Plans** **The Conservation Development Plan of HCC** (*Reyhan-Kayhan-Hanlar* Districts) - **Sketches** drawn by the reporters / experts	- **Sketch Plans and Photographs** from the literature - **Sketches** from Travellers' books	- **Sketches** - **Measured Drawings** - **Photographs** taken by the author

Fig. 6 Documentation phase (prepared by the Author, 2022)

in the following phases. Data transfer within the advanced documenting and modeling techniques, from CAD to GIS programs, was frequently used in this study to provide synchronization between different types of information about cultural heritage at different scales.

The scattered remains of the YGPH with complicated geometry or damaged structures required advanced digital documentation techniques, such as laser scanners and total stations, to get quick and accurate measured drawings (Fig. 7). It was possible to draw the plans and façades of the historic remains of the YGPH, which cannot be reached and measured manually while producing the silhouette and street sections of the study area. Meanwhile, it was easy to complete the section drawings according to a base map and optical measurements. The surface photographs were rectified in MSR (single image photogrammetry) to complete façade drawings in the CAD format. Hence, it was easy to document the current spatial and structural conditions of the remains in detail using both manual and digital methods.

3.2 Classification Phase

The collected documents are grouped systematically according to the information that differentiates about types and scales of cultural heritage, which would be a base map for preparing the attribute table to be used in the digitization phase (Figs. 8 and 9). The built environment within the study area was classified under the names of "historic buildings" and "new buildings" together with their surrounding "open areas". The historic buildings are composed of "the Ottoman khans", named "*Yeni Galle Pazarı Hanı*" and "*Yeni Tahıl Han*", together with the historic dwellings settled at the north side of the study area.

On the other hand, the open areas in the study area vary according to the land use, such as 'courtyards of the khans', 'gardens of the houses', and 'the streets'. The courtyards and the streets are open to public use while the gardens are known as private open areas. The overlap of various types of written and visual documents helps to prepare the

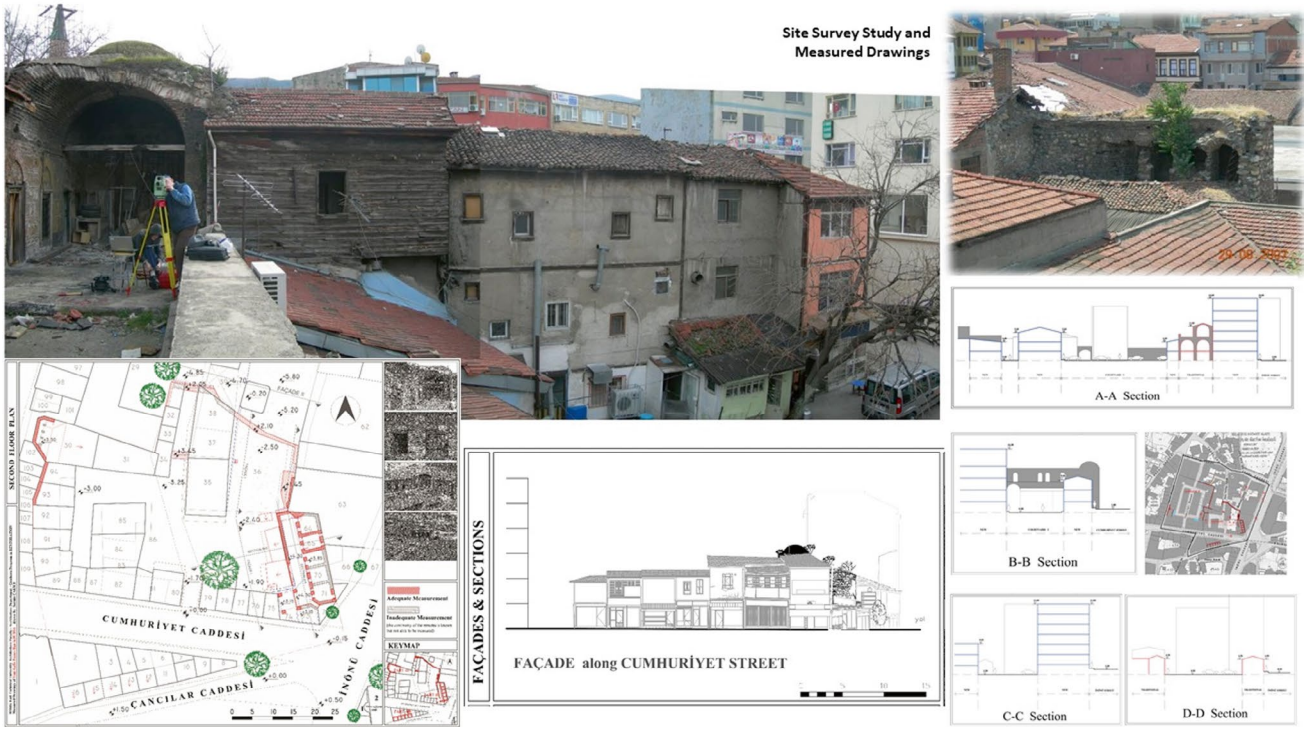


Fig. 7 The documentation process of the YGPH in the site survey study (prepared by the Author, 2008)

Fig. 8 Matrix revealing classification phase (prepared by the Author, 2022)

TYPE of INFORMATION		SPATIAL OBJECTS of the Study Area						
		Site Scale						
		Building Scale				Open Area(s)		
		Historic Building(s)			New Building(s)	Courtyard(s)	Garden(s)	Street(s) / Road(s)
		Yeni Galle Pazarı Hanı	Tahıl Han	Dwellings				
MASS	Original	✓	✓			✓		✓
	Current			✓	✓		✓	✓
FUNCTION	Original	✓	✓			✓	✓	✓
	Current	✓	✓	✓	✓	✓		✓
CONSTRUCTION TECHNIQUE	Original	✓			✓			
	Current	✓	✓	✓	✓			
PARCELLATION and OWNERSHIP	Original	✓	✓					
	Current	✓	✓	✓	✓	✓	✓	
STRUCTURAL MATERIAL	Original	✓	✓		✓	✓		
	Current	✓	✓	✓	✓			
FLOOR COVERING	Original					✓	✓	
	Current	✓	✓	✓	✓	✓	✓	✓
CONSERVATION STATUS	Registered	✓		✓		✓		
	Not Registered		✓	✓	✓		✓	

Fig. 9 Attribute table prepared for both the identification and alteration of spatial characteristics (prepared by the Author, 2022)

Attribute(s) List		Spatial Objects of the Study Area				
		Site Scale				
		Building Scale			Open Area(s)	Street(s) / Road(s)
		Historic Remain(s)	Historic Building(s)	New Building(s)		
Identifying Information	Name / Type	✓	✓	✓	✓	✓
	Location	✓	✓	✓	✓	✓
	Construction / Formation Date		✓			✓
	Period		✓		✓	✓
	Original Form		✓	✓		
	Current Form	✓		✓	✓	✓
	Original Function		✓		✓	✓
	Current Function		✓	✓	✓	✓
	Dimension(s)	✓	✓	✓	✓	✓
	Original Construction / Surface Material	✓		✓	✓	
	Structural Condition	✓	✓	✓		
	Condition of Construction Material	✓	✓	✓		
	Ownership Status	✓	✓	✓	✓	
	Conservation Status	✓			✓	
	Registration Status		✓		✓	
	Registration Date		✓		✓	
Changing Information	Type of Change	✓	✓	✓	✓	✓
	Date / Period of Change		✓		✓	✓
	Degree of Change	✓	✓			
	Reason of Change		✓	✓	✓	
	Result of Change	✓	✓		✓	✓

attribute list in this grouping phase. The collected information was used to classify the spatial objects, according to the original and current 'spatial characteristic', 'use', and 'structural condition' of the YGPH and its environs, while also being grouped according to the 'construction dates', 'conservation status', and 'ownership'.

3.3 Digitization Phase

Methods such as 3D scanning and rendering with high-fidelity graphics contribute to recreating scenes from the past. Laser scanners provide a method of capturing accurate information about the facades of historic buildings in

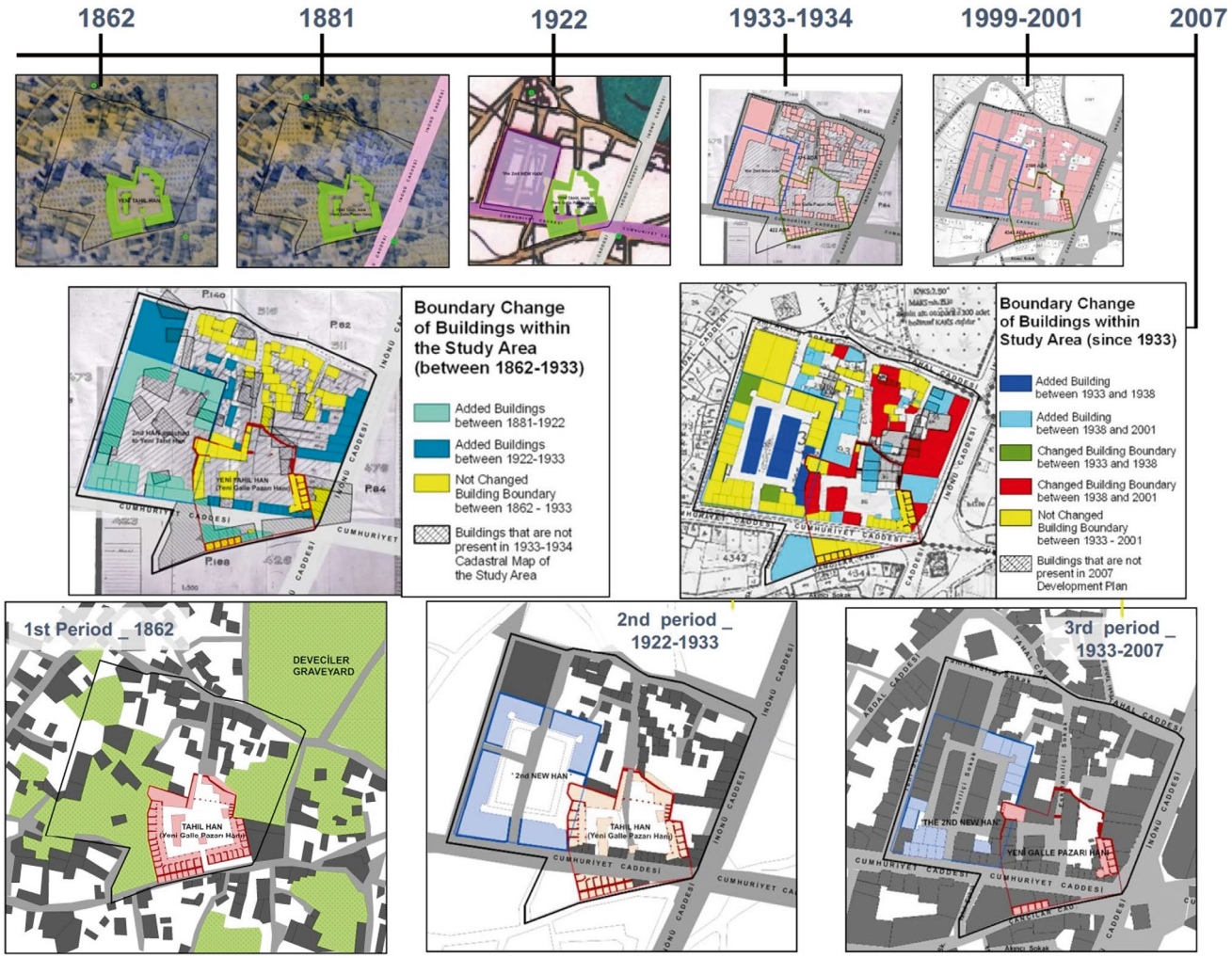


Fig. 10 Timeline of mobility in the study area as a result of digitization of different types of sources: **a** Chronological order of historical background, **b** Changes in buildings and building lots, and **c** Three periods of restitution processes in the study area (Çakıcı, 2008: 111–122) (the figure was re-designed by the Author, 2022)

the study area. Before entering the input data, the classified information was defined with a code number or code title to be recorded into the system according to the format of the designed database. This coding system helped establish a utilizable documentation system structure by indicating an ID number for each spatial object to record collected data related to the immovable cultural properties.[3] The printed visual data, including aerial photographs, sketch drawings, historical maps, and current cadastral maps, were converted into digital format to prepare for the attribute data entry of written information. It was also possible to visualize the past and present urban character by superimposing historical maps from 1862, 1881, 1907, 1922, and 1933, which are used to produce base maps defining the breakpoints in the rehabilitation history of the study area (Fig. 10).

The master plans, measured drawings, and cadastral maps needed to be rectified due to deviations in coordinate information. By doing this, both spatial (the graphics of traditional buildings and the new ones on site) and non-spatial data (ownership, usage, and conservation status of the land) were overlaid in GIS. Thematic clusters of spatial objects and attribute data could be combined in the GIS format to be shown on a base map, which provided a detailed assessment in the following phases of the study. Moreover, the literature information about the architectural character of the study area was adequate to be entered as input data. Consequently, it becomes easy to study all types of urban development and transformation activities seen in and around this architectural heritage site by preparing a digital map.

[3] For detailed information about the input data digitized by using GIS and Autocad, please look at (Çakıcı, 2008).

3.4 Evaluation Phase

This phase involved methods to make various queries over the developed database to visualize the problems, values, and potentials of the study area. It was essential to present the current state of the study area via the analyses and evaluation phases due to the superimposition of both visual and written documents in GIS. Accordingly, a multi-criteria analysis was performed to evaluate the development of the built environment for various land-use categories, including high-and low-storey buildings, together with waste disposal and natural conservation. For instance, the relationship between building lots, buildings, and streets needed to be evaluated through categories and queries like "ownership", "structural condition", and "functional continuity", which would describe the spatial and socio-economic characteristics of the study area (Fig. 11). It was essential to study the current condition of the new buildings constructed within the courtyard of the Hans to have an idea about how incompatible they are with the visibility of the traditional urban fabric of the study area. It was also easy to understand the conservation statuses of the historic buildings while also proposing strategies for the sustainability and integrity of the study area. Hence, the analyses were evaluated by overlapping several base maps in the GIS environment to determine the amount of deformation in the study area, which helped create sustainable solutions in the rehabilitation process (Çakıcı, 2008: 123–139).

4 Analysis and Results

The contribution of GIS is emphasized in the assessment of such complicated input data describing the study area while helping to understand the problems and the risks against the continuity of unity and authenticity of this architectural heritage, in this part. By this, it would be possible to follow the changes and originality in the study area, to produce a proper rehabilitation project including reconstruction and renovation of existing remains of the historic buildings by using advanced digital technologies. The results are composed of analyses that are mostly evaluated by using GIS to discuss the conservation approaches proposed in this study, which also emphasize the difference between the current and recommended physical condition of this historic built environment. In addition to the outcomes determining the alteration in original spatial character, the input data concerning the changes in parcellation reveal also the alteration in ownerships and functions of the historic buildings, which would help us to find out the breakpoints in the transformation and conservation history of the study area. Hence, the analyses explaining "the past and present conditions of the historic built environment in the study area" are compared to evaluate "the breakpoints in its urban transformation and conservation history".

In addition to the site survey analysis, archival documents, such as reports and sketches prepared by experts,[4] also give vital information to be used as input data to follow the changes in the structural characteristics of the YGPH. The vibration caused by the vehicular traffic passing through the gates, which mark the northern entrance of this monument, has been reported as a threat to the stability of the two arches in the gates. In another report, dated 1984, almost all of the brick vaults covering the balcony in front of the rooms on the upper floor of the Han had already collapsed. It becomes easy to follow these written data via the base map that describes transformation activities in the study area with the help of the collaborative use of GIS and other digital documentation techniques (Fig. 12). On the other hand, the accessibility by vehicular traffic passing through the site has become both a problem and potential for sustainability of this historic area, which would also be discussed to find out proper decisions conserving the traditional texture.

The changes in boundaries of buildings and building lots within the study area were discovered by overlapping two different cadastral maps together with aerial photographs in a standard base map, with the help of GIS (Fig. 13). As a result of comparing two cadastral maps, from the 1930s to the 2000s, the changes in ownership and use of the parcels, including the historical buildings and remains, can be determined to discover the transfer from public to private ownership in the study area. This analysis can explain why rapid and improper new building construction appeared in this historic built environment. As a result of overlapping the building height information from these cadastral plans, it is also clear that the architectural character and the current uses of the new incompatible building constructions in the study area have not been in harmony with the mass and height of the traditional commercial fabric since the 1970s. Consequently, these analyses reveal the loss of authenticity and unity of the Hans together with surrounding dwellings.

While discovering the historical background of the YGPH and its environs, a restitution practice was done to reveal the original design of this monument and describe the changes in its spatial character. Its reliability depends on the sources concerning the results of the site survey analyses, together with primary written and visual documents

[4]These experts are named Dursun Öcalan (archaeologist), Veysi Cengiz (reporter in BKTVKBK), Recep Okçu (researcher from the museum of Bursa), and Erdal Korkmaz (archaeologist). They prepared and gave the reports together with various photographs and sketches archived in the Conservation Council, between 1981 and 1987, just before the preparation of the CDP.

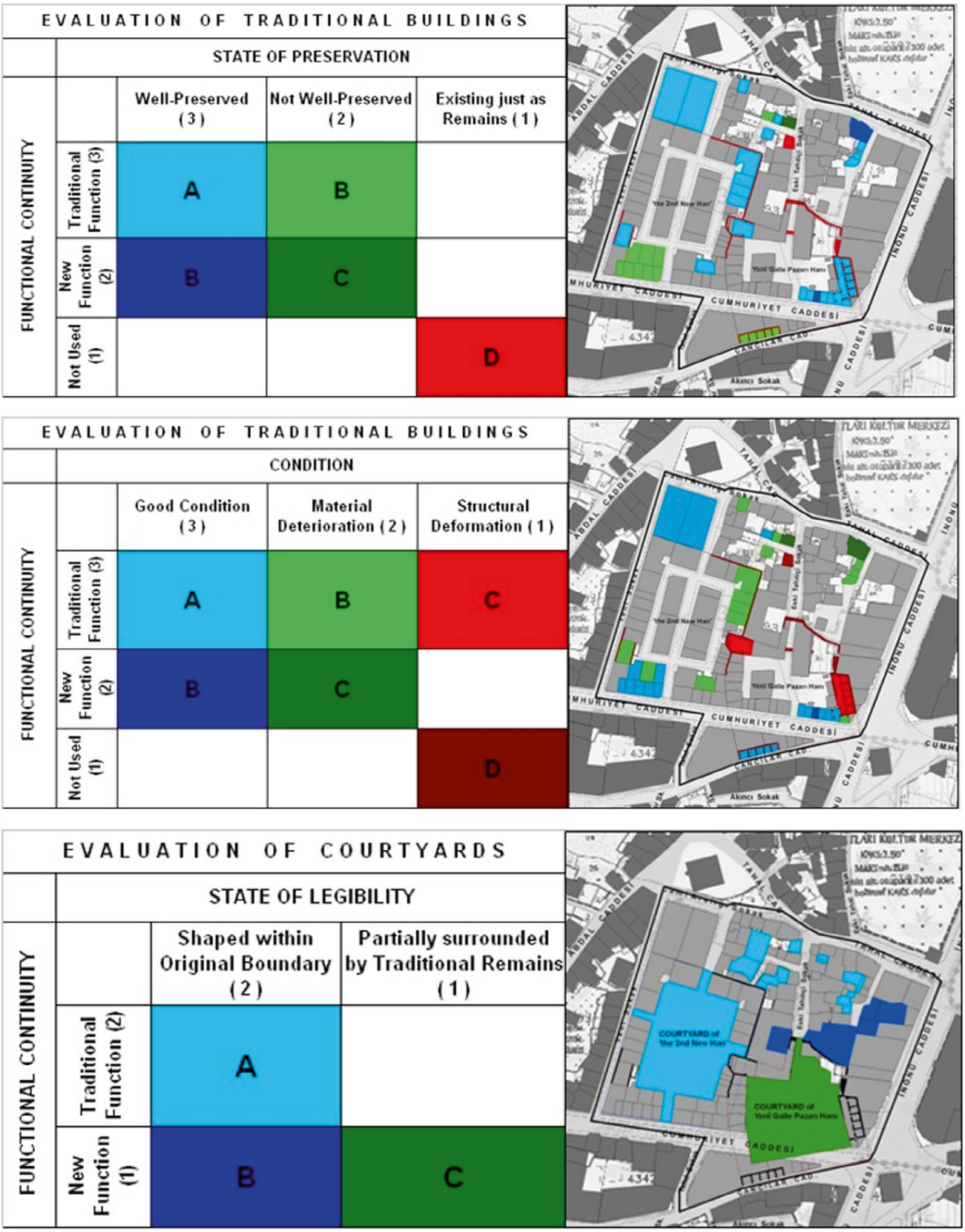


Fig. 11 The matrix tables reveal the relation between functional continuity, state of legibility, and conservation of historic buildings and courtyards as historic landscapes (Çakıcı, 2008: 125–128) (the figure was re-designed by the Author, 2022)

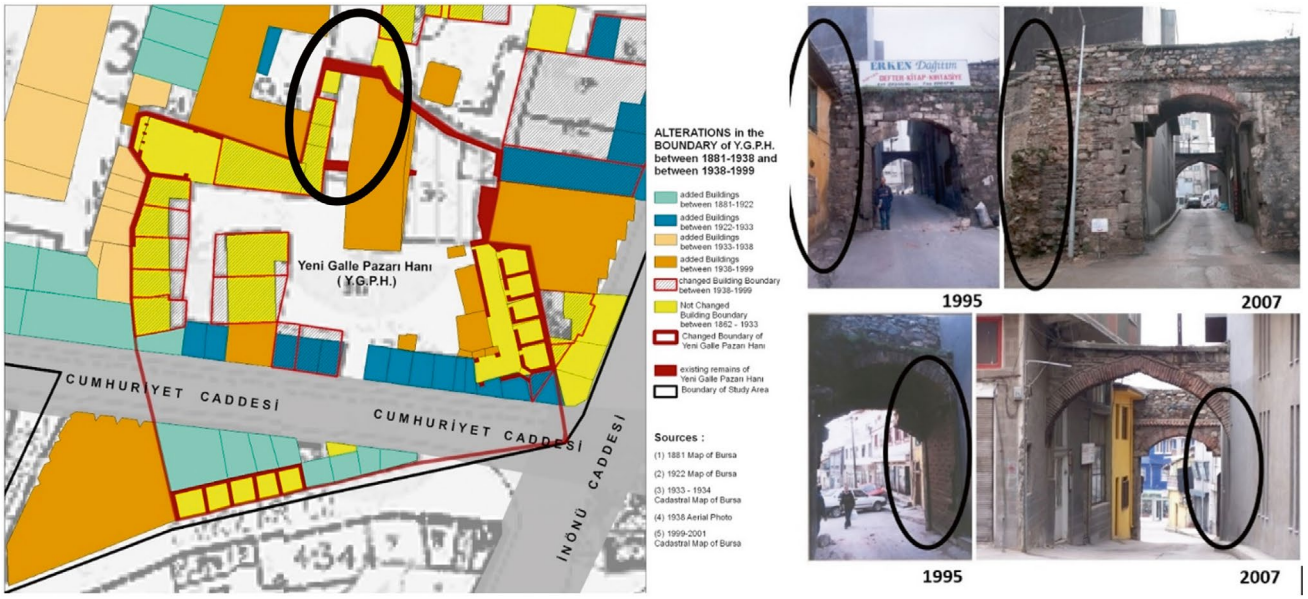


Fig. 12 Changes in and around the YGPH, according to the archival reports from the Conservation Council of Bursa (the 1980s to the 2000s) (*Source* photographs from the archive of the Regional Conservation Council of Bursa, 2007) (the figure was re-designed by the Author, 2022)

that were classified, overlaid, and combined as input data in digital formats using GIS. These data made it possible to draw the original form and structure of this architectural heritage and its surrounding historic built environment. Accordingly, there appear three breakpoints since the mid-nineteenth century in which we can classify the changes in the spatial and non-spatial character of the study area (Fig. 10):

- 1st period: The original state before the application of the Tanzimat Reforms (1862–1881).
- 2nd period: The first changes after the application of the Tanzimat Reforms (1881–1933).
- 3rd period: Changes in parcellation and the infill of new building constructions (since 1933).

As a result of this comprehensive evaluation phase, it would be easy to propose sustainability and perceptibility of such a multi-layered historic area, by using digital technologies in virtual ways of display in different environments, which would also raise public awareness about the historical and architectural value of this urban site. Accordingly, the basic principles that will maintain the historical character of the existing artifacts and monumental structures in the study area are defined below:

(1) The vehicular traffic through the study area should be managed.
(2) The public open areas should be reorganized in harmony with the traditional character of the courtyards of related two Han buildings.
(3) New constructions should be controlled to keep the historic urban areas still useful and sustainable.
(4) The existing traditional buildings and remains of the Caravanserai should be represented in contemporary ways of conservation.

The intervention decisions, proposed for the continuity of the historic architectural and urban character of the study area, are mapped in the 2D and 3D formats, concerning these basic conservation principles (Fig. 14). In addition to the documentation of the extant remains, excavating the underground parts of this monumental building would be helpful in corroborating information on the 1862 map of Bursa. If any foundations are found during excavations, they can be displayed via a raised pavement with transparent material, which would create public awareness about the original form of this cultural heritage. Some parts of the Han could also be physically reconstructed based on information uncovered from the excavations. On the other hand, in this part of the historical trade center of Bursa, it is essential to use digital methods and virtual techniques to obtain appropriate results in the preservation, presentation, and rehabilitation of the existing remains of the relevant architectural heritage with minimum error. As mentioned in the Venice Charter (1964),[5] the restoration work should stop at

[5]The 9th article of Venice Charter (1964); "*The process of restoration is a highly specialized operation. Its aim is to preserve and reveal the aesthetic and historic value of the monument and is based on respect for original material and authentic documents. It must stop at the point where conjecture begins*". (https://www.icomos.org/charters/venice_e.pdf).

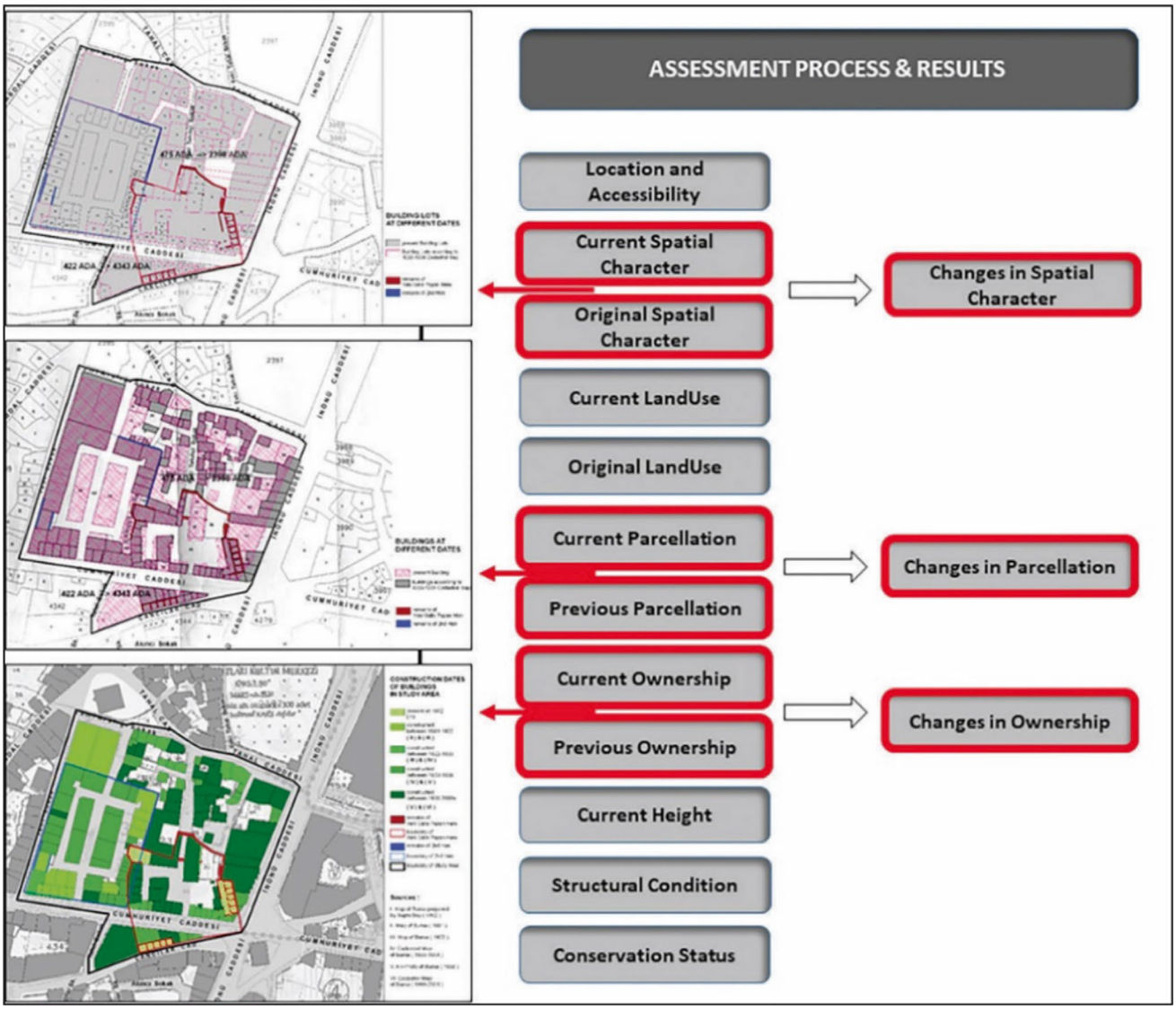


Fig. 13 The titles of analyses and results in evaluating process, visual sources: (1) old and new cadastral maps (1933–1934 and 1999–2000); aerial photographs of Bursa (1938, 1943, 1973, 1984, and 1997) (*Source* the plans from Çakıcı, 2008) (the figure was re-designed by the Author, 2022)

the point where conjecture begins. Moreover, reconstruction is defined in the Burra Charter of Australia ICOMOS (1999),[6] as possible to complete the heritage if it was damaged or altered. Hence, it is significant to have sufficient evidence to reproduce an earlier state of the fabric, which is not clear enough to complete the rest of the YGPH in physical reconstruction. Since it is hard to reconstruct the whole structure of mostly collapsed Hans, it is preferred to mention the continuity of the foundation on the ceiling over the street or by use of contemporary new construction techniques proposed to be differentiated from the others. During the whole process of the rehabilitation project, collaborative digital portals should be accepted as important technical tools to reconstruct the remains of the architectural heritage, whereas significant research tools are also regulated to promote the cultural structure of both existing and new audiences.

5 Conclusion

The use of advanced digital technologies in the presentation of an architectural heritage not only provides a proper preservation standpoint but also allows researchers to envision

[6]The 20th Article of revised version of Burra Charter of Australia ICOMOS (1999), "*… Reconstruction should be identifiable on close inspection or through additional interpretation*". (https://australia.icomos.org/publications/burra-charter-practice-notes/burra-charter-archival-documents/).

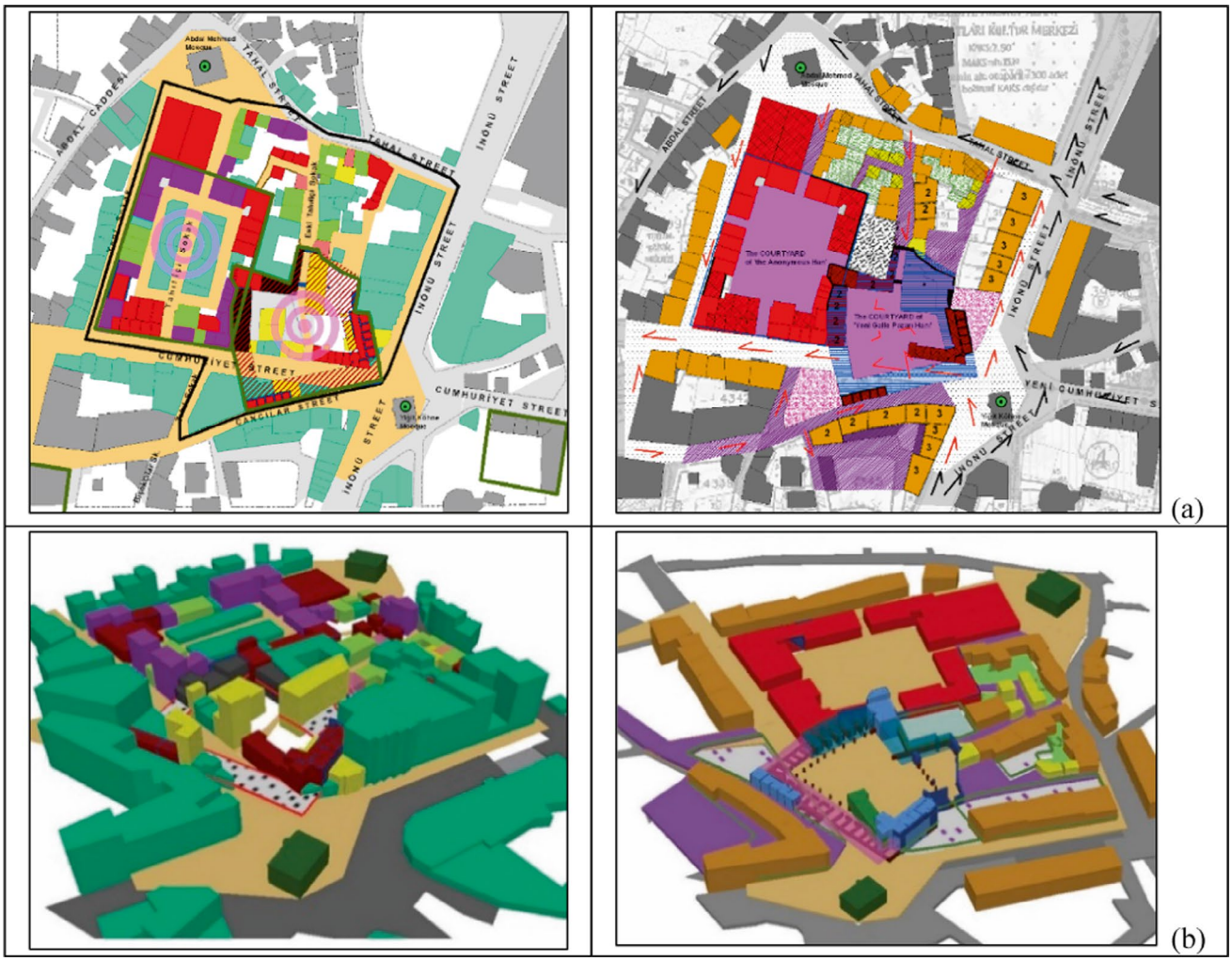


Fig. 14 The drawings of the study area revealing before and after the application of the proposed rehabilitation project for the study area (prepared by the Author, 2008)

the works in the context of their intellectual and historical significance. They help to find compatible solutions for the sustainability of architectural heritage through spatial analysis and the development of databases collected from the archives and the literature. This study aims to make a significant contribution to the re-presentation of the multi-layered urban identity of historic areas, including architectural heritage required to be restored, by using advanced digital technologies in addition to manual ones. It should be accepted as a model that exemplifies the systematic evaluation of a comprehensive and complex database on a digital platform to understand the past and present sustainability of the sixteenth-century Ottoman Caravanserai, constructed within the historical trade center of Bursa that is known as a World Heritage Site since 2014. As one of the most preferred tools for digitalization, Geographical Information Systems (GIS) enable researchers to develop sustainable strategies by evaluating the chronological and spatial distribution of changes in cultural heritage both on building and site scale. This would also aid in the systematic evaluation of complicated data comprising the geographical, architectural, social, and economic characteristics of a historic urban area.

Investigation of the urban transformation history in its surrounding landscape and built heritage makes it possible to evaluate the changes in a historic urban environment including different types of Ottoman period historic buildings. Accordingly, the original spatial characteristic of the study area becomes perceptible in digital ways of presentation together with the analysis formed with the help of GIS. This method would help to take basic conservation principles on the study area as following international regulations on the sustainability of such a multi-layered and badly destroyed traditional urban texture, including two historical monuments to be conserved and reused. Consequently, the related researchers and other stakeholders would discuss

the proper decisions for a management plan for such a continually developing and changing historic urban area. On the other hand, a wealth of sources must be analyzed and evaluated systematically to prepare a rehabilitation project. In this study, the inadequacy of the analyses on the deterioration of the façade material and the structural load-bearing situation meant that GIS was not sufficiently detailed for some queries.

According to the results of the assessment on this issue, it is suggested that the monumental buildings and their surrounding historic dwellings within the study area can be visualized and displayed with the help of virtual reconstruction. In this way, it is possible to interpret, analyze, and understand the complexity of this architectural heritage, without changing the structural condition of its standing urban texture. In the case of YGPH, which could not preserve its original form, the additions and still-existing parts of this monument can be marked and distinguished in a virtual space. Hence, it is possible to distinguish the elements, whose original form is known, within the physically non-existing, virtually reconstructed part in the 3D models of the site. Besides, there is no need to visit on-site, while it can be made available everywhere, not just for tourist presentations but also for scientific applications in the restoration process. More importantly, this type of technique can be performed in a significantly shorter time than physical restorations with architectural interventions. Consequently, the practice of the discovery of such a multi-layered historic area would be an excellent case study for the perceptibility and continuity of architectural heritage by digitally unifying resources concerning international and national declarations on the conservation of cultural heritage.

References

Boz L, Gülbay Y, Aygün HM, Erdoğmuş E (2014) Turkish national immovable cultural heritage inventory system. AgroLife Sci J 3(1):35–38

Çakıcı S (2008) A proposal for rehabilitation and preservation of Yeni Galle Pazarı Hanı (former Ali Paşa Kervansarayı) and its immediate surrounding in Bursa [Unpublished Master Thesis, Middle East Technical University]. Graduate Program in Restoration (Conservation of Cultural Heritage), Department of Architecture, METU. http://etd.lib.metu.edu.tr/upload/12609591/index.pdf

Ciski M, Rząsa K, Ogryzek M (2019) Use of GIS tools in sustainable heritage management—the importance of data generalization in spatial modeling. Sustain 11(20):5616. https://doi.org/10.3390/su11205616

Del Pozo S, Rodríguez-Gonzálvez P, Sánchez-Aparicio LJ, Muñoz-Nieto A, Hernández-López D, Felipe-García B, González-Aguilera D (2017) Multispectral imaging in cultural heritage conservation. Int Arch Photogramm, Remote Sens Spat Inf Sci 42:155–162

Droj G (2010) Cultural heritage conservation by GIS. In: GISOpen conference, p 6.

Duran Z, Toz G (2002) Using 3D GIS for documentation of historical monuments. In: ISPRS commission V symposium close range imaging, long-range vision, 2–6 September, Corfu. pp 354–359

Fletcher R, Johnson I, Bruce E, Khun-Neay K (2007) Living with heritage: site monitoring and heritage values in greater Angkor and the Angkor world heritage site, Cambodia. World Archaeol 39(3):385–405. https://doi.org/10.1080/00438240701465001

Gomes L, Bellon ORP, Silva L (2014) 3D reconstruction methods for digital preservation of cultural heritage: a survey. Pattern Recogn Lett 50:3–14

Heras VC, Wijffels A, Cardoso F (2013) A value-based monitoring system to support heritage conservation planning. J Cult Herit Manag Sus-Tainable Dev 3(2):130–147

Heuberger R, Leone LE, Evers R (2015) The challenges of reconstructing cultural heritage: an international digital collaboration. IFLA J 41(3):223–229

Karslı F, Ayhan E, Tunç E (2003) Building 3D photo-texture model integrated with GIS for architectural heritage conservation. In: CIPA, new perspectives to save the cultural heritage, commission V, WG II, four pages, Antalya, Turkey

Lewińska P, Zagórski P (2018) Creating a 3D database of Svalbard's historical sites: 3D inventory and virtual reconstruction of a mining building at camp asbestos, Wedel Jarlsberg Land, Svalbard. Polar Res 37(1):1485416

Ludwig M, Herbst G, Rieke-Zapp D, Rosenbauer R, Rutishauser SM, Zellweger A (2013) The advantages of parametric modeling for the reconstruction of historic buildings. Int Arch Photogramm, Remote Sens Spat Inf Sci 40:161–165

Pavlidis G, Tsiafakis D, Koutsoudis A, Arnaoutoglou F, Tsioukas V, Chamzas C (2007) Preservation of architectural heritage through 3D digitization. Int J Archit Comput 5(2):221–237

Piwowar JM (2005) Digital image analysis, remote sensing for GIS managers. In: Aronoff S (ed), ESRI Press, Redlands, California

Ragia L, Sarri F, Mania K (2015) 3D reconstruction and visualization of alternatives for restoration of historic buildings: a new approach. In: 1st international conference on geographical information systems theory, applications, and management (GISTAM), pp 1–9

Ruoss E, Alfare L, Vallone R, Moscatelli M (2013) GIS as tool for cultural heritage management. In: Ruoss E, Alfare L (ed) Sustainable tourism as driving force for cultural heritage sites development, planning, managing, and monitoring cultural heritage sites in South East Europe, project name: CHERPLAN, Chapter: 3.5. CNR IGAG, pp 42–46

Perez ASH, Diego TM, Carreras NP (1999) Digital photogrammetry integration possibilities to heritage record by an architectural information system. In: Proceedings of CIPA 1999 international symposium. Olinda, p 5. https://www.cipaheritagedocumentation.org/wp-content/uploads/2018/11/Hern%C3%A1n-P%C3%A9rez-e.a.-Digital-photogrammetry-integration-possibilities-to-heritage-record-by-an-architectural-information-system.pdf

Vileikis O, Cesaro G, Quintero MS, Van Balen K, Paolini A, Vafadari A (2012) Documentation in world heritage conservation: towards managing and mitigating change-the case studies of Petra and the silk roads. J Cult Herit Manag Sus-Tainable Dev 2(2):130–152

A Metaheuristic-Based Subspace Search Approach for Outlier Detection in High-Dimensional Data Streams

Imen Souiden, Zaki Brahmi, and Mohamed Nazih Omri

Abstract

The continuous progress in technology is leading to the widespread existence of data streams with high dimensions. Identifying outliers in this particular scenario presents a notably difficult task. The unique characteristics of data streams, combined with the effect of the dimensionality curse in high-dimensional space, create constrained mining requirements, and a current challenge is to simultaneously address them. A common approach to handle high dimensionality is to identify outliers only within subspaces of space of features that contain interesting knowledge, where outliers are typically found. However, in the realm of data streams, this area of study has not been well explored. In this article, our objective is to discover interesting subspaces for outlier detection while accommodating the needs of data streams, including limited time and memory, and addressing the adaptation to data changes (concept drift), as well as providing better performance than the closely related approaches. In this context, we used a metaheuristic-based approach (Adapted Binary Gravitational Search algorithm) to discover high-contrast subspaces comprised of independent features, within which the outlier detection will be performed. To deal with data streams, we adopted the sliding window structure together with a modified version of the N-Dimensional Kolmogorov–Smirnov WindoWin (NDKSWIN) concept drift detector. We conducted experiments on both synthetic and real-world data and the results demonstrated its effectiveness and superiority over the competitors.

I. Souiden (✉) · M. N. Omri
MARS Laboratory, Sousse, Tunisia
e-mail: imen.sui@gmail.com

M. N. Omri
e-mail: mohamednazih.omri@eniso.u-sousse.tn

Z. Brahmi
Taibah University, Madina, Saudi Arabia
e-mail: zakibrahmi@gmail.com

Keywords

High dimensionality · Data streams · Subspace search · Outlier detection · Metaheuristics

1 Introduction

Detecting outliers in high-dimensional data streams is fundamental in many fields, including intrusion detection (Xie et al., 2018), fault detection and prevention (Zhang et al., 2017), and more. This process involves identifying irregular patterns that exhibit high deviation compared to normal instances in datasets. The data stream is characterized by being infinite, evolving, and continuously arriving at a rapid pace with an important number of features. The core dilemma in this context is to address simultaneously the challenges posed by the data stream characteristics and the curse of dimensionality. On one side, the outlier detection algorithms must be able to handle the data non-stationary, where the underlying distribution changes with time (Salehi & Rashidi, 2018). Furthermore, it is important to consider the constraints of computational resources, such as limited memory and time, while still maintaining a high level of accuracy. On the other hand, in high-dimensional spaces, identifying outliers becomes challenging due to sparse data distribution and the equidistant nature of the data points. Furthermore, these outliers are often masked by the influence of irrelevant features. Another issue arises, which is the computational resource burden resulting from the high number of features.

A commonly used strategy to tackle these problems is to conduct outlier detection within different subsets of the original feature space. These subsets referred to as subspaces have a lower number of dimensions compared to the full feature space. In a d-dimensional space, there are $2^d - 1$ potential subspaces, a number that grows

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*,
Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_3

exponentially with dimensionality (Souiden et al., 2022b). Consequently, searching for an outlier within these subspaces constitutes an NP-hard problem (Fouché et al., 2021). Due to this combinatorial increase, conducting an exhaustive search over all subspaces is impractical, particularly when dealing with data streams.

A way to tackle this problem is by creating subspaces from randomly selected features. However, this introduces the possibility of including irrelevant attributes (Lazarevic and Kumar, 2005), (Pevnỳ, 2016), (Manzoor, et al., 2018), (Zhang et al., 2022), etc. This technique often sacrifices accuracy with lower computational resources. Alternatively, others opt for a deterministic strategy to find relevant subspaces consisting of meaningful attributes, which allows for more accurate results. However, it faces the disadvantage of having an expensive subspace search operation (Fouché et al., 2021). Additionally, the subspace relevance measures influence the results. Another problem is concept drift, where the relevance of features undergoes changes over time. Thereby, the development of an approach that can address these manifold issues and provide a compromise between the effectiveness of the results and the computational resources becomes crucial.

In this context, we have proposed in an earlier work a Binary Gravitational Subspace Search Algorithm (BGSSA) (Souiden et al., 2022a). In BGSSA, we relied on a deterministic subspace strategy and looked for relevant and general subspaces, which are not constrained by an outlier criterion and can be used for different downstream tasks. For the subspace search, we used a meta-heuristic-based subspace search strategy to mitigate the NP-hard problem associated with this operation. Metaheuristics have shown substantial effectiveness in resolving this problem and have been undergoing significant revolution lately. However, the application of metaheuristics within the domain of data streams remains relatively unexplored. More specifically, we have introduced a customized adapted version of the binary GSA algorithm to obtain the topQ subspaces having the highest contrast measure, which is our measure of relevance. We adopted the definition of contrast proposed by (Fouché & Böhm, 2019) since it possesses several favorable properties for data streams, such as efficiency and robustness. High-contrast subspaces are subspaces constituted of independent features that exhibit distinct separation between outliers and inliers.

To track the change of subspace relevance and data points outlierness with time, we have used a non-overlapping sliding window structure. This structure serves a dual purpose: data stream processing and blind adaptation of concept drift (Bayram et al., 2022). Blind drift adaptation entails the model retraining either upon the arrival of new data or at fixed time intervals, without relying on any drift-detection mechanism. However, a notable limitation of this method is its potential resource intensiveness and unnecessary model updates. Experiments on BGSSA have revealed that BGSSA has satisfactory performance and outperforms existing competitors. Yet the processing time remains an open issue.

To handle data stream evolving nature, various techniques are available, including windowing techniques as used in BGSSA and the integration of concept drift detectors (Bayram et al., 2022). The idea of integrating a concept drift detector involves detecting when concept drift has occurred and then adapting the model to the new distribution. Thus, it is generally less resource-consuming than windowing techniques, and this often depends on the type of concept drift detector employed. Another advantage of this method is its ability to offer information about the dynamics of generating data (Riess, 2022).

This idea has motivated us to propose a novel version of BGSSA denoted by the Adapted Binary Gravitational Subspace Search Algorithm (ABGSSA). This approach aims to ABGSSA incorporate a concept drift detector to improve efficiency by eliminating unnecessary model updates. Specifically, the subspace search is only performed when a concept drift is detected. In this context, we adopted and modified the NDKSWIN concept drift detector proposed in (Togbe, et al., 2021). This detector stands out as a novel and efficient tool used within the context of high-dimensional data streams.

Our main contributions encompass the following: (1) Proposing ABGSSA (Adapted Binary Gravitational Subspace Search Algorithm) algorithm for subspaces search in high-dimensional data streams, which discover high-contrast subspaces using a proposed adapted version of Binary GSA. (2) Proposing a customized version of the NDSWIN concept drift detector, and (3) Evaluating the performance of our proposed approach on both real and synthetic datasets.

The remainder of this article is organized in the following manner: In Sect. 2, we discuss the primary works closely related to our approach. Section 3 outlines the key concepts employed. Section 4 formulates the problem. In Sect. 5, we present the proposed solution. In Sect. 6, we conduct the experimentation on the proposed solutions. At last, Sect. 7 summarizes the work and provides some future directions.

2 Related Work

In recent years, there has been a proliferation of research focused on outlier detection within high-dimensional data streams (Vanea et al., 2012), (Sathe and Aggarwal, 2018), (Manzoor, et al., 2018), (Bhatia et al., 2022), (Zhang et al., 2022), etc. These studies can be divided into two main

groups: full-space-based and subspace-based techniques. Full-space-based methods (Cai et al., 2020), (Chen et al., 2021), (Khalique & Kitagawa, 2021), (Degirmenci & Karal, 2022) assume that all features are equally important. As a result, these algorithms can be influenced by irrelevant features, leading to potential biases. Thereby, diminishing their capability to accurately pinpoint outliers. As for subspace-based methods, they solely depend on one or multiple subspaces of the original set of features to conduct outlier detection.

The one subspace-based methods; Feature extraction (Bhushan et al., 2015), (Xie et al., 2018), (Bhatia et al., 2021), and section (Almusallam et al., 2017), (Yang et al., 2016), (Bhatia et al., 2022) are powerful techniques and widely adopted as pre-processing steps. However, these methods often struggle to detect outliers that are hidden in different subsets of the feature space. To address this issue, multiple subspace-based methods are employed. They identify outliers with subspaces obtained deterministically or randomly.

Random methods (Salehi et al. 2016), (Sathe & Aggarwal, 2016), (Sathe & Aggarwal, 2018), (Manzoor, et al., 2018), (Zhang et al., 2022) opt for a randomized strategy for subspace selection bypassing the resource-intensive subspace search. Nonetheless, there remains a potential challenge of incorporating irrelevant features into subspaces, while potentially omitting relevant ones. This issue becomes particularly pronounced when irrelevant features dominate. Opposite to random methods, deterministic methods have substantially worse efficiency. They rely on search operations to find subspaces that meet specific search criteria and compute outlier scores within these obtained subspaces (Souiden et al., 2022b).

The deterministic methods either seek sparse or relative subspaces. Sparse subspace-based methods (Zhang et al. 2008), (Zhang et al., 2009), and (Zhang et al., 2015) aim to identify subspaces with low density, as low-density regions often indicate the existence of outliers. In this situation, the measures based on density estimation are commonly used to quantify the sparseness of subspaces. SPOT method (Zhang et al. 2008) and its variants (Zhang et al., 2009) and (Zhang et al., 2015) utilized meta-heuristics algorithms, specifically the Genetic algorithm, to discover subspaces that maximize the sparsity measures for each data point. Relative methods, (Vanea et al., 2012), (Zhang et al., 2017), (Fouché et al., 2021) intend to find subspaces composed of relevant attributes. ABSAD-SW (Zhang et al., 2017) looks for relevant subspaces for each data point. A relevant subspace in this context is the set of dimensions that capture most information with regard to the discordance of the data point to its neighbors. As for HPC-StreamMiner (Vanea et al., 2012) and SGMRD (Fouché et al., 2021), they quantify subspace relevance based on the relations between attributes with no reference to any particular data point.

Sparse subspace-based methods as well as the relative subspace-based approaches seeking for relevant subspaces based on outlier criteria are primarily destined for outlier detection, making them task-specific. In contrast, relative subspace-based methods relying on the relation between the features are general, enabling their application across various tasks, which gives them additional importance. HPC-StreamMiner (Vanea et al., 2012), and SGMRD (Fouché et al., 2021) are notable contributions in this context. Both of them leverage the contrast measure to identify subspaces consisting of independent features. The contrast quantifies the disparity between the marginal and conditional density distributions. In HPC-StreamMiner (Vanea et al., 2012), authors discuss a method where they gradually explore high-contrast subspaces over time. They also continuously update the ranking of these subspaces to account for changes in the data. On the other hand, SGRMD (Fouché et al., 2021) focuses on identifying optimal subspaces for the individual features, rather than the entire feature space. It used a Hill-climbing greedy heuristic, together with a multivariate dependency estimator and a novel monitoring method based on bandit theory (Fouché et al., 2019). Experiments in (Fouché et al., 2021) showed that the approach is characterized by good effectiveness due to its ability to find and monitor high-quality subspaces over time. Yet, this comes with an overhead in computation.

On our side, we have introduced in a prior study (Souiden et al., 2022a), Binary Gravitational Subspace Search Algorithm, tailored over windowed data streams. This algorithm aimed to identify subspaces with high contrast, where outlier detection is conducted. This work has been tested and compared to different competitors including SGRMD on different synthetic and real datasets and gave the best performance. Yet, to keep the model updated, the subspace search operation is elaborated for every window, which induces more computational resources. To improve BGSSA, we present the current approach: Adapted Binary Gravitational Subspace Search Algorithm (ABGSSA). In this approach, we intend to reduce the processing time as well as better handling the concept drift aspect. In this context, we have modified and integrated the NDKSWIN concept drift detector (Togbe, et al., 2021).

3 Background

During this section, we will introduce the main concepts needed in our work.

3.1 Adapted Binary GSA

The Gravitational Search Algorithm (GSA) is a population and physics-based metaheuristic algorithm that draws inspiration from Newton's gravity law (Rashedi et al., 2009). It uses a population of particles (search agent), where each particle's mass corresponds to the fitness value of the potential solution (in our situation is the contrast measure). These particles interact via gravitational forces, causing lighter objects with lower mass to move toward heavier objects with higher mass. The heavier objects, which represent better solutions, move more slowly.

The GSA was originally designed to tackle optimization problems with continuous search spaces. However, in binary optimization problems, the search space is designed as a hypercube, and an agent's position is updated by altering one or more bits within its position vector. The binary version of GSA was first introduced in (Rashedi et al., 2010).To convert the continuous algorithm to its binary version, a transfer function (TF) bounded between 0 and 1 is used. This TF determines the probability of changing the position's elements to 0 or 1.

The Adapted Binary GSA presented in our previous work (Souiden et al., 2022a), represents an adapted version of BGSA in the context of data streams. The adaptation consists in obtaining the topQ best solutions instead of a single one. Additionally, we proposed a customized initial population for every algorithm's consecutive generation. This is achieved through the integration of a portion of previously obtained best solutions, enabling us to track the best solutions generated in the previous generation. An algorithm generation represents a separate algorithm call. Within this context, a solution is defined as subspace S. All solutions are encoded using binary strings of uniform length d, where d corresponds to the number of features in F. Each bit within the binary string takes a value of 0 or 1, indicating whether the corresponding feature is included (1) or excluded (0) within a specific subspace. The binary representation serves to capture the selection or omission of individual features within the subspace.

3.2 Sliding Window Structure

A sliding window is a popular stream processing structure that presents an interval set including only a part of data. (recent data) (Nguyen et al., 2015). Consider a window SW with size *ws*, at any time *t*, then we only consider the data in the SW [t−ws+1, t] sliding window. Moreover, sliding the window forward indicates the arrival of the new data. These windows can be overlapping or non-overlapping. If the size of the slide *sz* is less than the window size. Then, each windowed data will experience ws/sz slides during its lifetime. In this case, we can talk about overlapping windows. As the window slides, the algorithm requires a model update. Contrary, when $sz = ws$, we are talking about non-overlapping windows where the algorithm continually operates in one window, i.e., the latest window. Non-overlapping sliding windows are simpler and require fewer model updates, thus, they are used in our context (Dehghani et al., 2019).

3.3 Contrast Measure

The contrast is the measure of subspaces' relevance used in our context. It is designed to detect dependence within a given space, which is a strong indicator of the relevancy of a subspace for data mining tasks. Given a feature space F, a high contrast subspace $S \subseteq F$ denotes a selection of features exhibiting a data distribution having a high dependency among the selected features. To calculate the contrast of a subspace, we adopted the definition outlined in the Monte Carlo Dependency Estimator (MCDE) framework proposed in (Fouché & Böhm, 2019). It quantifies the dependency of a set of features (i.e., subspace) as the degree of independence violation based on marginal and conditional distributions. The more the conditional PDF deviates from the marginal PDF, the more the independence assumption is violated. In this situation, the contrast represents the average deviation between the marginal and conditional distributions estimated through M Monte Carlo iterations and determined using a statistical test. To facilitate the calculation, the authors proposed the creation of an index structure containing feature-related information as a preprocessing step. For more details refer to the original paper (Fouché & Böhm, 2019).

4 Problem Formulation

Consider a sequence of high-dimensional stream tuples $D = \{\langle X_1, t_1\rangle, \langle X_2, t_2\rangle, \langle X_3, t_3\rangle, \ldots, \langle X_\infty, t_\infty\rangle\}, X_n = \{x_1^1, \ldots, x_n^d\}$. t_j is a timestamp indicating the current time and n corresponds to the number of data points $X_n = \{x_1^1, \ldots, x_n^d\}$ arriving at time t. x_n^d is a d-dimensional vector. The full feature space is represented by set of features $F = \{F_1, F_2, , \ldots F_d\}$. A subspace $S \subseteq F$ is a non-empty subset of this feature space with dimensionality $d' < d$.

The goal is to determine if each data point $X_n \in D$ is an outlier or an inlier at any given time t. The scoring of X_n should be performed before $X_n + 1$ and within the existent memory capacity. In high dimensional space, outliers are generally embedded in lower-dimensional subspaces.

For that reason, we consider detecting outliers within relevant subspaces RS rather than the full-dimensional space F. Our approach relies on the notion of contrast presented by (Fouché and Böhm, 2019) as a measure of relevance. In essence, high-contrast subspace rs⊆RS entails chosen dimensions exhibiting a data distribution with a strong dependence. This dependency results in clear clustered structures vs. individual outliers. To compute the contrast of a subspace, we adopted the MCED dependency estimator similar to our previous work. As well, we also adopted the same search strategy; the Adapted binary GSA previously described (Souiden et al., 2022a). Further details on contrast calculation and the adapted binary GSA can be found in the background section as well as in our previous work.

Once the relevant subspaces have been found, the data points undergo scoring via an outlier detection algorithm within each of the identified subspaces. Thereafter, the individual scores assigned to each point $sc(X_n)_{rs}$ across various subspaces $rs \subseteq RS$ are computed and combined to generate a unified score $sc(\text{Xn})$ as mentioned in the Eq. 1.

$$sc(X_n) = \frac{1}{RS} \sum_{rs \subseteq |RS|} sc(X_n)_{rs} \tag{1}$$

The outlier detection decision function, denoted by f (see Eq. 2), labels a point as an outlier if it is associated outlier score sc falls below a predefined threshold; otherwise, it will be flagged as a regular data.

$$f(X_n \in \text{DS}) = \begin{cases} outlier\ if\ sc(X_n) < \beta \\ normal\ and\ otherwise \end{cases} \tag{2}$$

5 Proposed Solution

During this section, we will present the proposed approach and elucidate the improvements we brought to our previous work.

5.1 Solution Description

ABGSSA (Adapted Binary Gravitational Subspace Search Algorithm) represents an improvement of our previously proposed approach, BGSSA (Binary Gravitational Subspace Search Algorithm). The improvement consists of the application of a more efficient data stream handling strategy, manifested in the integration of a concept drift detector. Instead of performing the subspace search in every window, the concept drift detector permits elaborating this action only when needed, i.e., when a change takes place in the data distribution. This results in a major reduction in computational cost due to fewer subspace search operations.

Figure 1 describes the architecture of the proposed approach. It consists of three logically distinct modules:

- **The data stream handling module,** where the data stream will be processed and checked for concept drift,
- **The subspace search model,** where the subspace search method can be used to find relevant subspaces for outlier detection,
- **The outlier detection model,** where an outlier detection algorithm can be applied.

Our modification resides in the data stream handling module. The rest of the modules are the same as the BGSSA. The different modules will be detailed in further sections, with more concentration on the data stream handling module, where our measure contribution resides.

5.2 Data Stream Handling Module

Due to the data stream constrained characteristics concerning the resource requirements and the evolving distribution nature, an efficient data stream handling scheme is required. In the previous work, we relied solely on a non-overlapping sliding window model, because of its generality and efficiency. The sliding window can be useful for dealing with data stream constraints. Yet, it requires periodic model updates. In the current work, we integrated a concept drift detector to avoid the useless updates. In this context, we divided the stream handling module into two sub-modules: Data stream processing and concept drift detection.

The Stream Processing Module

We used the non-overlapping sliding window described in Sect. 2. The algorithm consistently operates within the most recent window. At the initial stage, the algorithm creates the first window, which is used for model initialization. Subsequently, it processes the incoming instances. Different from BGSSA, in ABGSSA, the instances are examined for concept drift, and the model is updated only upon the detection of a concept drift. However, as soon as the window becomes full, outlier detection is performed. The process is finished either by the end of the stream or by any other stopping criteria.

The Concept Drift-Detection Module

We have introduced a modified version of the NDKSWIN concept drift detector to handle the data stream's non-stationarity. The fundamental idea behind concept drift detectors is to monitor the stream and trigger model updates when changes in the data distribution occur. Several drift detectors exist like ADWIN (ADaptive WINdowing) (Bifet & Gavalda,

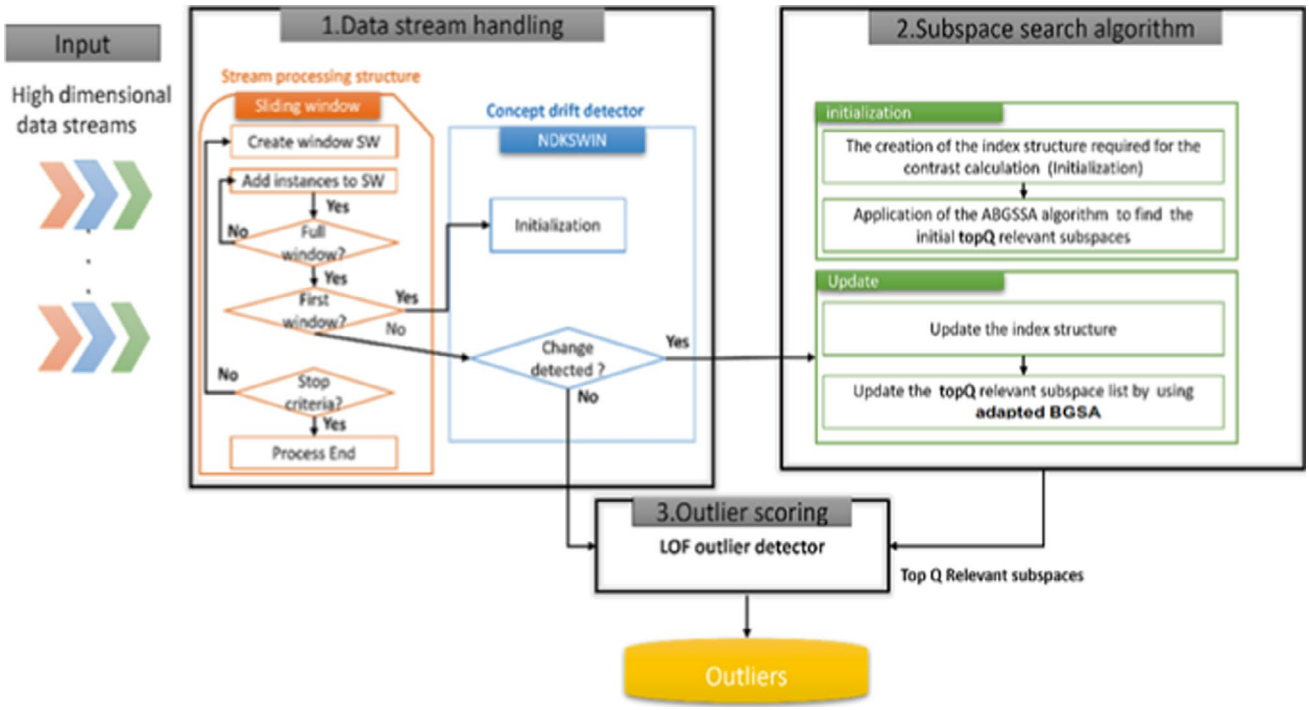


Fig. 1 Solution architecture. *Source* The author

2009), KSWIN (Kolmogorov–Smirnov Windowing) (Raab et al., 2020), and more. A more recent addition is the N-Dimensional KSWIN (NDKSWIN) introduced by Tobge in (Togbe, et al., 2021). NDKSWIN declares a drift when a change is observed in at least one of the n dimensions using KSWIN, which accepts a one-dimensional array as input and hence it gives the possibility to deal with multidimensional data streams. KSWIN (Raab et al., 2020) is a concept change detector that uses the Kolmogorov–Smirnov (KS) statistical test. It maintains a sliding window SW of fixed size ws. The most recent r samples of SW are assumed to represent the last concept R. From the first $n-r$ samples of SW, r samples are uniformly drawn, representing an approximated last concept W. The KS-test is then applied to windows R and W of the same size. It compares the distance of empirical cumulative data distributions dist (R, W). If this distance is significant, i.e., the difference between R and W is large, a concept drift is detected.

NDKWIN chooses the dimensions to verify concept drift randomly from the full set of features F. In our context, we performed a modification at this level: We verify the changes in a selected set of features denoted SD. It represents the union of features composing the obtained set of relevant subspaces RS.

When a drift is detected, a model update is triggered. The update concerns the index structure used for the contrast calculation of the set of relevant subspaces as well as the relevant subspace set.

Subspace Search and Outlier Detection Modules

Within the subspace search module, we employed the adapted binary GSA proposed and used in BGSSA. After every algorithm generation, the topQ subspaces that exhibit high fitness values (measured by the contrast measure) among all subspaces are selected as the best subspaces. However, in contrast to BGSSA, this process is executed exclusively upon the detection of concept drift. In this case, a new subspace set is created.

As for the outlier detection module, we employed the density based LOF algorithm (Breunig, et al., 2000). Outliers are identified within each window within the same subspace set. Nevertheless, when drift occurs, the outlier detection process will be conducted within the newly generated subspaces. The outlier score for each data point is obtained according to Eqs. 1 and 2.

5.3 Solution Overview

The purpose of this section is to provide a comprehensive overview of our approach by presenting the algorithm's pseudo-code as illustrated in Fig. 2. The algorithm starts by receiving a data stream DS comprising n-d-dimensional features. Additionally, it takes the required global parameters essential for the various components. Q represents the number of top subspaces and ws depicts the window size. As for binarygsa_params, contrast_params, and

Algorithm 1 ABGSSA

1: **Data**: DS: data stream, Q: top subspaces, ws: Window size, binarygsa_params: Binary GSA related parameters, contrast_params: Contrast calculation related parametres, ndk_params: NDKWIN parameters, LOF_params
2: **Result**: set of outliers and topQ subspaces
3: **Begin**
4: Subspaces ← ∅
5: Outliers← ∅
6: i←1
7: **while** (DS.hasData()) **do**
8: $W_i \leftarrow createwindow(ws)$
9: **if** (i=1) **then**
10: $I \leftarrow CreateIndexstructure(W_1)$
11: $Init\_NDKDWIN(ndk\_params, W_1)$
12: $Subspaces_1 \leftarrow AdapatedBGSA(W_1, binarygsa\_param, contrast\_params)$
13: **else**
14: **if** NDKDWIN.detectchange() **then**
15: $I \leftarrow UpdateIndexstructure(W_i)$
16: $Subspaces_i \leftarrow AdapatedBGSA(W_i, binarygsa\_param, contrast\_params)$
17: **end if**
18: **end if**
19: $Outliers_i \leftarrow OutlierDetection(Subspaces_i,)$
20: $Clearwindow(W_i)$
21: i←i+1
22: **end while**
23: **END**

Fig. 2 Pseudocode of the proposed solution. *Source* The author

LOF_params, they indicate the Binary GSA, the contrast calculation, and outlier detection related parameters.

At the start of the algorithm, the sets called "Subspaces" and "outliers" have no element. "Subspaces" is destined to hold subspaces employed for outlier detection, while "outliers" serves as a repository for the detected outliers.

During the initial phase, the algorithm initialization takes place, involving the creation and processing of the initial window W1, the creation of the index structure required for the contrast calculation (Line 10), the initialization of the concept drift detector using the data from the first window (Line 11), and the application of the adapted binary GSA for the first time to identify the topQ subspaces (Line 12). These identified subspaces are then utilized for outlier detection (Line 19). Subsequently, following the completion of processing the current window, the window is cleared (Line 20) and a new sliding window is generated and handled. As new data elements are provided to the window, the concept drift detector will check for any modification (Line 14). When a change is detected, a model update is triggered to handle potential changes. The model update encompasses two crucial steps; updating the index structure based on new incoming inputs and conducting the subspace search. These steps are iterated for each subsequent data window until the end of the stream. Notably, the integration of the concept drift detector allows a reduction in the frequency of subspace searches and hence reduces the computational cost.

6 Experimentation and Result Analysis

6.1 Experimentation Setup

During this section, we delve into the assessment of how well the proposed method's efficiency and effectiveness by conducting experiments on both synthetic and real-world datasets. We also compared it with our previous work (Souiden et al., 2022a) to showcase the improvements we have made. As well, we tested it against LOF-stream adaptation of LOF (Breunig, et al., 2000) to the stream setting, SGRMD (Fouché et al., 2021), and xStream (Manzoor et al., 2018) algorithms.

SGMRD is closely related to our work. It focuses on identifying an optimal subspace set comprising the best subspace for each dimension. The subspaces relevance is based on the discrepancy between the marginal and conditional distribution of each dimension in comparison to other features within the subspace. On the other hand, our approach considers the

Table 1 Datasets used

Dataset	Dimension	Instances	Outlier percentage (%)
Real datasets			
Kddcup99	38	25,000	7,12
Activity	51	22,253	10
Synthetic datasets			
Synth10	10	10,000	0,86
Synth20	20	10,000	0,88
Synth50	50	10,000	0,81

relevance of the subspace by calculating the average deviation of random features within the subspace. Both approaches employ statistical tests to quantify this deviation. During this experimentation, the Kolmogorov–Smirnov test (Siegel & Castellan Jr, 1988) is employed.

SGRMD utilizes diverse monitoring strategies to maintain the optimal subspaces during each window slide. Among these, the authors recommend the MPT strategy since it gave the best performance (Fouché et al., 2021).

To evaluate the effectiveness, we relied on AUC (Area Under ROC Curve) together with the average precision (AP) metrics. The former presents how well outliers are ranked in relation to inliers (Fkih et al., 2012), while the latter evaluates the algorithm's ability to correctly identify outliers. For efficiency assessment, we considered the processing time, measured in seconds. For SGRMRD and ABGSSA, we further compared monitoring time and subspace search time. To ensure an efficient processing time evaluation, we implemented the proposed algorithms and SGMRD, xStream, and LOF in Python. For LOF and xStream, we used the implementations provided by the PySAD framework (Yilmaz and Kozat, 2020). The conducted experiments were carried out on a laptop Intel(R) Core (TM) i7-7700HQ CPU @ 2.80 GHz, 8 GB RAM, Windows 10 (Professional).

6.2 Datasets Used

Table 1 depicts the datasets we relied on in our study. The real-world datasets KDDcup99[1] and Activity[2] are two frequently used real world datasets that come with ground truth information about outliers. We performed pre-processing and selection to these datasets as in (Sathe & Aggarwal, 2018) and (Fouché et al., 2021). As for the Synthetic data, we chose to use the three benchmark datasets generated by (Fouché et al. 2021): Synth10, Synth20, and Synth50. They compromiser 10, 20, and 50 dimensions, respectively.

6.3 Parameter Settings

During our experimentation, we employed the parameter setting suggested d by the authors of the competing approaches. However, for approaches involving LOF, namely, ABGSSA, BGSSA, LOF-stream, and SGRMD, we conducted calculations with different values of K parameter, including {1, 5, 10, 20, 50, 100}. Subsequently, we reported the most effective result. We calculated the average of results obtained from 7 distinct runs for each experiment. Regarding our own approaches ABGSSA and BGSSA, we set the size of the window to 1000 and the number of BGSA iterations to 10. For the parameters ps (population size) and topQ subspaces, we set them to d/2, since this configuration yields to the lowest processing time and an acceptable AUC. Furthermore, we employed the recommended setting for the dependency estimator and NDKSWIN concept drift detector.

6.4 Results and Analysis

In this section, we will introduce the results of our experiments, focusing on both efficiency and effectiveness.

Efficiency Results

The objective of this experiment is to evaluate the processing time of ABGSSA and compare it to various competitors. The number of dimensions greatly affects the performance; the processing time grows as the dimensionality increases for all the approaches, as illustrated in Fig. 3. However, it is clear that the random-based and full space-based approaches are faster because they do not require an additional subspace search operation, like ABGSSA, BGSSA, and SGRMD. Actually, the increase in the number of attributes results in a bigger search space and a larger data size, therefore, approaches based on subspace search require more time.

Now, let's consider the approaches based on subspace search. As illustrated in Fig. 3, ABGSSA outperforms both our previous work and SGRMD, which validates our proposed improvement. It's worth noting that BGSSA does not employ any monitoring strategy to track the subspace relevance; it only performs the subspace search operation for every window. In contrast, in ABGSSA, we relied on a concept drift detector, and we only applied the subspace search operation when a concept drift is detected in at least one of the features composing the subspaces. This strategy

[1] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[2] A https://archive.ics.uci.edu/ml/datasets/PAMAP2+Physical+Activity+Monitoring

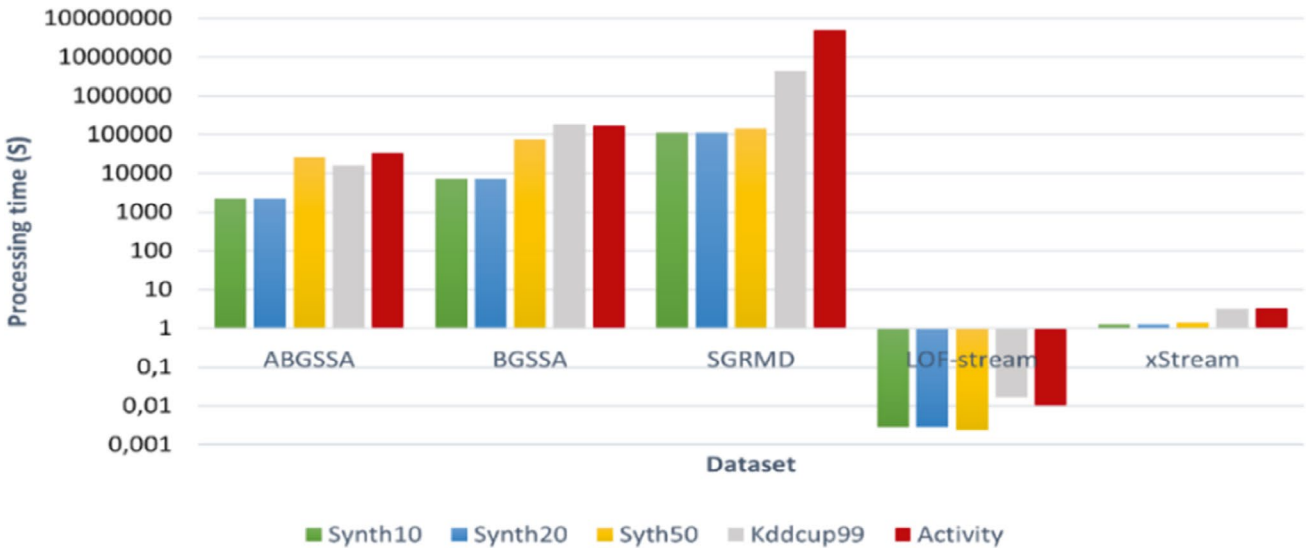


Fig. 3 Approaches processing time. *Source* The author

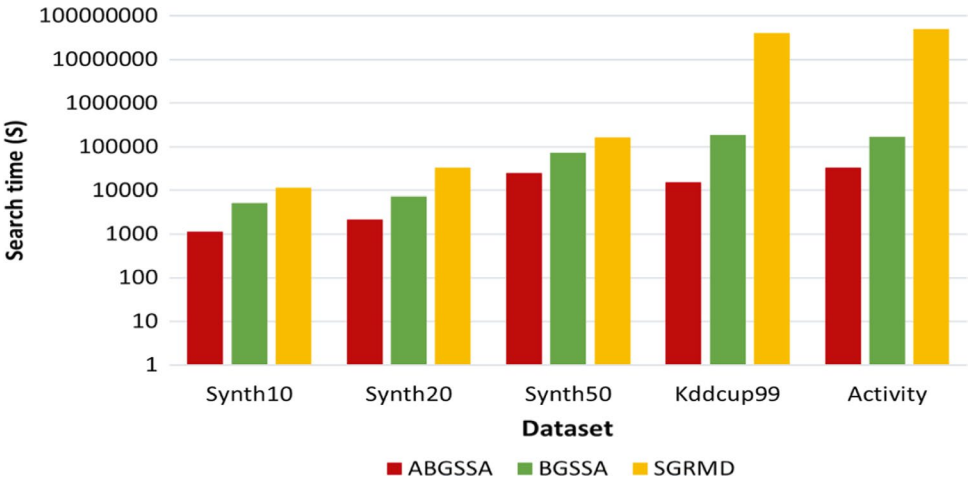


Fig. 4 Subspace search time (SGRMD versus BGSSA, versus ABGSSA). *Source* The author

eliminates the need for repetitive subspace searching operation, contributing to the observed reduction in processing time.

Practically, the indexing time (time for index structure creation and update for contrast calculation), the subspace search time (time for the subspace search operation), the monitoring time (time for monitoring data changes), and the scoring time (time for the outlier scoring) make up our approach's processing time. Yet, BGSSA does not have a monitoring time, and SGRMD adds a decision time. Generally, all approaches share the same strategy for indexing. As for the outlier scoring, we also employed the same LOF algorithm. ABGSSA and BGSSA require lower scoring time compared to SGMRD since the outlier detection is performed on d/2 subspaces. However, in SGRMD, outliers are found in d subspaces. This results in additional time compared to our approach.

In fact, the subspace searching strategy significantly contributes to the overall processing time of the algorithms. Thus, it has a substantial influence on the efficiency. Figure 4 presents the subspace search time for all examined approaches. As we can see, the strategy adopted in BGSSA and ABGSSA is faster than the one used in SGRMD, even though it conducts the search solely when the window slides across specific features. Also, we noticed that even though we used the same search strategy in BGSSA, ABGSSA required less time, since it required fewer search operations due to its concept drift-detection mechanism.

As for the monitoring time, we can notice from Fig. 5 that we have a faster monitoring strategy compared to SGMRD.

In our monitoring strategy, we modified the NDKSWIN concept drift detector by verifying the existence of the concept drift from a selected set of features instead of the full

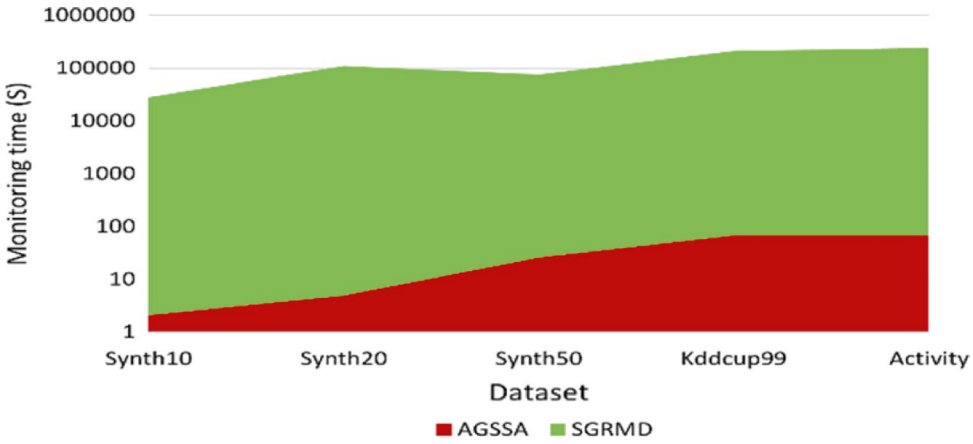


Fig. 5 Monitoring time (SGRMD versus ABGSSA). *Source* The author

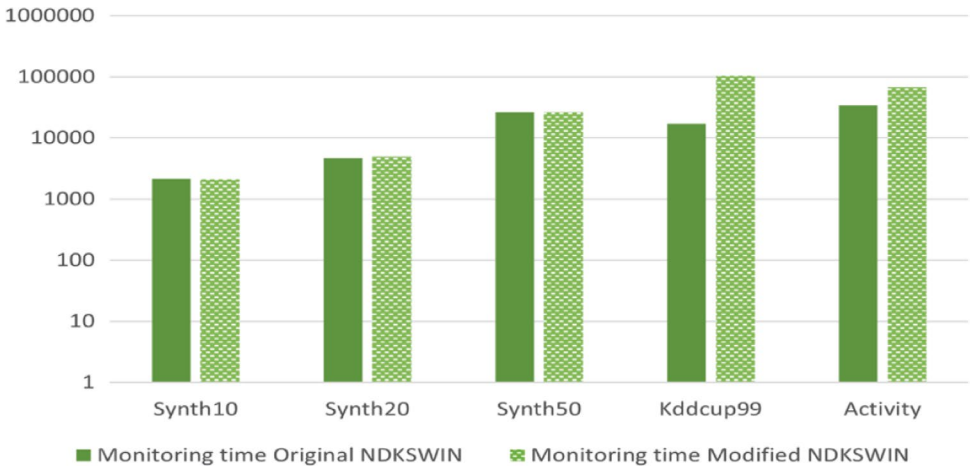


Fig. 6 Monitoring time with the original and modified NDKSWIN. *Source* The author

feature space. Our modification gave lower monitoring times compared to the original NDKSWIN, as shown in Fig. 6. Practically, usually, the number of features will be lower than the original feature space, which can explain the time reduction.

Effectiveness Results

During this section, our focus is on presenting the effectiveness results of our approach in comparison to those of our competitors and our previous work. As well, we will introduce the impact of our modification on the NKDSWIN concept drift detector. As shown in Fig. 7, with or without our modification of NDKSWIN, our approach has good performance for the majority of the datasets. Notably, our modification has brought further improvements compared to the original version.

Figure 8 summarizes the AUC and AP across the different datasets. As expected, deterministic subspace-based approaches exhibit superior performance, followed by the LOF stream. Conversely, random-based methods have the worst performance due to their prioritization of speed over accuracy. Notably, our approach delivers improvement compared to the previously proposed approach, BGSSA, while also showing comparable results with SGRMD. We outperformed SGMRD for the lowest dimensionality data (Synth10) as well as Activity dataset, while SGMRD exhibits a slight superiority over our approach across the rest of datasets with an increase ranging between 0.01 and 0.02% for the AUC measure. The algorithm's capability to correctly identify outliers decreases as dimensionality increases, and this is a result of the curse of dimensionality observed across the synthetic datasets. However, our approach excels with Activity dataset, and demonstrates a relatively low performance with the KDDcup99 dataset.

Regarding AP, our algorithm shows strong performance across the used datasets, with values ranging between 0.92 and 0.99. This signifies that our algorithm has the capability to accurately identify outliers.

Based on the obtained results, we can conclude that our approach demonstrates a satisfactory performance in the

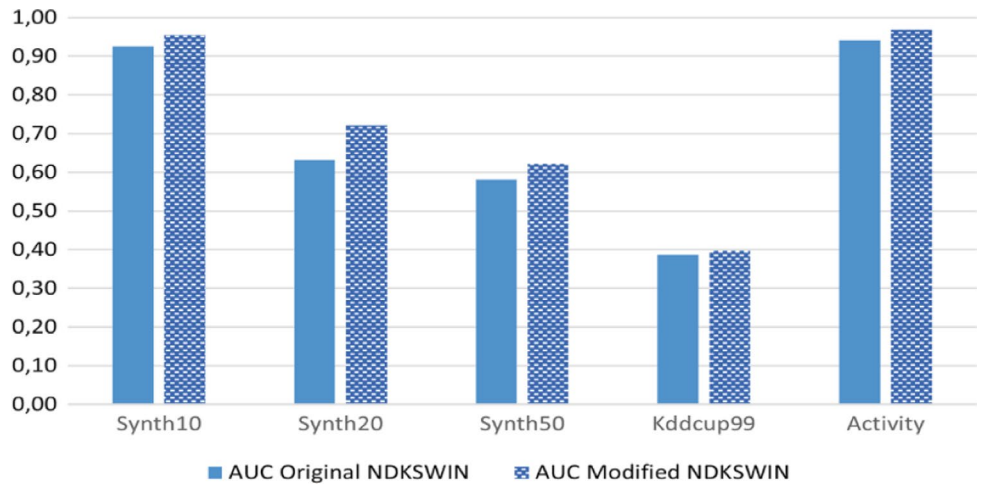
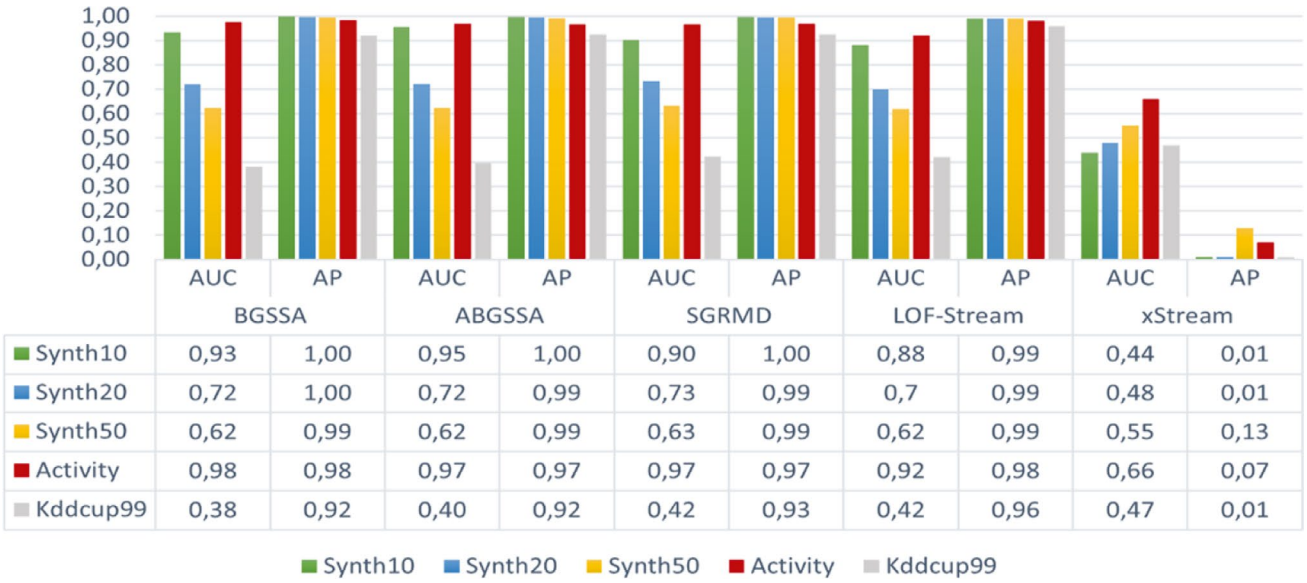


Fig. 7 AUC with the original and modified NDKSWIN. *Source* The author



	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
	BGSSA		ABGSSA		SGRMD		LOF-Stream		xStream	
Synth10	0,93	1,00	0,95	1,00	0,90	1,00	0,88	0,99	0,44	0,01
Synth20	0,72	1,00	0,72	0,99	0,73	0,99	0,7	0,99	0,48	0,01
Synth50	0,62	0,99	0,62	0,99	0,63	0,99	0,62	0,99	0,55	0,13
Activity	0,98	0,98	0,97	0,97	0,97	0,97	0,92	0,98	0,66	0,07
Kddcup99	0,38	0,92	0,40	0,92	0,42	0,93	0,42	0,96	0,47	0,01

Fig. 8 Effectiveness results. *Source* The author

level of efficiency and effectiveness. The use of a concept detector permits improving the processing time while maintaining accuracy compared to our previous work BGSSA. Nevertheless, it requires critical parameter tuning, especially the population size, the number of parameters, and topQ subspaces, which have a significant impact on the performance. Furthermore, there is a need for more exploration of our algorithm's capability to adapt to various types of concept drift, e.g., sudden, gradual, etc. In summary, ABGSSA showcased superior performance compared to BGSSA and the closely related SGMRD approach, as it is faster and gives comparable results in terms of effectiveness.

7 Conclusion and Future Work

High-dimensional data streams are becoming ubiquitous nowadays. Detecting abnormal patterns from these data is crucial in different domains. For that reason, there is an excessive need to develop tools capable of performing this task while providing a compromise between efficiency and effectiveness. In this paper, we have proposed a novel approach for outlier detection within relevant subspaces. This approach builds upon the previously proposed approach; BGSSA. In both approaches, relevant subspaces are derived through an adapted binary gravitational search algorithm. However, ABGSSA further enhances BGSSA by

integrating a modified version of the NDKSWIN concept drift detector. The modification we brought to NDKSWIN serves to reduce the amount of time it takes for processing compared to the original NDKSWIN. Furthermore, the incorporation of the modified NDKSWIN enabled the improvement of BGSSA's performance, especially in terms of processing time.

Our approach upholds the notion of generality in the subspace definition while coupling the subspace search and outlier detection. Notably, outlier detection is carried out in every window, whereas the subspace search is only conducted when a concept drift is detected. We conducted a comparative analysis against BGSSA, SGRMD, xStream, and LOF-stream in terms of efficiency and effectiveness. The results indicated that our approach performs well. It has superior performance than BGSSA and outperforms the closely related work SGMRD, as it is faster by 4 orders of magnitude and delivers comparable results in terms of effectiveness. Additionally, it exhibits superior effectiveness compared to xStream and LOF-stream. However, even though we have improved the processing time compared to BGSSA, it remains an ongoing concern. In our future work, we plan to enhance the current approach by incorporating alternate stream processing structures and employing different statistical tests, as well as providing a better concept drift handling technique. Additionally, we will improve the experimentation by incorporating more datasets and more competitors.

References

Almusallam NY, Tari ZA, Bertok P, Zomaya AY (2017) Dimensionality reduction for intrusion detection systems in multi-data streams—a review and proposal of unsupervised feature selection scheme. Emergent Computation: a Festschrift for Selim G. Akl, pp 467−487

Bayram F, Ahmed BS, Kassler A (2022) From concept drift to model degradation: an overview on performance-aware drift detectors. Knowl-Based Syst 245:108632. https://doi.org/10.1016/j.knosys.2022.108632

Bhatia S, Jain A, Li P, Kumar R, Hooi B (2021) MStream: fast anomaly detection in multi-aspect streams. In: Proceedings of the web conference 2021. pp 3371–3382

Bhatia S, Jain A, Srivastava S, Kawaguchi K, Hooi B (2022) Memstream: memory-based streaming anomaly detection. In: Proceedings of the ACM web conference 2022. pp 610−621

Bhushan A, Sharker MH, Karimi HA (2015) Incremental principal component analysis based outlier detection methods for spatiotemporal data streams. ISPRS Ann Photogramm, Remote Sens Spat Inf Sci 2:67–71

Bifet A, Gavalda R (2009) Adaptive learning from evolving data streams. In: Advances in intelligent data analysis VIII: 8th international symposium on intelligent data analysis. pp 249−260

Breunig MM, Kriegel HP, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data, pp 93–104

Cai S, Li S, Yuan G, Hao S, Sun R (2020) MiFI-outlier: minimal infrequent itemset-based outlier detection approach on uncertain data stream. Knowl-Based Syst 191:105268. https://doi.org/10.1016/j.knosys.2019.105268

Chen L, Wang W, Yang Y (2021) CELOF: effective and fast memory efficient local outlier detection in high-dimensional data streams. Appl Soft Comput 102:107079. https://doi.org/10.1016/j.asoc.2021.107079

Degirmenci A, Karal O (2022) Efficient density and cluster based incremental outlier detection in data streams. Inf Sci 607:901–920. https://doi.org/10.1016/j.ins.2022.06.013

Dehghani A, Sarbishei O, Glatard T, Shihab E (2019) A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors. Sensors 5026

Fkih F, Omri MN, others (2012) Learning the size of the sliding window for the collocations extraction: a ROC-based approach. In: Proceedings of the 2012 international conference on artificial intelligence (ICAI'12), pp 1071–1077

Fouché E, Kalinke F, Böhm K (2021) Efficient subspace search in data streams. Inf Syst 97:101705. https://doi.org/10.1016/j.is.2020.101705

Fouché E, Böhm K (2019) Monte Carlo dependency estimation. In: Proceedings of the 31st international conference on scientific and statistical database management, pp 13–24

Fouché E, Komiyama J, Böhm K (2019) Scaling multi-armed bandit algorithms. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp 1449–1459

Khalique V, Kitagawa H (2021) VOA*: fast angle-based outlier detection over high-dimensional data streams. In: Pacific-Asia conference on knowledge discovery and data mining. pp 40−52

Lazarevic A, Kumar V (2005) Feature bagging for outlier detection. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery in data mining. pp 157–166

Manzoor E, Lamba H, Akoglu L (2018) Xstream: outlier detection in feature-evolving data streams. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery and data mining. pp 1963−1972

Nguyen H-L, Woon Y-K, Ng W-K (2015) A survey on data stream clustering and classification. Knowl Inf Syst 45:535–569. https://doi.org/10.1007/s10115-014-0808-1

Pevnỳ T (2016) Loda: lightweight on-line detector of anomalies. Mach Learn 102:275–304. https://doi.org/10.1007/s10994-015-5521-0

Raab C, Heusinger M, Schleif F-M (2020) Reactive soft prototype computing for concept drift streams. Neurocomputing 416:340–351. https://doi.org/10.1016/j.neucom.2019.11.111

Rashedi E, Nezamabadi-Pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inf Sci 179(13):2232–2248. https://doi.org/10.1016/j.ins.2009.03.004

Rashedi E, Nezamabadi-Pour H, Saryazdi S (2010) BGSA: binary gravitational search algorithm. Nat Comput 9:727−745. https://doi.org/10.1007/s11047-009-9175-3

Riess M (2022) Automating model management: a survey on metaheuristics for concept-drift adaptation. J Data, Inf Manag 4:211–229. https://doi.org/10.1007/s42488-022-00075-5

Salehi M, Rashidi L (2018) A survey on anomaly detection in evolving data. ACM SIGKDD Explorations Newsl 20(1):13–23. https://doi.org/10.1145/3229329.3229332

Sathe S, Aggarwal CC (2016) Subspace outlier detection in linear time with randomized hashing. In: 2016 IEEE 16th international conference on data mining. pp 459–468

Sathe S, Aggarwal CC (2018) Subspace histograms for outlier detection in linear time. Knowl Inf Syst 1–25

Siegel S, Jr C, John N (1988) Nonparametric statistics for the behavioral sciences, 2nd edn. Nonparametric statistics for the behavioral sciences, Second, McGraw-Hill, New York

Souiden I, Omri MN, Brahmi Z (2022b) A survey of outlier detection in high dimensional data streams. Comput Sci Rev 44:100463. https://doi.org/10.1016/j.cosrev.2022.100463

Souiden I, Brahmi Z, Omri MN (2022a) Binary gravitational subspace search for outlier detection in high dimensional data streams. In: International conference on advanced data mining and applications. pp 157–169

Togbe MU, Chabchoub Y, Boly A, Barry M, Chiky R, Bahri M (2021) Anomalies detection using isolation in concept-drifting data streams. (MDPI, Ed.) Computers 10(1):13. https://doi.org/10.3390/computers10010013

Vanea A, Emmanuel M, Keller F, Klemens B (2012) Instant selection of high contrast projections in multi-dimensional data streams. In: Proceedings of the workshop on instant interactive data mining (IID 2012) in conjunction with ECML PKDD

Xie KA, Cao J, Xie G, Wen J, Zhang D, Qin Z (2018) On-line anomaly detection with high accuracy. IEEE/ACM Trans Netw 26(3), 1222−1235. https://doi.org/10.1109/TNET.2018.2819507

Yang D, Wang Y, Li Y, Ma X (2016) A variable Markovian based outlier detection method for multi-dimensional sequence over data stream. In: 2016 17th international conference on parallel and distributed computing, applications and technologies. pp 183–188

Yilmaz SF, Kozat SS (2020) PySAD: a streaming anomaly detection framework in python. arXiv preprint arXiv:2009.02572

Zhang J, Li H, Gao Q, Wang H, Luo Y (2015) Detecting anomalies from big network traffic data using an adaptive detection approach. Inf Sci 318:91–110. https://doi.org/10.1016/j.ins.2014.07.044

Zhang L, Lin J, Karim R (2017) Sliding window-based fault detection from high-dimensional data streams. IEEE Trans Syst, Man, Cybern: Syst 47(2):289–303. https://doi.org/10.1109/TSMC.2016.2585566

Zhang J, Gao Q, Wang H (2008) SPOT: a system for detecting projected outliers from high-dimensional data streams. In: International conference on database and expert systems applications. pp 1628–1631

Zhang J, Gao Q, Wang H, Liu Q, Xu K (2009) Detecting projected outliers in high-dimensional data streams. In: International conference on database and expert systems applications. pp 629–644. https://doi.org/10.1007/978-3-642-03573-9_53

Zhang S, Ursekar V, Akoglu L (2022) Sparx: distributed outlier detection at scale. In: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. pp 4530−4540

Artificial Intelligence and Crowdsourced Social Media Data for Biodiversity Monitoring and Conservation

Nathan Fox, Enrico Di Minin, Neil Carter, Sabina Tomkins, and Derek Van Berkel

Abstract

Environmental resilience is intrinsically tied to the conservation and promotion of biodiversity at multiple scales, spanning from local ecosystems to the global biosphere. Biodiversity assumes a pivotal role in the capacity of ecosystems to endure and recuperate from diverse perturbations. Human-induced stressors are causing unprecedented losses to biodiversity. Preventing and reversing the global biodiversity crisis necessitates targeted conservation endeavors, yet monitoring efforts are expensive, and conservation resources are limited. This lack of information on biodiversity statuses and trends may obscure population declines and potential extinctions. As a result, there is a pressing need for cost-effective and scalable solutions to monitor biodiversity. Here, we carried out a systematic literature review focusing on the use of artificial intelligence (AI) methods to assess social media data for biodiversity and conservation, identifying 32 articles. Our review focused on capturing which AI approaches were used, and where relevant how studies used multiple AI methodologies for a multimodal approach. Our results highlight significant recent developments in computer vision, natural language programming, and spatial analysis, and discuss their exciting applications to big data from social media for biodiversity monitoring, which hitherto have been underexplored. Social media uniquely allows for multimodal analysis offering a rich understanding of conservation issues by combining multiple data types, such as audio, video, and text. Compared to previous ecological research harnessing AI and social media, a multimodal approach offers additional insight relevant to biodiversity monitoring, including tracking the changes in timing and distribution patterns of biodiversity events and identifying areas affected by invasive species. By harnessing the capabilities of computer vision, natural language processing, and spatial–temporal analysis, we can unlock valuable insights from social media posts and guide conservation strategies for enhancing environmental resilience in an efficient and scalable manner.

Keywords

Environmental resilience · Machine learning · Computer vision · NLP · Citizen data

1 Introduction

Human activities have greatly decreased biodiversity over the past century, degrading crucial ecological functions, diminishing vital genetic variability, and threatening socio-economically important species. This alarming decline in biodiversity has raised concerns about the cascading repercussions it may trigger, potentially leading to the degradation and collapse of ecosystems (Peterson et al. 1998; Chambers et al. 2019). Prioritizing the conservation of biodiversity has become imperative for resilience across both local and global contexts (GBF 2022). Preventing and reversing the global biodiversity crisis requires targeted

N. Fox (✉)
Michigan Institute for Data & AI in Society, University of Michigan, Ann Arbor, USA
e-mail: foxnat@umich.edu

N. Fox · N. Carter · D. Van Berkel
School for Environment and Sustainability, University of Michigan, Ann Arbor, USA

E. Di Minin
Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

S. Tomkins
School of Information, University of Michigan, Ann Arbor, USA

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_4

conservation efforts informed by large-scale monitoring of wild plant and animal species.

One of the greatest challenges facing biodiversity conservation is that large-scale monitoring efforts are expensive and resources for conservation scarce. They also tend to be concentrated in certain geographies (e.g., Global North) and focus on iconic species. Lack of information on biodiversity statuses and trends, especially from understudied regions or species, may conceal population declines and potential extinctions, and consequently affect human well-being. Therefore, cost-effective and scalable solutions to monitor biodiversity are needed.

Computer Vision (CV) and Natural Language Processing (NLP) are poised to transform how we collect and validate biodiversity sightings by providing scalable methods for analysis of wildlife images and textual content (Tuia et al., 2022). Signs of this imminent revolution can be seen in global-scale efforts to crowd-source biodiversity monitoring with popular citizen science projects 'ebird' (ebird.org) and 'iNaturalist' (inaturalist.org) that take advantage of human observational capacity and machine learning methods. Furthermore, spatial and temporal analysis of such datasets has the potential to unlock new knowledge frontiers in biodiversity spatial–temporal patterns.

Social media is increasingly popular, and millions of users post content, often daily, that can be used to monitor biodiversity and human-nature interactions (Toivonen et al., 2019). As social media sites are diverse in their usage, they have different metadata types available for access (Table 1). Spatiotemporal data (e.g., geolocated images), for example, can be used to understand species' distributions and trends, species life cycles, and ecological niche spaces over space and time (Fox et al., 2020); phenomena that are exceedingly challenging to evaluate using traditional monitoring methods. Promising research has demonstrated the capacity for the evaluation of images through CV, how information can be gleaned from text posted through NLP, and how coordinates can bring spatiotemporal behavioral insights (Toivonen et al., 2019). However, truly benefiting from this unique data source in conservation science likely requires tapping into the multiple media and data, modes of social media posting, and automated data analysis methods. Here, through a systematic literature review, we demonstrate that a multimodal approach to using the diverse metadata provided by social media posts, including images, text, and spatial–temporal information can be harnessed to complement and enhance biodiversity monitoring and conservation by providing additional in situ context (Fox et al., 2020; Toivonen et al., 2019).

2 Literature Review

Definitions of AI often shift based on context and application, ranging from basic automation tasks to advanced cognitive functions. Examples include autonomous robotics in manufacturing, recommendation engines in e-commerce, and predictive analytics in finance. In the context of social media data for biodiversity and conservation research, we define AI as any methodology emulating human-like behaviors, such as learning and reasoning. We also note that terms like machine learning, neural networks, and deep learning are frequently, though incorrectly, used interchangeably with AI. We position these terms within a hierarchical framework, where they are subsets nested within the broader domain of AI. Therefore, our broad use of AI in this article encompasses machine learning methods, neural networks, deep learning, natural language processing, and computer vision, among others.

The use of social media data for sustainability research is a nascent field that has great potential for helping to shape future conservation research, policy, and practice (Toivonen et al. 2019; Ghermandi et al. 2023). The Sustainable Development Goals (SDGs) are a set of 17 global goals set forth by the United Nations in 2015 as part of its 2030 Agenda for Sustainable Development. These goals address various global challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice. Of particular relevance to environmental resilience, is SDG 15 "Life on Land", which aims to "Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat

Table 1 Selected examples of metadata available from social media sites

Social media	Social media type	Selected available data
Flickr	Image and video sharing	Photographs, videos, text, spatial data, temporal data
Twitter	Microblogging	Photographs, videos, text, spatial data, temporal data
Instagram	Image and video sharing	Photographs, videos, text, spatial data, temporal data
YouTube	Video sharing	Videos, text, temporal data
Reddit	Social news aggregation and discussion	Photographs, videos, text, temporal data
Weibo	Microblogging	Photographs, videos, text, spatial data, temporal data

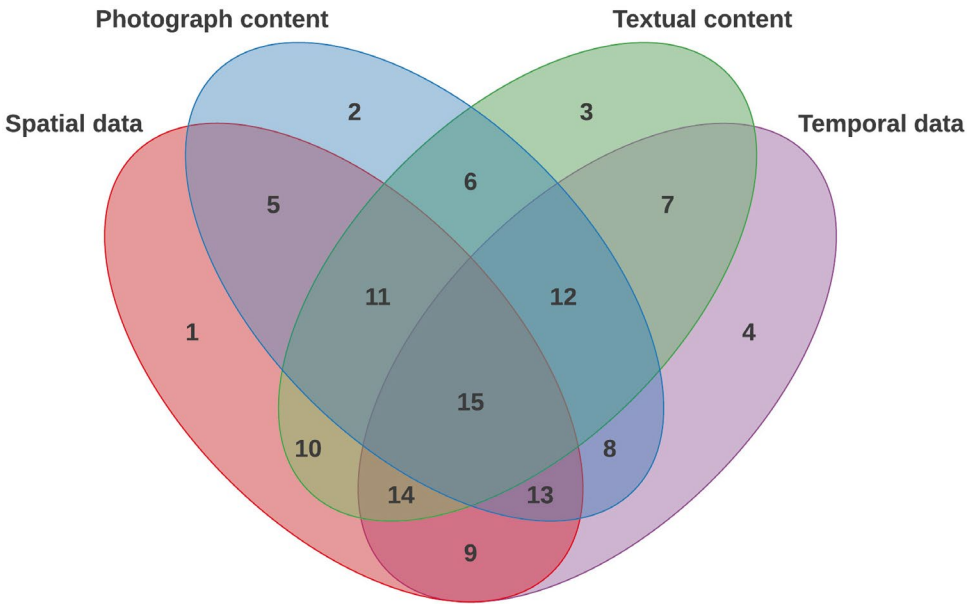


Fig. 1 Multimodal framework: a Venn diagram demonstrating the multiple ways metadata from social media data can be combined. *Source* The authors

desertification, and halt and reverse land degradation and halt biodiversity loss" (Sayer et al. 2019). Ghermandi et al. (2023) identified that social media studies have started to address this SDG, with research focusing on, loss of biodiversity (SDG 15.5), protected species trafficking (SDG 15.7), Invasive alien species (SDG 15.8), among others.

In particular, social media data has been demonstrated to be a good source of data for creating primary biodiversity datasets (Barve 2014; Fox et al. 2020). Methods for extracting primary biodiversity data from social media often follow a simplistic method of searching for a given species name and plotting the georeferenced images. Because social media data is messy, with mistagged or misidentified images, a simple search may not produce accurate biodiversity datasets (Fox et al. 2020), leading to the need for validation of species sightings. Methods used for validation often use experts to assess the contents of images and text manually. However, this is time-consuming and often impractical to apply to the large datasets available from social media images. Several studies have started to use CV methodologies to automatically identify species within social media photographs (August et al. 2020), yet implementation of such AI methodologies to automate this process is limited.

Moving beyond simple species identification and mapping, more nuanced biodiversity and conservation research methods have been developed (Toivonen et al. 2019). For example, NLP can be applied to textual data to retrieve information on threatened species from online sources (Kulkarni and Di Minin 2021) and assess public sentiment on environmental conditions (Becken et al., 2017), and CV can be used to detect the illegal trading of species and wildlife products on social media (Di Minin et al., 2018; Kulkarni and Di Minin 2023). Though advances in AI methodologies, such as CV and NLP, show great promise, their implementation into social media analysis remains scarce (Toivonen et al. 2019). For instance, a search of the Scopus database for "*(social media) AND biodiversity AND ((artificial intelligence) OR (machine learning) OR (deep learning) OR (neural network))*" returns only 11 related articles.

Social media content frequently comprises visual and textual data, necessitating multimodal analysis techniques that harness multiple types of metadata (Fig. 1). The values in Fig. 1 represent the 15 possible combinations of these most commonly accessible social media metadata types. The use of a multimodal approach can improve the performance of the analysis, for example, sentiment analysis, concurrently evaluating an image alongside its accompanying text can provide a more robust understanding of user thoughts and opinions than examining each component in isolation (Toivonen et al. 2019). This article therefore aims to review how a multimodal approach leveraging multiple AI methodologies could be used to enhance how social media data is used to address environmental resilience through biodiversity monitoring and conservation.

3 Methods

In our systematic literature review, we sought to explore the intersection of social media, biodiversity, and artificial intelligence, combining three sources of data, a literature search, and two previously curated databases on social media and biodiversity. First, we began with a focused search on Scopus using the search '*(social media) AND biodiversity*

AND ((artificial intelligence) OR (machine learning) OR (deep learning) OR (neural network))'. Second, Ghermandi et al. (2023), compiled a database of published articles using social media for environmental sustainability, categorizing each according to which SDG it addressed. From this database, we identified papers that addressed SDG 15 "Life on Land". Third, Toivonen et al. (2019) curated a database of papers that used social media data for conservation science.

To focus our review of papers using AI methodologies, the combined list of papers underwent a two-step screening: first, we reviewed each paper's titles and abstracts, and searched the full paper for any mentions of four key phrases "Artificial Intelligence", "Machine Learning", Neural Network", or "Deep Learning". Papers relevant to AI methodologies were then read in full and were further screened to only include papers that used an AI methodology themselves. While screening we only kept papers that explicitly state AI usage in their methodology. For example, if a paper stated that it used "image content analysis" with no qualifier on whether it was manually assessed or through a computer vision model, it was assumed that AI was not implemented. For each article, we assessed and classified (1) which AI methodologies were employed, (2) the media or data from the metadata of the social media used in the analysis, and (3) the biodiversity monitoring and conservation applications. For the most commonly used social media sites, we assessed the accessibility of their data for researchers.

4 Results

Our review returned 197 possible papers using AI and social media for biodiversity and conservation research (Fig. 2). With duplicates and inaccessible articles removed, our initial database for screening contained 168 articles. Of the screened articles 46 articles discussed AI. The final database contains 32 articles using AI methodologies to assess social media data for biodiversity and conservation research. We also further reviewed the 14 articles that discussed AI but did not implement it in their methodologies.

Articles using AI and social media addressed a wide variety of biodiversity-related issues including, tracking spatial and temporal distribution of biodiversity events as well as assessing public sentiment of conservation through analyzing varying textual posts from differing social media modalities (Table 2). By and large, these articles are confined to one or two of the available media or data available in the metadata for their analysis (e.g., photographs and captions). Where AI methodologies are applied, we identified three broad use cases, CV, NLP, and spatial analysis, with no article using an AI methodology to assess temporal trends. The most used AI method was CV, followed by NLP and spatial analysis (Fig. 3). Though many studies do use multimodal approaches, assessing multiple metadata types, they generally only assess one data type using AI. For example, many studies use CV or NLP to generate primary biodiversity datasets but then do not use an AI methodology to map or assess their spatiotemporal distribution and patterns, instead relying on traditional statistical modeling. Only four (12.5%) of the reviewed studies used multiple AI methods, all of which used both CV and NLP.

The studies that mention AI but do not use it, often highlight in their discussion how AI could improve the efficiency or the robustness of their research, with them unsurprisingly often opting for manual evaluation of social media posts. Of the studies reviewed, the majority use a single social media source for their analysis. Alarmingly, a number of social media platforms, prominently used for the study of conservation issues, are becoming or are now permanently unavailable to researchers without special permission (Table 3).

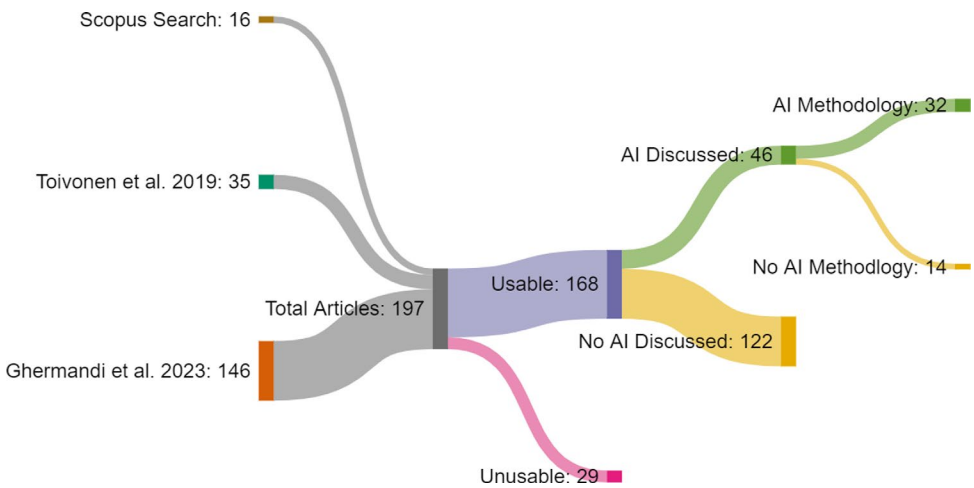


Fig. 2 Systematic literature review process. The values indicate the number of articles kept at each stage of the review. *Source* The authors

Table 2 Selected examples of social media and AI usage for biodiversity and conservation research

Study	Biodiversity application	Social media used	Multimodal famework section: data assessed with AI
Jeawak et al. (2017)	Predicting species distributions	Flickr	1; Spatial
Walden-Schreiner et al. (2018)	Protected area management	Flickr	1; Spatial
August et al. (2020)	Generating biodiversity observations	Flickr	2; Photograph content
Kulkarni and Di Minin (2023)	Identifying wildlife trade	Flickr	2; Photograph content
Feddema et al. (2021)	Identifying wildlife trade	Facebook	3; Textual
Hausmann et al. (2020)	Public sentiment towards nature	Instagram	3; Textual
Gosal and Ziv (2020)	Public sentiment towards nature	Flickr	6: Photograph content; Textual
Rocha et al. (2023)	Generating biodiversity observations	Instagram	6: Photograph content; Textual

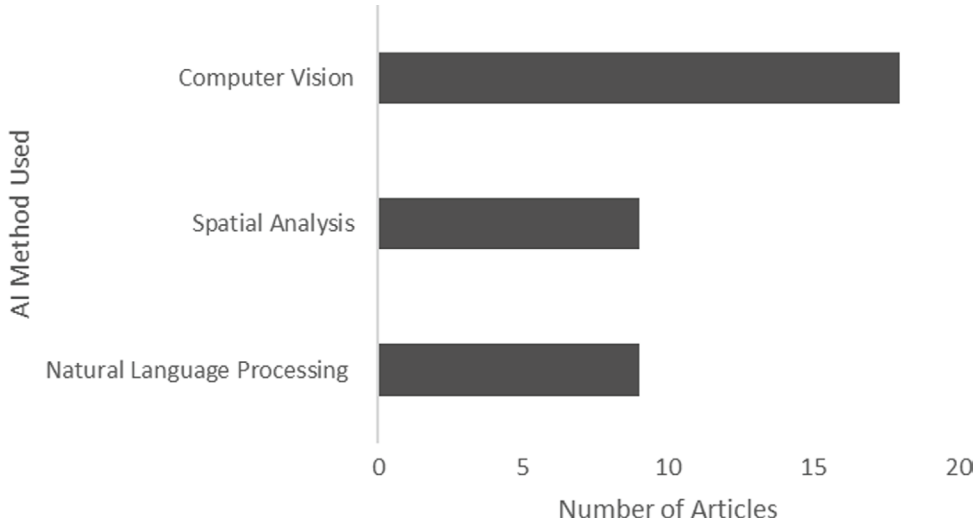


Fig. 3 Number of articles using different AI methodologies. *Source* The authors

Table 3 Accessibility of social media

Social media website	Accessibility to researchers (as of November 2023)
Flickr	Yes—accessible API supported by open-source tools in a number of programming languages
Twitter	Future unknown—currently undergoing changes to API access
Instagram	No—previous API access has been restricted
YouTube	Yes—accessible API supported by open-source tools in a number of programming languages
Reddit	Future unknown—currently undergoing changes to API access
Weibo	Yes—accessible API supported by open-source tools in a number of programming languages

5 Discussion

Mobilizing social media data and machine learning methods can help enhance existing knowledge of biodiversity with the potential to better inform conservation (e.g., the Kunming-Montreal Global Biodiversity Framework) and broader sustainability (e.g., the UN Sustainable Development Goals) decision-making. Innovative research can use these AI-validated species datasets to provide crucial information that helps guide conservation priorities across spatiotemporal scales, from local site-specific studies to identify global biodiversity priorities. Previous work, for example, has harnessed such machine learning approaches to successfully identify species across different taxa and environmental settings (August et al., 2020; Nunes et al., 2020), thereby delivering scalable insights for managers that were previously overlooked. Building upon these successes and adding multiple machine learning domains will increase the robustness of social media posts for biodiversity monitoring.

Our study demonstrates that conservation science studies using social media have yet to fully tap into the multimodal potential of social media. Limiting the study of these issues to a single social media or one media or data type may add unwanted bias to results. For example, the study of spatial and temporal trends in social media use from coordinate and timestamp alone (Wood et al. 2013), may miss important contextual information about the motivation to visit certain areas. By enriching spatial and temporal data with contextualized information, for example, extracted from photographs or textual content, we can gain more robust insights into biodiversity conservation. These scalable methods excel in such big data environments, allowing for the efficient processing of millions of images and vast text data, broadening the scope for research across varied spatial and temporal scales (Schirpke et al. 2023).

Multimodal approaches that leverage all available metadata provide innovative approaches that allow for fully realized applications in biodiversity monitoring. Still, our results show that these multimodal approaches will require flexibility that accounts for heterogeneous, often changing, social media data types (e.g., audio, video, and text on a Twitter post) (Toivonen et al., 2019). Current multimodal approaches only use a combination of images and text and do not use AI methodologies to assess the other metadata associated with social media posts. Moreover, results indicate that these multimodal approaches underutilize AI approaches. For example, though they did not use any AI methods, one of the best examples of such a multimodal approach is presented by Allain (2019). In their study, they were able to map invasive species from photograph interpretation and the coordinates of these social media images. Such studies would benefit from computer vision techniques, allowing for scalable methods that could extend these insights geographically. By utilizing AI for filtering and validating data on species sightings shared on social media platforms by users, we can incorporate these observations in larger biodiversity databases for research and conservation purposes. The use of AI labeling may also help to reduce the amount of noisy data caused by incorrectly tagged images (Fox et al. 2020).

Moving forward, multiple metadata types should be incorporated into multimodal workflows to increase the robustness of the conservation outcomes. Computer vision tools can be effectively used to investigate the contents of images such as detecting and identifying and counting individuals of a species, assessing human–nature interactions, and understanding the posture and behavior of animals (Kulkarni and Di Minin, 2023; Tuia et al., 2022). Similar methodologies can be applied to the associated spectrograms of audio recordings or the sound from videos to detect species such as songbirds (Tuia et al., 2022). These AI-identified species, or human-nature interaction, datasets from image or sound content should be combined with coordinate and timestamp information to enable spatiotemporal analysis. Though we did not review any papers using AI for temporal analysis, conservation science could harness established AI time series analysis methodologies and apply them to social media data. The implementation may enable more robust assessments of population dynamics and monitor changes over time.

Textual data associated with social media posts also provides the opportunity to assess public interest, sentiment, perceptions, attitudes, and behaviors for biodiversity, natural areas, recreational activities, or conservation efforts (Toivonen et al., 2019). Conservation efforts should be intrinsically linked to supporting the sustainable development of both people and nature. By using NLP alongside CV to gauge public opinions on multiple facets of biodiversity and conservation, we can better inform conservation choices to be beneficial not just to biodiversity, but to individuals, society, and the economy (Toivonen et al., 2019). Moreover, the combination of social media and georeferenced locations with spatial AI methods can provide powerful predictive modeling, such as predicting, species distribution, landscape esthetic qualities, and environmental variables (Jeawak et al. 2017, 2020).

Conservation science could further benefit from novel AI approaches and applying these to social media data for biodiversity monitoring and conservation. First, AI could be better leveraged to detect changes in timing or distribution patterns by tracking the frequency and location of social media posts about specific biodiversity events, such as bird migrations or flowering events (August et al., 2020). Second, by analyzing geotagged posts or user-generated content related to sightings or encounters with invasive species, AI can help rapidly identify areas where invasive species are prevalent or spreading, track their distribution, and raise awareness among the public and relevant authorities for management actions (August et al., 2020). Third, AI can add critical insights into public knowledge, awareness, and engagement with biodiversity and human-nature interactions. By identifying relevant hashtags, keywords, and user-generated content, AI can provide insights into public perceptions, misconceptions, and knowledge gaps, which can inform educational campaigns and outreach efforts to raise awareness and promote sustainable interactions with nature (Ghermandi et al., 2023). Fourth, for species that are not captured by a large number of social media posts, reinforcement learning approaches (Sutton and Barto, 2018) can nudge social media users to provide additional information on existing posts of these data-sparse species or to collect new photographs when they see them. In this way, social media data combined with machine learning can

usher in novel ways of detecting threats and changes to biodiversity, engaging the public, and supporting conservation efforts related to human-nature interactions.

For these potentials to be realized in the conservation sciences researchers newly using AI and social media should follow ethical guidelines and in full respect of data privacy and data protection regulations (Di Minin et al., 2021; Ghermandi et al., 2023). For example, sensitive information regarding the citizen who uploaded the information must be protected, ensuring that all data is anonymized and unidentifiable (Ghermandi et al., 2023). Furthermore, it is critical that work using these methodological approaches does not inadvertently contribute to biodiversity declines. For instance, providing detailed maps of vulnerable species locations may enable poachers to find them. Though we should strive for open and transparent research, this sensitive topic requires some nuanced approaches to presenting results. For example, where researchers have mapped the distribution of at-risk species, or human–nature interactions, it may be better to present their results in an aggregated manner, that while still helpful for guiding conservation strategies, does not reveal detailed information that could be used for malicious purposes.

Continued data accessibility is the biggest challenge facing the use of social media data for biodiversity monitoring and conservation (Ghermandi et al., 2023). Though some social media websites such as Flickr and YouTube maintain open and accessible APIs, other sites such as Twitter and Reddit are changing their API access which could impact the accessibility of these datasets for all researchers, regardless of the scientific domain. For example, Twitter (now X), once used to assess and communicate conservation efforts (e.g., Di Minin et al. 2018), has undergone tremendous changes after its sale to Elon Musk. Access has diminished for researchers, and the composition of users has resulted in an increasingly less constructive environment for sharing information on conservation-related topics (Ghermandi et al., 2023). This reflects historic changes to websites such as Facebook and Instagram which now restrict researchers from accessing their invaluable datasets. Researchers may have a role to play in ensuring long-term accessibility to such data, by having open conversations with social media companies, our shared commitment to sustainability may foster extensive collaboration for mutual benefit and progress (Ghermandi et al., 2023).

6 Conclusion

In contrast to traditional ecological research utilizing AI, a multimodal approach using diverse metadata from social media to biodiversity monitoring enables innovative methods for tracking changes in biodiversity and informing conservation for ecosystem resilience. Leveraging computer vision, natural language processing, and spatial–temporal analysis, this approach unlocks valuable insights from social media posts, offering an efficient and scalable means to inform conservation strategies and bolster environmental resilience. Through the integration of cutting-edge AI technologies, we can now envision a more comprehensive and data-driven approach to safeguarding our planet's biodiversity for a sustainable future. Future work should therefore leverage the full range of metadata available from social media for contextualized insights into biodiversity monitoring and conservation for building environmental resilience. We envisage harnessing a multimodal AI approach is best suited to (1) assessing complex spatiotemporal biodiversity patterns, such as invasive species, migration, and niche shifts under changing climates and (2) soliciting public opinions to better include societal values in conservation decisions.

Acknowledgements This project is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program.

References

Allain SJ (2019) Mining flickr: a method for expanding the known distribution of invasive species. Herpetol Bull 148(Summer 2019):11–4

August TA, Pescott OL, Joly A, Bonnet P (2020) AI naturalists might hold the key to unlocking biodiversity data in social media imagery. Patterns 1(7)

Barve V (2014) Discovering and developing primary biodiversity data from social networking sites: a novel approach. Eco Inform 24:194–199

Becken S, Stantic B, Chen J, Alaei AR, Connolly RM (2017) Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. J Environ Manage 203:87–97

Chambers JC, Allen CR, Cushman SA (2019) Operationalizing ecological resilience concepts for managing species and ecosystems at risk. Front Ecol Evol 7:241

Di Minin E, Fink C, Tenkanen H, Hiippala T (2018) Machine learning for tracking illegal wildlife trade on social media. Nat Ecol & Evol 2(3):406–407

Di Minin E, Fink C, Hausmann A, Kremer J, Kulkarni R (2021) How to address data privacy concerns when using social media data in conservation science. Conserv Biol 35(2):437–446

Feddema K, Harrigan P, Wang S (2021) The dark side of social media engagement: an analysis of user-generated content in online wildlife trade communities. Australas J Inf Syst 25

Fox N, August T, Mancini F, Parks KE, Eigenbrod F, Bullock JM, Graham LJ (2020) "photosearcher" package in R: an accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX 12:100624

GBF (2022) The kunming-montreal global biodiversity framework. https://www.cbd.int/doc/c/e6d3/cd1d/daf663719a03902a9b116c34/cop-15-l-25-en.pdf

Ghermandi A, Langemeyer J, Van Berkel D, Calcagni F, Depietri Y, Vigl LE, Wood SA (2023) Social media data for environmental

sustainability: a critical review of opportunities, threats, and ethical use. One Earth 6(3):236–250

Gosal AS, Ziv G (2020) Landscape aesthetics: Spatial modelling and mapping using social media images and machine learning. Ecol Ind 117:106638

Hausmann A, Toivonen T, Fink C, Heikinheimo V, Kulkarni R, Tenkanen H, Di Minin E (2020) Understanding sentiment of national park visitors from social media data. People Nat 2(3):750–760

Jeawak SS, Jones CB, Schockaert S (2020) Predicting the environment from social media: a collective classification approach. Comput Environ Urban Syst 82:101487

Jeawak S, Jones CB, Schockaert S (2017) Using Flickr for characterizing the environment: an exploratory analysis

Kulkarni R, Di Minin E (2021) Automated retrieval of information on threatened species from online sources using machine learning. Methods Ecol Evol 12(7):1226–1239

Kulkarni R, Di Minin E (2023) Towards automatic detection of wildlife trade using machine vision models. Biol Cons 279:109924

Nunes JAC, Cruz IC, Nunes A, Pinheiro HT (2020) Speeding up coral reef conservation with AI-aided automated image analysis. Nature Machine Intelligence 2(6):292–292

Peterson G, Allen CR, Holling CS (1998) Ecological resilience, biodiversity, and scale. Ecosystems 1:6–18

Rocha HF, Nascimento LS, Camargo L, Noernberg M, Hara CS (2023) Labeling portuguese man-of-war posts collected from instagram. In: European conference on advances in databases and information systems. Springer Nature Switzerland, Cham, pp 369–381

Sayer J, Sheil D, Galloway G, Riggs RA, Mewett G, MacDicken KG, Edwards DP (2019) SDG 15 Life on land–the central role of forests in sustainable development. In: Sustainable development goals: their impacts on forest and people. Cambridge University Press, pp 482–509

Schirpke U, Ghermandi A, Sinclair M, Van Berkel D, Fox N, Vargas L, Willemen L (2023) Emerging technologies for assessing ecosystem services: a synthesis of opportunities and challenges. Ecosyst Serv 63:101558

Sutton RS, Barto AG (2018) Reinforcement learning: an introduction. MIT press

Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, Di Minin E (2019) Social media data for conservation science: a methodological overview. Biol Cons 233:298–315

Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Berger-Wolf T (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13(1):792

Walden-Schreiner C, Leung YF, Tateosian L (2018) Digital footprints: incorporating crowdsourced geographic information for protected area management. Appl Geogr 90:44–54

Wood SA, Guerry AD, Silver JM, Lacayo M (2013) Using social media to quantify nature-based tourism and recreation. Sci Rep 3(1):2976

Parallelism in Architectural Design: New Perspectives, Workflows and Tools Involving Robotics and Artificial Intelligence

Client Brief to 3D Printed Construction—An Artificial Intelligence Workflow for Architectural Design Process

Preyan Mehta

Abstract

The pressing need to provide housing for a projected global population growth mandates the construction of 327 million new independent houses over the next three decades. Achieving this formidable task entails either a tenfold augmentation in the population of practicing architects or a tenfold acceleration of design and construction processes. Notably, the latter objective emerges as an attainable target when compared to the former. Nevertheless, the Architecture, Engineering, and Construction (AEC) industry grapples with an array of challenges emanating from the prevailing Architectural Design Process (ADP). These predicaments encompass substantial investments in software skill acquisition and maintenance, the absence of real-time feedback bridging the conceptual and construction phases, the dearth of real-time physics simulations, the challenges of translating architects' visions into on-site constructed projects, the absence of automation for repetitive tasks, the complexities associated with converting drawing typologies, to name a few. This research adopts a qualitative research paradigm, characterized by an exploratory investigation grounded in personal experiences, empirical observations, dialogues, and experimental data. The study introduces a Smart System poised to redefine the ADP by integrating Artificial Intelligence (AI) comprehensively. This innovative system aspires to be both foundational and open, fostering continuous development and improvement. To facilitate this vision, the realm of Building Information Modeling (BIM) and the Industry Foundation Classes (IFC) file format are extended, capitalizing on the Python programming language. The process of prototyping and experimentation unfolds within the open-source 3D software Blender, exemplifying the viability of an open system. It is crucial to underscore that the research places a distinct emphasis on the refinement of extant AI models, prioritizing this avenue over the creation of new models. This preference is underscored by the formidable time and computational resources requisite for training novel AI models, constituting a central constraint within the scope of this inquiry.

Keywords

Architectural design · System design · Smart system · Artificial intelligence · Open source · System optimization

1 Introduction

Current demographic trends and population projections indicate that the global human population is anticipated to increase to 9.7 billion by the year 2050, from its present estimate of 8.1 billion in 2023 (UNITED NATIONS DEPARTMENT FOR ECONOMIC AND SOCIAL AFFAIRS, 2023). This population growth necessitates the creation of new housing for approximately 1.6 billion individuals within the forthcoming three decades. By considering this figure alongside the average global household size, it is evident that a minimum of 327 million new independent houses must be designed and constructed to accommodate this growing populace.

When accounting for the existing number of practicing architects, the average number of projects managed by each architect per year, and the typical duration required to design and construct a new house per architect, it becomes

P. Mehta (✉)
Architecture Department, I.I.D.E.A, Indus University, Rancharda, Ahmedabad, Gujarat 382115, India
e-mail: preyanmehta.edu@gmail.com

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_5

apparent that achieving this objective necessitates either a tenfold increase in the architect population or a tenfold acceleration in the rate of house design and construction. According to calculations, architects would need to design and construct one new house per month to merely meet the housing demand, a stark contrast to the current standard of designing and constructing three new houses per year.

However, a tenfold increase in the number of practicing architects within the next decade is practically implausible due to waning interest in the field, prohibitive infrastructure costs, and sheer volume. The most feasible approach lies in the development of an efficient and more intelligent ADP. To achieve this, we must first scrutinize the shortcomings inherent in the existing ADP.

The conventional ADP consists of six primary stages as seen in Fig. 1: Project Brief, Conceptual Design, Pre-Design, Design, Design Execution, and Project Feedback. Each stage is beset by specific challenges, as outlined below:

1. Project Brief: During this stage, the client introduces the project to the architect, outlining requirements and providing inspiration, if any. Key issues include inadequate documentation, extended contract bidding times, and information loss between the client's project brief and the architect's understanding thereof (Norouzi et al., 2015).
2. Conceptual Design: Here, initial design concepts emerge through case studies, contextual analysis, and the architect's intuition. Challenges include difficulties in sourcing, structuring, and organizing case studies, time-consuming abstraction processes, and a lack of real-time feedback through simulation at this early design stage.
3. Pre-design: This stage involves further refinement of conceptual ideas, analyzing volumes based on technical requirements, climatic considerations, and human circulation patterns. It includes the conversion of rough sketches into preliminary technical drawings, such as plans, sections, and elevations. Challenges encompass the repetition of logical processes for each project, the absence of actionable guidelines for personalized spatial experiences, time-consuming sketch-to-technical drawing conversion, and a lack of real-time simulation and data processing.
4. Design: In this phase, presentation-ready technical drawings are generated, encompassing details such as fenestrations, basic interior layouts, and exploration of materials and construction techniques. These drawings serve to explain the design to clients. Issues include a lack of real-time simulation and manual calculation of budget constraints (Lawson, 2010).
5. Design Execution: Construction-ready MEP (Mechanical, Electrical, Plumbing) details are integrated, and technical drawings prepared for construction. Once finalized, these drawings are sent to conventional construction agencies for implementation. Challenges include coordination gaps between agencies, disparate file formats, limited real-time design reflections post-MEP integration, and the absence of automated tendering and Bill of Quantities (BOQ) generation.
6. Project Feedback: Post-construction, feedback from clients is solicited concerning the architect's ADP, construction and design quality, and the efficiency of translating design concepts into reality. Challenges involve the lack of post-occupancy feedback using sensors and documentation for continuous improvement.

To effectively and efficiently address the challenges inherent in the ADP, particularly in the context of accommodating the expanding global housing demand, the present study advocates the development of a smarter ADP system. The system introduces an entirely novel approach, characterized by extensive integration of AI. This innovative approach aims to systematically tackle major ADP issues, thereby enhancing overall efficiency. Prior to delving into the proposed solution, the subsequent section offers a comprehensive exposition of the research methodology adopted in this investigation.

1.1 Research Methodology

The present study introduces a novel system enriched with AI integration to enhance the existing ADP. This endeavor aligns with a qualitative research approach, concentrating on the exploratory and conceptual examination of AI integration within the ADP framework. The research methodology encompasses multifaceted data collection methods, outlined as follows:

1. Personal Experiences and Observations: Primary insights and knowledge are derived from active engagement in architectural design projects and academic pursuits in architectural design. These personal experiences provide a tangible foundation for comprehending the ADP and the intricate nuances associated with the integration of AI within this context. Concurrently, observations are conducted during the execution of architectural design tasks, even when the adoption of AI systems or tools is under consideration, even if not fully implemented. These observations afford valuable insights into the practical challenges and opportunities inherent in the architectural design realm when AI is introduced.
2. Discussions and Interviews: A comprehensive array of discussions and interviews, both structured and informal, were conducted with a diverse panel of 15 experts. This panel comprised five practicing architects, five

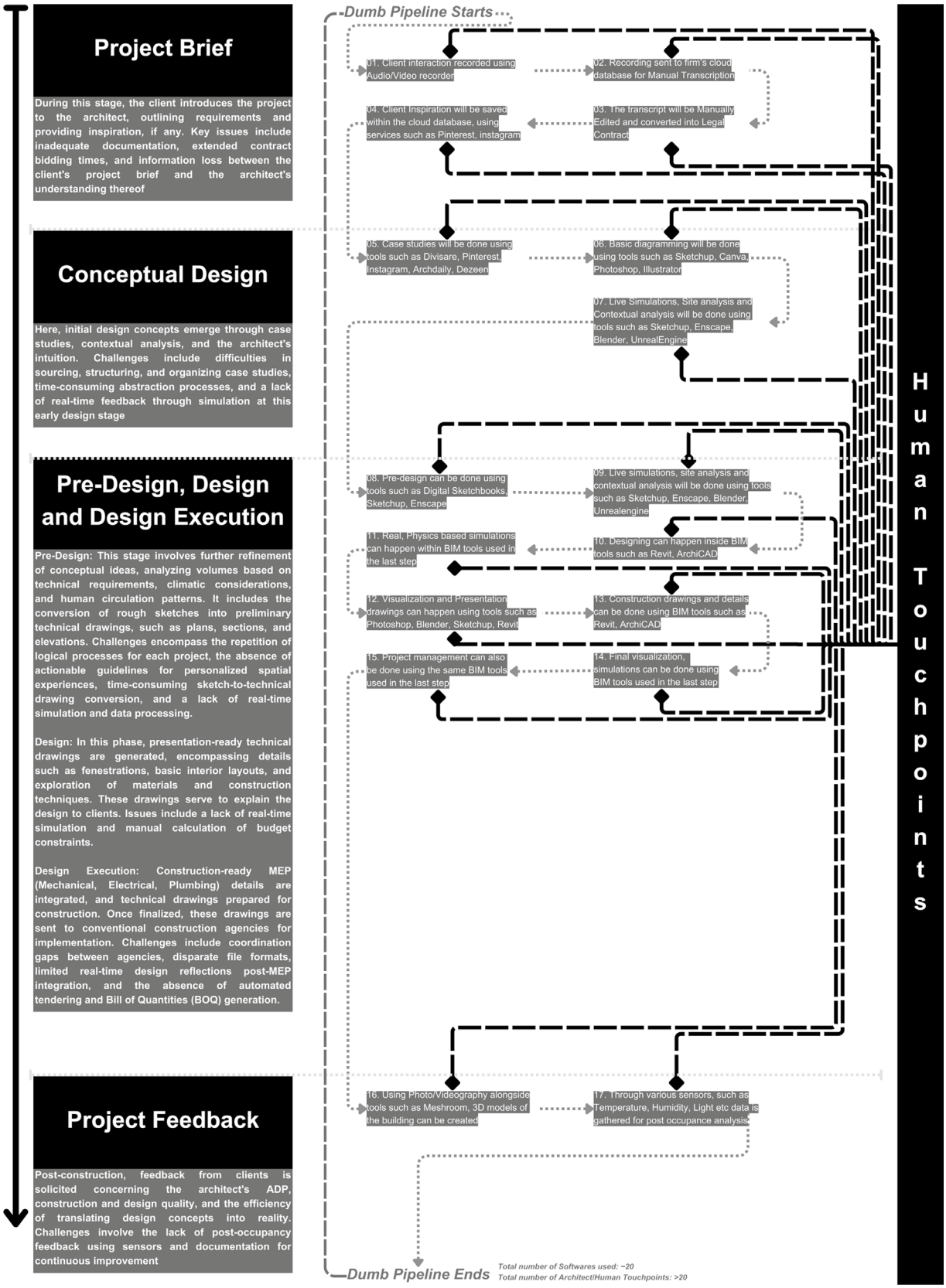


Fig. 1 Existing Dumb ADP—Six Stages, their issues and existing pipeline. *Source* Author

esteemed professors, and five industry professionals, hailing from geographically diverse regions including India, Australia, Egypt, the United States, Poland, and Germany. Six of the participants were female, while nine were male, ensuring a balanced gender representation. These interactions were designed to capture a spectrum of perspectives, expert opinions, and industry insights pertaining to the role of AI in architectural design.

3. Experimental Data: The research incorporates experimental data derived from the utilization of various existing software, tools, and AI models across various stages of the ADP. Data was meticulously observed and collected, while novel AI models were subjected to testing to assess their applicability within distinct components of the proposed system.
4. Literature: The study integrates a comprehensive literature review encompassing books, research papers, social media posts, blog entries, and video content. This extensive literature study augments the research's theoretical foundation and provides critical context.

The data acquired through the aforementioned methods was subjected to a mixed-methods analysis. This approach combines thematic understanding of the literature with hands-on experimentation involving AI models. The Python programming language served as the common analytical tool, complemented by the open-source 3D software Blender. Notably, the research conceptualizes a framework rather than a fully operational system due to the resource-intensive and time-consuming nature of training novel AI models. This limitation underscores a key facet of the research.

To address inherent limitations and potential biases, several strategic measures were implemented:

- The true intent of interviews and discussions was concealed to minimize participant bias.
- Diverse ethnic demographics were included to ensure a broader perspective.
- The term "AI" was intentionally avoided when proposing the new ADP system, as the term carries implicit bias, particularly within the industry (Lee, n.d.).
- The experiment's output was presented to a diverse group of individuals to assess its output quality, with the deliberate omission of any indication that it was generated by AI.

In summary, the research methodology employed herein amalgamates diverse data sources and analytical techniques to comprehensively explore the integration of AI within the architectural design process. This approach underscores the research's rigorous and systematic pursuit of enhancing architectural design through AI innovation, while diligently addressing potential limitations and biases.

2 Proposed ADP

Every workflow or system possesses an organizational structure, typically comprising two fundamental components: the “Shoulder” and the “Body.” To elucidate this concept, consider the analogy of an office environment, where the “Shoulder” corresponds to the C-level executives responsible for decision-making. These individuals collectively represent a cross-sectional view of the organization, encompassing its various departments, such as finance, sales, marketing, and operations. The “Shoulder” element, characterized by its horizontal nature, symbolizes the breadth of the organization. Conversely, the “Body” is embodied by the individual departments aligned vertically, signifying the depth of the organization. For any organizational system to function effectively, both breadth and depth are indispensable components (Wang & von Tunzelmann 2000).

In the context of a conventional ADP, the “Breadth” is exemplified by distinct stages, including “Brief,” “Concept,” “Design,” “Execution,” and “Feedback.” Meanwhile, the “Depth” manifests through the myriad tasks nested within each stage, encompassing activities such as brief documentation, program extraction, case studies, visualization, and BOQ preparation. This hierarchical structure also extends to the realm of AI, with different agents responsible for diverse tasks.

The AI agent entrusted with managing the “Breadth” must possess the capacity to oversee all aspects, mediate interactions, facilitate communication, and interconnect various agents seamlessly. This AI entity, referred to as “Assistant Architect AI” (AAI), harnesses Natural Language Processing (NLP) capabilities, leveraging advanced Large Language Models (LLMs) which are also multi-modal in nature (OpenAI 2023). AAI's proficiency in comprehending natural human language eliminates the need for specialized code inputs. Additionally, alongside the AAI, a Visualization Agent (VA) assumes the role of a generalist agent, consistently available to facilitate the visualization process at various stages.

In pursuit of heightened operational efficiency, all subsidiary agents, representing the “Depth” or specialized aspects of the system, engage in collaborative efforts wherever practicable. The principal verticals of this parallel process encompass the Context Class Agents (CCA), Design Class Agents (DCA), and Execution Class Agents (ECA) as can be seen in Fig. 2. CCA and DCA fall under the jurisdiction of the Architect, while ECA is affiliated with the construction agency. The CCA encompasses a range of agents, including the Building Typology Agent (BTA), Context Agent (CA), Site Agent (SA), Sensorial Experience Agent (SXA), and Case Study Agent (CSA). This class provides essential foundational information to the DCA.

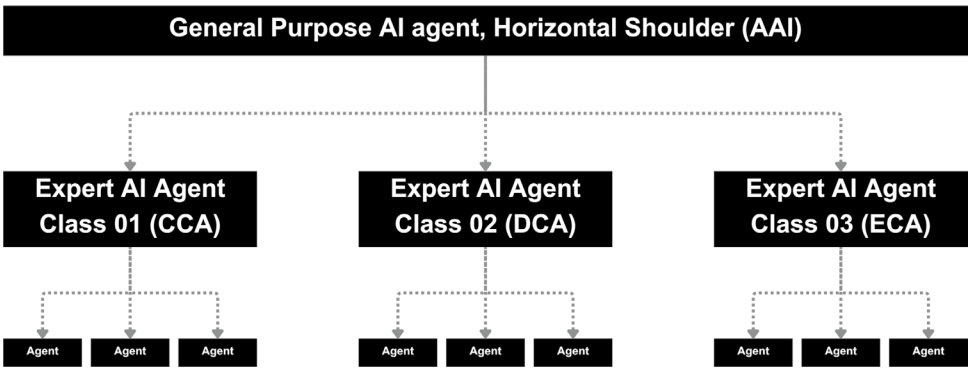


Fig. 2 Diagram representing AI agent's structure. *Source* Author

The DCA comprises agents such as the Design Agent (DA), Simulation Agent (SMA), Construction Agent (CNA), and Budget Agent (BA). This class is responsible for the design process and its iterative refinement based on real-site data. The ECA is composed of agents such as the Executor Agent (EA), Bidding Agent (BDA), Logistics Agent (LA), Worker Agent (WA), and Quality Check Agent (QAA). This class oversees the practical execution of the design on the construction site.

AAI and other specialized agents communicate via Application Programming Interface (API), a common mechanism employed in software applications for inter-communication, data sharing, and information exchange (Boateng et al. 2019). APIs comprise a set of protocols and routines facilitating communication among services, agents, or other software components in the background processes. These APIs operate both internally within the system and externally over the internet, enabling connections with various pre-built services or software solutions. The Graphical User Interface (GUI or UI), defining the user interaction with the system, assumes the form of a web application compatible with any device or operating system. The back-end operations are executed using Blender, offering architects the flexibility to access and utilize the system from any location, regardless of their possession of a personal computer.

The subsequent section elucidates the functioning of the enhanced ADP, commencing with a hypothetical conversation between an Architect and a Client.

Client: I want to design a University, which will be situated near Hogwarts School of wizardry, in Harry Potter's Universe. I have a budget of 100 million De (Demo-Currency or Currency of Democratically Smart World) (Mehta, 2019), and I want the campus to be able to host at least 900 students within the campus, at once.

Architect: Ok, and do you have a site in mind near Hogwarts, or we're supposed to choose that as a part of the process? The budget and number of users seems ok right now.

Client: Well, it is a little embarrassing that you asked. It would be nice if you could show us a few sites around, and then we can decide which site to purchase and make the university on.

Architect: No worries. Can you share a little bit about your university's vision and mission? We would like to know more about it.

Client: Sure! Our mission is to prepare our students for the next 2 decades in terms of Jobs and Entrepreneurship, through innovative, technologically enhanced teaching pedagogies as well as infrastructure. And the vision is to show the World the power of innovation as disruption.

Architect: That is amazing to hear. What types of institutes are you planning to have?

Client: for starters, we would like to have Institutes of Metaverse Architecture, eVTOL Engineering, and Parametric Marketing. But this will only be the first phase. We do want to expand later into 5 more institutes. All the institutes must be located within the same campus, in such a way that it maintains individuality, be private but also promote and support transdisciplinary research and knowledge sharing.

Architect: Understood. That will help promote your mission and vision to people. Anything else you want to share with us today, which will help us design better?

Client: Not really, but we do have some images, videos and articles for reference as to how we perceive the university to be. Would you like to see them?

Architect: Definitely! Send the resources and assets post our conversation. Last important step, before we go ahead with the project. Based on what we talked about, our experience and our brand, the estimated fees for the project will be 1% of the total budget, i.e. 1 Million De. Here is a stages wise split of our fees-

Stage 01: 10% of the total fees, i.e. 1,00,000 De
Stage 02:40% of the total fees, i.e. 4,00,000 De
Stage 03, 04 & 05: 45% of the total fees, i.e. 4,50,000 De
Stage 06:05% of the total fees, i.e. 50,000 De
The fees include the following major Job roles:
Site selection and surveying
Designing stage
Construction stage

It doesn't include Site Visits. If you want to include anything else beyond what is discussed above, and mentioned in this invoice, extra charges will be implied. Do you agree to the invoice created above? Or would you like to negotiate either the fees or the terms of service?

Client: That is pretty standard, and I knew before talking to you as to whom we are hiring. Everything seems fine, we accept to pay the proposed fees, and acknowledge the terms and service contract.

> Architect: Acknowledged. Post our conversation you'll be sent the detailed invoice and contract, upon signing which our time period of 01 month will begin. I've a good feeling towards this project. Thank you for choosing us.

AAI generated the conversation transcript using the web application and subsequently processed, analyzed, filtered, and organized it. After the information has been organized, AAI will proceed to conduct a sentiment analysis on the conversation. This analysis aims to evaluate the overall tone of the conversation, serving as a means for both the Client and the Architect to identify any potential issues. In the event that concerns arise on either end, the respective party will be allotted a specified timeframe to address these matters. In the absence of a satisfactory resolution within the stipulated timeframe, the project may be subject to cancellation, thus safeguarding the interests of both the Architect and the Client (Fig. 3).

The client has also furnished sources of inspiration in the form of images, videos, and articles. AAI will systematically examine these references, and the resultant data will be methodically organized and incorporated into the project's database for reference. After the completion of these steps, a formal and legally binding contract, structured in accordance with the architect's predefined terms and services, will be forwarded to the client for their endorsement. Upon the client's signature and acceptance of the contract terms, the ADP advances to the subsequent stage.

The BTA, CA, SA, and SXA play crucial roles in data extraction, collecting information regarding building typology, user count, contextual details, site specifics, program prerequisites, and sensory experiential demands. This information is then systematically structured and presented as input to the CSA, which proceeds to search for pertinent case studies corresponding to the keywords identified by the aforementioned agents. CSA generates new structured data encompassing text, images, and technical drawings based on the retrieved case studies. Simultaneously, the architect contributes personal inspirations, ideas, and concepts, in the form of case studies, sketches, and drawings, which are also relayed to CSA for the identification of relevant case studies and further data structuring.

Subsequently, this amalgamated information is presented to the VA to formulate refined design concepts, which are then presented to the client for evaluation. Upon receiving client feedback and approval, the data is channeled back to AAI, which disseminates it to the relevant agent classes (CCAs and VA) for the development of design visualizations. During this iterative process, the agents engage in inter-agent communication. VA concurrently produces technical drawings, such as Plans, Sections, and Elevations, while consulting SMA for real-time AI simulation data regarding climate, energy, and sustainability factors. It also interfaces with CNA to access information on material science, construction techniques, and construction details. BA oversees adherence to budget constraints.

Upon completing a single iteration, 50 design options are presented to the architect, who reviews and provides input. The design alternatives are presented in a tabular format, facilitating a comparative analysis of strengths and weaknesses. Design modifications extend beyond mere drawings and encompass 3D models, amenable to parametric adjustments through the web application. The architect iterates, potentially introducing new references, which prompts CSA to intervene and restart the cycle until a consensus is reached. Upon selection of three design options by the architect, these choices are presented to the client. The client may accept, reject, or request revisions, potentially initiating further iterations until satisfaction is achieved.

Upon the client's acceptance of one proposal, the information is relayed to AAI and subsequently to the architect for any final adjustments. This marks the last design phase, after which AAI distributes the data across the agent classes, ultimately reaching VA to generate construction-ready technical drawings for on-site execution. These drawings are shared with the architect for review and, if no revisions are necessary, passed to the client for final approval.

BA compiles a tender document for the finalized design, which AAI forwards to construction agencies. EA initiates BDA for project bidding. The winning BDA is awarded the construction project, pending approval from the architect and client. Once approved, EA commands the 3D printing of construction components. EA and AAI maintain ongoing communication, keeping the architect, client, and construction agency informed. EA calls LA to oversee the delivery of 3D-printed materials and autonomous construction robots managed by WAs on-site. EA provides live-streamed updates of the construction process to the construction agency and AAI, which subsequently relays real-time updates to the architect and client. Depending on the contract, the architect may visit the site for real-time inspections.

Upon project completion, EA and AAI facilitate quality checks conducted by the QAA, involving the architect, construction agency, and client. Following satisfactory design and construction, the client formally signs off on project completion, providing feedback. EA and AAI record this feedback to inform process enhancements and create appropriate documentation. As per the initial design contract, the client had agreed to post-occupancy surveys and feedback. EA and AAI collect this data, which includes sensor-based insights and user feedback, contributing to continual process improvement and project execution enhancements as can be seen in Fig. 3.

This concludes the proposed Smarter ADP. The subsequent section will offer a comprehensive discussion on existing technologies, required innovations, and the path toward the realization of this ADP.

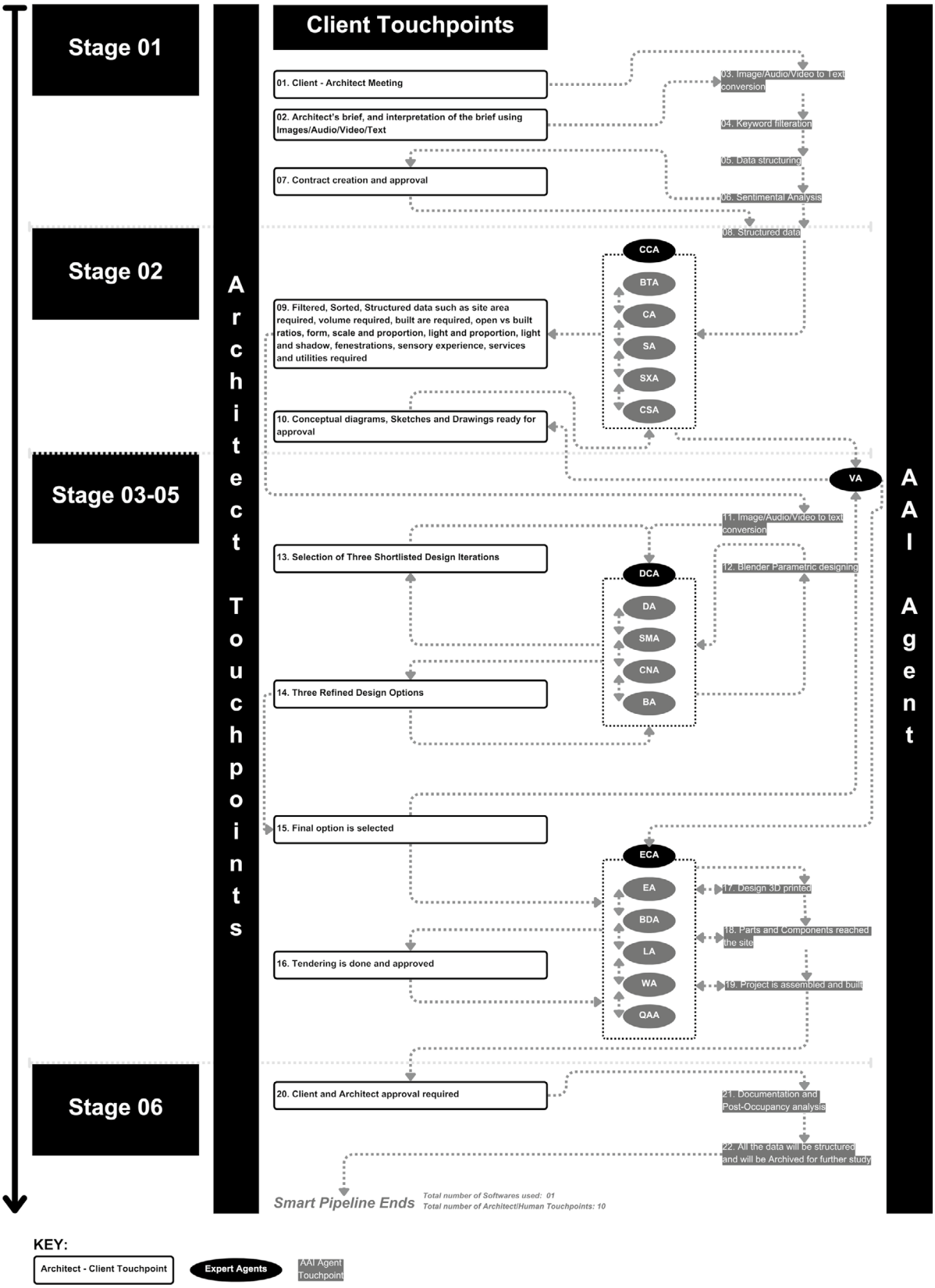


Fig. 3 Proposed ADP. *Source* Author

3 Conclusion

The need for an efficient Architectural Design Process (ADP) is paramount in light of the growing global population. The demand for housing is expected to surge, putting immense pressure on existing infrastructure and the architectural community. However, the architectural profession finds itself at a crossroads, grappling with divisions and challenges that hinder progress.

The schism within the architectural community is conspicuous, with two distinct camps emerging. On one side, there are those who approach architecture with a factory-like mentality, focusing solely on construction without due consideration for design and sensorial experiences. On the other side, there are those who view construction as a craft, valuing the deliberate pace of creation. This divide is exacerbated by the profession's reluctance to embrace new technologies, creating further fragmentation and lack of unity. The existing ADP is fraught with challenges that hinder its efficiency and effectiveness. Architects are burdened with multiple roles, and the industry often adheres to a rigid, waterfall framework for conducting business. Automation remains limited in various ADP tasks, necessitating significant investments in software proficiency. Real-time simulations and feedback between different stages of design are elusive, leading to information gaps and time-consuming processes. The translation of design ideas into execution often results in significant losses, while poor business practices persist.

Despite these challenges, the architectural field is making strides in adopting technological advancements. Efforts to standardize BIM (Autodesk, n.d.) and the use of AI models like GAN (Generative Adversarial Networks) (Nathan 2023), Diffusion, and CLIP (Contrastive Language-Image Pre-Training) for diverse tasks are underway. There is growing interest in parametric design (TestFit, n.d.) and nature-inspired computational design (Oxman, n.d.). However, these advancements encounter their own set of hurdles, including fragmentation, proprietary software development, and the need for programming skills.

This research took a qualitative approach to address these challenges, striving to develop an open, foundational system that can cater to diverse architectural needs. While the research was conducted diligently, it faced limitations, primarily related to the extensive resources required to fine-tune existing AI models.

The proposed ADP offers a promising solution to the existing challenges. The introduction of AAI, an AI assistant capable of understanding natural language, deciphering images, and translating between languages, can revolutionize the ADP. AAI's ability to oversee the entire ADP and convert structured language data into IFC file format streamlines standardization, eliminating the need for format conversions or multiple software installations. AAI addresses numerous existing ADP issues and technological limitations. Moreover, the utilization of pre-trained AI models minimizes time and resource investments, and AAI's adaptability facilitates seamless communication between various specialist AI models. AI-based simulations, already under development in AEC software, offer further enhancements. AAI promotes better coordination between stages and verticals, implementing an Agile framework where processes run in parallel. AI-driven image models can efficiently convert sketches into technical drawings. AAI's supervision curtails information loss from ideation to execution, preserving finer details. Leveraging open-source software like Blender enhances transparency, adaptability, and foundationality, reducing resource consumption.

The proposed ADP, while advantageous, is not without limitations. Fine-tuning AI models demands time and expertise. The system's reliance on API communication presents potential interruptions if technical issues arise. Architects may have reduced control due to increased automation. Furthermore, commercial-grade 3D printers and autonomous construction robots are prerequisites.

In conclusion, this research addresses the stagnation in the field of architecture and the need for an efficient ADP. Architects have the opportunity to transition from service providers to business leaders, leveraging AI to transform the industry. The proposed ADP, despite its limitations, offers a viable path forward, capable of significantly enhancing efficiency and productivity. The architectural profession stands at the threshold of a new Renaissance, driven by AI, which has the potential to democratize the field, remove barriers to entry, and empower clients to engage in design. The trajectory of this transformation depends on the industry's ability to adapt swiftly, minimize friction, raise awareness, and embrace the next era of architectural innovation.

Acknowledgements This work owes its existence to the invaluable insights and inspiration imparted by Professor Károly Zsolnai-Fehér, the curator of the "Two Minute Papers" YouTube channel. His unsupervised and informal guidance, coupled with his extensive coverage of AI research papers from around the world, illuminated the boundless potential of AI and research papers in affecting global change. My profound gratitude goes to him.

I extend my sincere appreciation to the Department of Architecture at Indus University, where I had the privilege of conducting this research. The university's commitment to fostering research as an integral component of academic pursuits has been instrumental in this endeavor. I am also thankful for the stimulating brainstorming sessions with my colleagues and fellow students.

My heartfelt thanks and acknowledgment are reserved for my family members, including my brother, wife, and parents, whose unwavering support and nurturing environment have enabled my growth and research pursuits.

References

Autodesk A (n.d.) Sweco streamlines its data- and workflows for project and business success. Retrieved September 25, 2023, from https://www.autodesk.com/customer-stories/sweco-forma-data-workflows

Boateng R, Ofoeda J, Effah J (2019) Application Programming Interface (API) research: a review of the past to inform the future. Int J Enterp Inf Syst 15(3):76–95. https://doi.org/10.4018/IJEIS.2019070105

Lawson B (2010) How designers think: The design process demystified (Reprint). Elsevier Arch Press

Lee, C. (n.d.). Awareness as a first step toward overcoming implicit bias

Mehta P (2019) Democratically smartWorld. Acad Res Community Publ, 3(2), 199–211. https://doi.org/10.21625/archive.v3i2.512

Nathan S (2023). How to improve a floor plan. Rayon. Rayon.Design. https://www.rayon.design/blog/how-to-improve-a-floor-plan

Norouzi N, Shabak M, Embi MRB, Khan TH (2015) The architect, the client and effective communication in architectural design practice. Procedia Soc Behav Sci 172:635–642. https://doi.org/10.1016/j.sbspro.2015.01.413

OpenAI (2023). GPT-4 Technical Report (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Oxman O (n.d.) OXMAN. OXMAN. Retrieved September 25, 2023, from https://oxman.com

TestFit T (n.d.) Testfit overview. Real estate feasibility platform. Retrieved September 25, 2023, from https://www.testfit.io/overview

United Nations Department For Economic And Social Affairs (2023) World population prospects 2022: Summary of results. U N.

Wang Q, von Tunzelmann N (2000) Complexity and the functions of the firm: Breadth and depth. Res Policy 29(7):805–818. https://doi.org/10.1016/S0048-7333(00)00106-2

SketchPLAN Recognition and Vectorization of Floor Plan Sketches for Building Information Modelling Design Environment

Ilyas Abdelmoula, Jens-Uwe Schulz, and Thomaz da Silva Lopes Vieira

Abstract

Our work builds upon general-purpose sketch recognition research, proposing a framework to integrate sketches within architectural design software. The use case being freehand-drawn floor plans as input for the Building Information Modelling environment. For that, we developed the SketchPLAN process, which recognizes, vectorizes, and contextualizes floor plan sketches. For recognition, SketchPLAN leverages the power of conditional Generative Adversarial Networks (pix2pix) to generate semantic segmentation maps out of floor plan sketch inputs, after training it with our own annotated dataset. The next step consists of vectorizing the segmented raster output through a floor plan vectorization library. A graphical user interface was designed to streamline the process. Bringing its output into Revit environment goes through the pipeline of "Rhino. Inside", converting geometry objects into Revit type instances. SketchPLAN demonstrates how artificial intelligence tools can be harnessed to bring floor plan sketches as input into BIM environments. In terms of contribution, we have collected and annotated an image dataset of floor plan sketches. The annotation system was thought right from the start as a means to vectorize those images. Moreover, our developed floor plan vectorization library performed well in converting those segmentation maps into usable geometry. The proposed solution shows nevertheless some deficiencies: struggling to omit all background noise when recognizing, and not considering some special cases when vectorizing. SketchPLAN holds however a bigger potential, considering the possible improvements on the process in terms of data collection, model architecture, more comprehensive vectorization, as well as implementation in other scenarios such as design recommendation tools.

Keywords

BIM · Floor plan · HCI · Sketch recognition · Vectorization

1 Introduction

Graphical representation is an essential component of architectural design. Across all project stages, a variety of representation modes are involved, mainly within the framework of descriptive geometry (Perez-Gomez & Pelletier 1992). Sketches in turn come as a malleable tool, which still functions under such a codified framework while giving the designer enough liberty to smoothly portray their thoughts. Thanks to their flexibility and level of abstraction, they still hold their relevancy even today faced with the dominance of digital technology, serving to channel internal thoughts, as well as to communicate ideas to others (Nalbach & Figa 2003).

From a platform standpoint, digital tools opened new doors to virtually navigate through designed space, with computational design tools taking a data-driven approach, and BIM ones focusing on building semantically informed 3D models, seeking better interoperability. Indeed, Building Information Modelling came as an AEC-specific framework of collaboration, which goes beyond what CAD workflows

I. Abdelmoula (✉)
BIM Automation Specialist at GOLDBECK GmbH, Bielefeld, Germany
e-mail: ilyas.abdelmoula@goldbeck.de

J.-U. Schulz
Professor at Department of Structural Engineering and Design, Ostwestfalen-Lippe University of Applied Sciences and Arts, Detmold, Germany
e-mail: jens-uwe.schulz@th-owl.de

T. da Silva Lopes Vieira
Scientific Researcher in Computational Design at Ostwestfalen-Lippe University of Applied Sciences and Arts, Detmold, Germany
e-mail: thomaz.da@th-owl.de

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_6

have to offer in terms of the amount of incorporated information within a project's digital model. In BIM jargon, this can be understood through the Levels of Detail (LOD), which indicates the level of both geometry and non-graphical information in a project model.

From a usage perspective, both computational tools and BIM software rely partly on user input in the form of commands and code. However, supporting human cognition in an even direct way constitutes a challenge to be addressed (Kensek & Noble 2014). Therefore, bringing sketches as software input could help in assisting the designer further by interpreting conceptual hand drawings and translating designer's intention more directly into the virtual environment.

Bridging the gap between hand drawings and digital tools could lead to an even better Human–Computer Interaction in the architectural context. Two principles substantiate the idea of a sketch-based architectural interface, namely "Know the User" and "Understand the Task" (Kim 2015). Both principles of Human–Computer Interaction point us to investigate bringing architects' sketch reliance to the digital realm, aiming to make a practical contribution while shedding light on the potential of sketch recognition in architectural design.

1.1 Framework and Scope

Thinking of freehand drawing as an additional input to architectural design tools leads us to conceptualize a framework which could handle different use cases (Fig. 1). Depending on the envisioned hardware interface, human input could be paper-based or digital.

From a processing perspective, the first step is to recognize sketch components and to extract semantic information. As this information must fit within already established design platforms, a crucial task is to contextualize it, by converting it to a compatible data structure, mainly vector-based. The next step is to harness this newly acquired information by processing it and generating corresponding output, which may also give valuable feedback to the user.

By projecting this framework on different sketching representation modes, several plausible usage scenarios emerge (Fig. 2). They range from input for design recommendation algorithms, as suggested by Chaillou (2019), to 3D geometry reconstruction or query based on sketches (Gryaditskaya et al. 2020).

In our work, we will focus on floor plans as a representation medium. While floor plan recognition is a trending research problem, sketches in that regard are yet to be explored, hence the interest for this use scenario. The scope of the work will be on investigating ways to recognize, vectorize, and contextualize floor plan sketches in order to integrate them within a BIM environment.

2 State of the Art

From a Human–Machine Interaction standpoint, sketch-based input has been a subject of interest in different domains. Semantically interpreting them was demonstrated in user interface design (Huang et al. 2019; Suleri et al. 2019), slide making (Muneeb & Jeff 2014), electrical engineering (Reddy & Panicker 2021), and chemistry (Ouyang & Davis 2011). Hence our interest is to see how it could apply to an architectural context.

To approach our task in terms of related work, two major categories need to be considered. On the one hand, research works dealing with sketch recognition in general, which include sketch-specific tasks and their proposed algorithms. On the other hand, research which tackled the problem of floor plan recognition and vectorization.

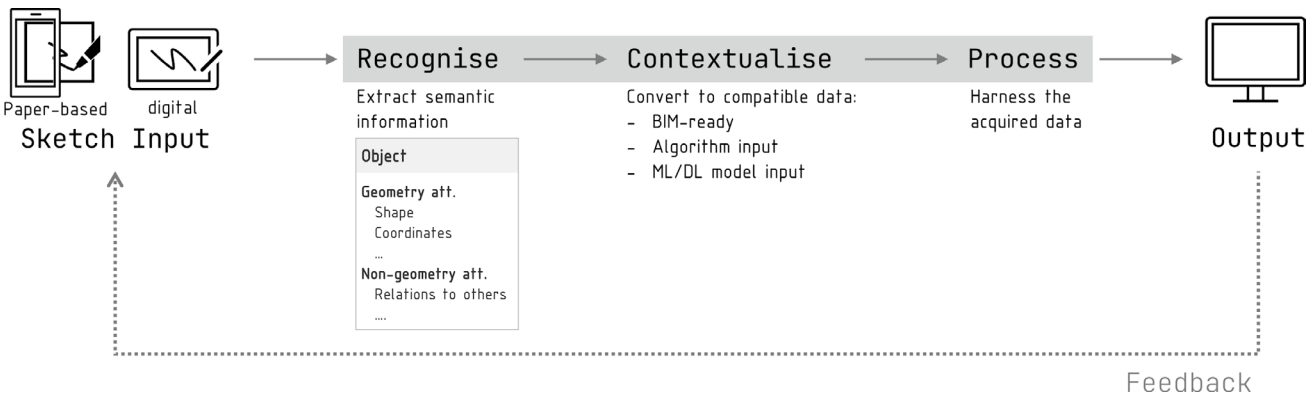


Fig. 1 Sketch integration within architectural design tools (*Source* author)

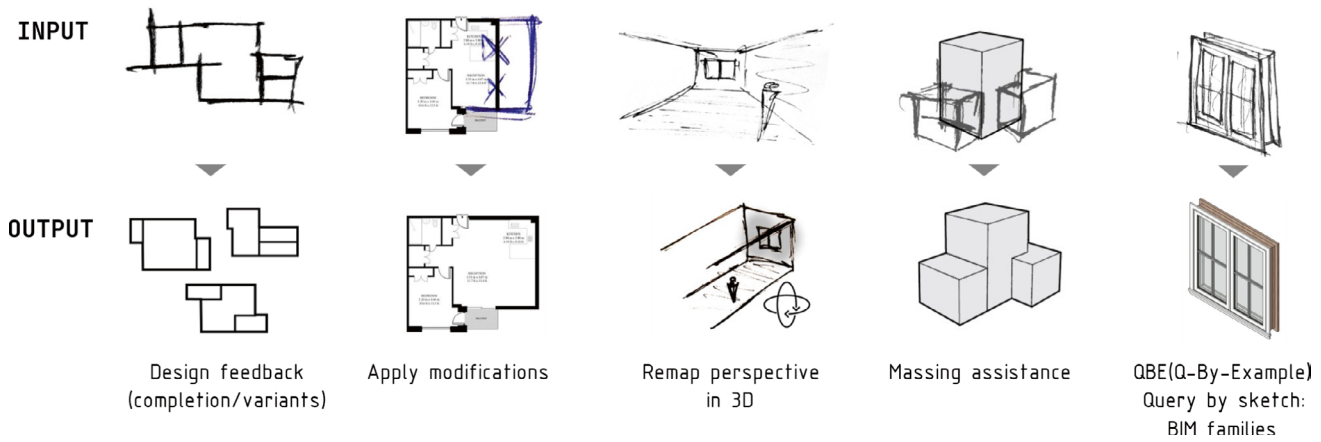


Fig. 2 Potential use cases for sketches (*Source* author)

2.1 Sketch Related Works

When it comes to sketches, comprehensive task taxonomies are available through specialized research survey (Xu et al. 2020). From that, we can distinguish two families with potential usage in an architectural context, namely 2D-oriented and 3D-oriented tasks. Since our topic is floor plans, we will lean more towards 2D-based papers.

One key task is Sketch recognition, which can be described as a classification problem. It was initially approached through hand-crafted features that showed their limitations faced to the sparsity of sketches, their level of abstraction, in addition to the variety of personal styles and level of expertise (Xu et al. 2020). Research has been lately stepping towards Deep Learning approaches Across the years. Convolutional neural networks (CNN) dealt with raster sketches (Yu et al. 2015). Recurrent Neural Networks (RNN) came after to harness the temporal component of vector sketches, namely the sequence of drawing strokes. Some works even combined both architectures to leverage pixel-based features and temporal data (Xu et al. 2018).

A more relevant task for us is Sketch Semantic Segmentation, in which a sketch is not classified as a whole, but rather its sketch strokes are classified based on predefined labels. Similarly to recognition, both CNN and RNN were used for that (Jiang et al. 2019; Wang et al. 2019). We can see how this can match our purpose, as dealing with a floor plan implies distinguishing its elements. Similarly, others considered element distinction as an object detection problem, by spotting hand-drawn symbols within a sketch (Thoma et al. 2021). Creswell even included generative model techniques to train with less labelled data (Creswell 2019).

2.2 Floor Plan-Related Works

In the lack of literature dealing with floor plan sketches, we will dedicate this part to discuss previous work in relation to regular floor plan images. Multiple datasets were compiled targeting mainly residential floor plans. A milestone was set by Lilfull, a large-scale dataset (National Institute of Informatics 2015). Moreover, bringing variety to datasets was of a high priority, either by including different graphical styles (Dong et al. 2021; las Heras et al. 2015), or by picking samples from a regional pool, like CubiCasa5k, with plans of a Finnish origin (Ahti Kalervo et al. 2019a, b). However, the closest to a dataset of hand-drawn floor plans was provided by ROBIN++dataset (Sharma et al. 2017), with 510 synthetic sketch-representations of floor plans. Being synthetically generated, its limitations lie in symbol repetitiveness, as well as single-style wall representation.

Regarding works related to floor plan recognition, two types of Neural Networks were implemented: Convolutional Neural Networks (CNN) and conditional Generative Artificial Networks (cGAN), aiming to generate corresponding segmentation maps, by colour coding each element of the floor plan. For Convolutional Neural Networks, some opted for encoding semantic elements (walls, openings, furniture) in a single map with additional maps for wall junctions (Kalervo et al. 2019a, b; Liu et al. 2017; Radne & Forsberg 2021), whereas others detached symbols from the map, spotting them through object detection architectures (Dodge et al. 2017; Lv et al. 2021). As per cGANs, they were used with the intention of harnessing their adversarial nature to bring the variety of graphical floor plan styles into a common domain (Cho et al. 2020; S. Kim et al. 2018). The obtained segmentation maps can

include more visual cues (e.g., door opening directions). Viewing it this way, the segmentation problem is tackled as an image translation task. This same type of neural network proved its suitability as well in floor plan generation (Chaillou 2019; Pan et al. 2021). A recent example (Swahn 2019) showed how the model was capable of generating increasingly readable floor plan suggestions out of rough hand-drawn contours.

As noticed above, all Deep Learning recognition methods lead to raster outputs, hence the need for a subsequent step of vectorization. This is usually done by constructing a grid of walls and then inferring doors and windows into it. Works based on CNNs, specifically on wall junction detection, use their detected junctions to iteratively build that wall grid (Kalervo et al. 2019a, b; Liu et al. 2017). This method, in spite of its effectiveness, does not perform well with non-orthogonal walls. In a related manner, we would like to hint also to other methods used outside of the floor plan context. Muneeb and Jeff (2014) used a general-purpose corner descriptor in the process of converting hand-drawn shapes into vector geometry for digital slides. Similarly, de et al. (2019) developed specific algorithms for vectorizing the lines of electrical circuit images.

It can be noticed that vectorization outputs were not necessarily oriented towards an eventual integration in design software. In some cases, it was conceived as a stand-alone representation in the form of SVG or high-resolution image output (Chaillou 2020; Kalervo et al. 2019a, b). In others, the aim was to obtain vector contour outputs, used to build 3D models for visualization purposes (Liu et al. 2017; Lv et al. 2021). And here, we can see room to develop outputs contextualized for design environments.

3 Method

Fitting into our suggested framework, our developed process aims at integrating hand-drawn floor plans in BIM environment. This process, named SketchPLAN, follows three main phases: recognition, vectorization, and contextualization (Fig. 3). Recognition relies on a deep learning model to identify the sketch components, vectorization brings that information into vector domain, and finally contextualization converts it into Revit type instances.

From a development perspective, the implementation was achieved in three parts (Fig. 4): a backend including recognition and vectorization, a frontend consisting of a graphical user interface, and finally BIM integration within Autodesk Revit. Having a stand-alone backend increases the tool's autonomy, which opens more possibilities for its integration in various workflows and scenarios.

Sections 3.1 and 3.2 will explain respectively recognition and vectorization parts. Section 3.3 is dedicated to the user interface overview, whereas Sect. 3.4 will go through the details of integration within the Revit environment.

3.1 Recognition

Recognition Strategy

Choosing an appropriate recognition strategy was crucial to implement our method, as it affects data collection, annotation, and the rest of the steps. Two factors have to be considered: the possibility to encode geometric information and the effort needed to collect and annotate a suitable dataset, with the lack of a proper one. Based on previous works, an interesting choice stands out: semantically segmenting sketches through image translation, using a cGAN. It deals with sketches as raster images, making the dataset collection relatively easier in comparison to vector data, and it outputs one segmentation map using a single deep learning model, in comparison to combining an object detector to spot floor plan symbols, and a segmentation map to recognize walls.

The used architecture is pix2pix. It consists of two models, a generator and a discriminator, competing against each other during training (Isola et al. 2016). There are other cGAN architectures fine-tuned for sketch input and adapted to its sparsity, faced to pix2pix which has a more one-to-one correlation between its input and output (Gao et al. 2020). We deem pix2pix nevertheless more fitting for us, as our use case is a colour coding task, which is more of a one-to-one translation task.

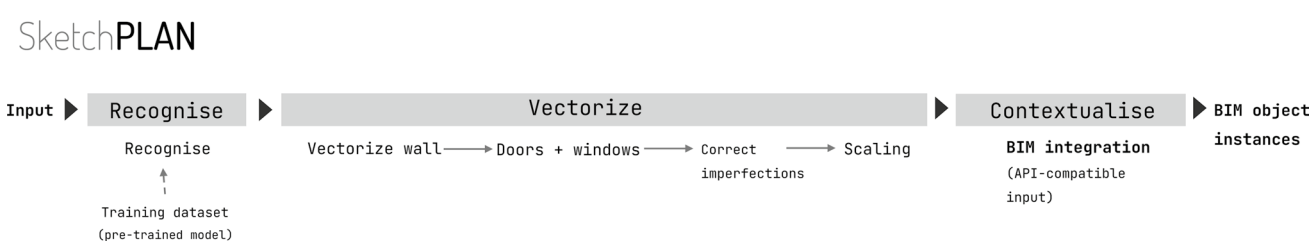


Fig. 3 Overview of the developed process SketchPLAN for integration of floor plan sketches in BIM environment (*Source* author)

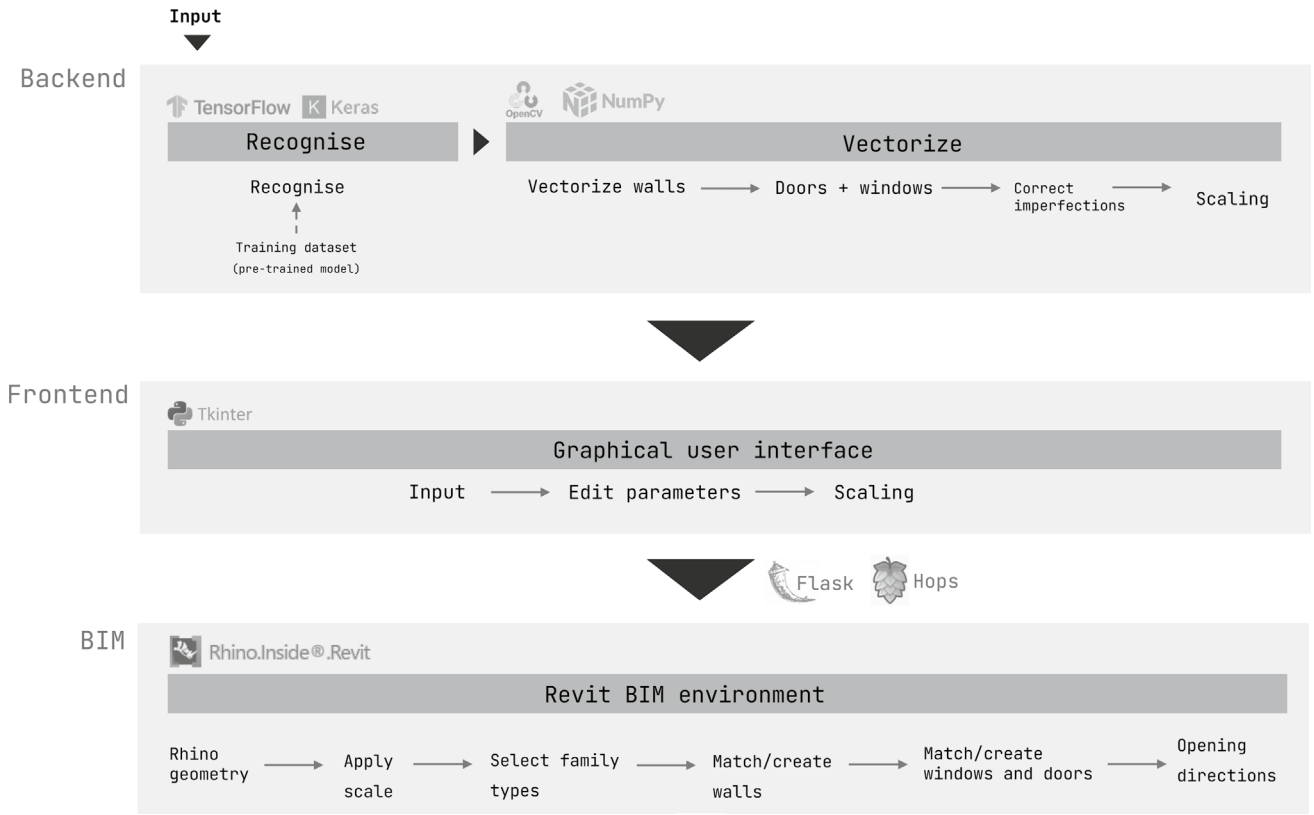


Fig. 4 SketchPLAN process implementation (*Source* author)

Dataset

In order to have a variety in sketches, we collected and annotated a dataset of 80 floor plan sketch images from different sources, mainly by assigning a group of students to reproduce existing floor plan samples with their own style within a short time frame (about 5 min, mimicking conceptual rough sketching settings), with the sample images picked from online datasets, mostly residential with straight walls (a total of 39 images). Adding to that contributions from volunteers through an online form (7 images), as well as four sketches from the ROBIN++dataset (Sharma et al. 2017). While all the abovementioned can be eventually published as open-access dataset, we had to add 30 web-crawled floor plan sketches, in order to increase the dataset size and variety, aiming to help the machine learning model to better generalize. As per the dataset size of 80 images, even though having a bigger dataset would be recommended, the total number of images was mainly dictated by the obligation of manually annotating them, a task which we tackled reasonably with respect to our available resources.

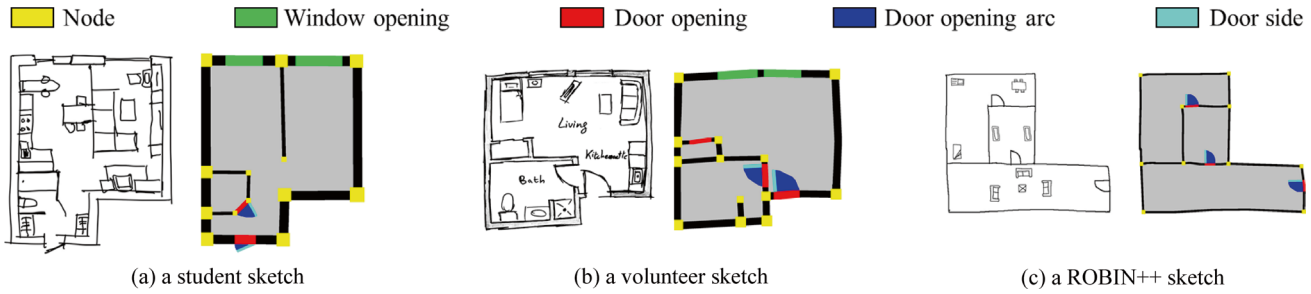
Regarding annotation, since training our model requires having image pairs, we had to annotate each sketch by making a matching segmentation map (Fig. 5). We have focused on the core elements of a floor plan: walls, doors, and windows. The principle behind it is to colour-code those objects in a way that conveys useful information to the next phase of vectorization. For instance, a straight wall is defined by its endpoints or nodes, hence the need for a particular colour for them (yellow). Likewise, doors contain dimension properties as well as geometric ones, such as opening side and hinge position, which must be vectorized afterwards.

Model Training

The training was done on the prepared dataset using laptop hardware, with 20% held as a test set. With the limited dataset size, we had to augment it using different techniques as summarized in Table 1. We have experimented with both offline and online augmentation, applying a combination of operations: rotation, flipping, jittering, and shifting. By comparing their output, we try to bring some insights to our specific use case.

In this part, we will discuss the training outcome, as Sect. 4 will be dedicated to showing the results of the SketchPLAN process as a whole.

Table 2 presents an overview of quantitative training results in terms of loss values and qualitative ones in the form of image segmentation samples. We can interpret quantitative values following some pix2pix-specific rules, such as reaching balanced loss values between the



(a) a student sketch (b) a volunteer sketch (c) a ROBIN++ sketch

Fig. 5 Samples of dataset image pairs (*Source* author)

Table 1 pix2pix model training settings

Training setup	Offline Augmentation	Online augmentation	Dataset size	Epochs
(1) Reference	/	Flip (horiz/vert); shift	80	1800
(2) Rotation	/	flip (horiz/vert); shift; rotation	80	1800
(3) Scaling Nearest	/	flip (horiz/vert); shift	80	1500
(4) Hybrid	flip (horiz/vert); squeeze (horiz/vert)	jitter; shift	960	900

generator and the discriminator (TensorFlow 2021). The reference setup (1) shows a significant imbalance between Discriminator loss and Adversarial loss, Setup (2) demonstrates how including rotation in the augmentation helped to reduce that effect, as both graphs similarly fluctuate.

Setup (3) tested another aspect related to the nature of pix2pix. Since it requires 256×256 input images, our sketch samples had to be for the most part scaled down. In this setup, we used the Nearest neighbour scaling method instead of the bicubic default one, leading to aliased images with no transition pixels. We can see that the performance drops drastically as the model generates more false positives, which are especially observable at the level of wall nodes. This can be attributed to the small input resolution. By applying no interpolation during scaling, we lose plenty of meaningful pixels and correspondences between input and target images. Later on, pix2pixHD could be an interesting architecture to test as it has higher resolution inputs of 1024×1024 (Wang et al. 2017).

The last setup (4) was trained for a longer period of time, with a mix of on-the-fly and offline augmentation. With limited hardware performance, it was a compromise to avoid memory saturation. We can see that the results got cleaner, especially with more intricate floor plans. One thing to point out, this setup did not include rotation augmentation, which means there might still be room for improvement. Additionally, all augmentation has virtually a limit which requires real new data in the form of additional sketches and their annotation. This is deducible as our loss values could not, in any case, settle at the recommended 0.69 value of both discriminator and adversarial loss (TensorFlow 2021).

3.2 Vectorization

Plan Vectorization Library

The output of the trained model is a raster image. A vectorization Python library as part of SketchPLAN.

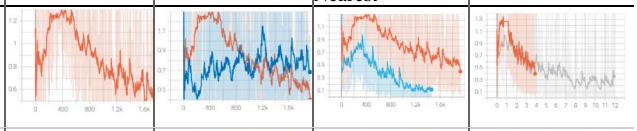
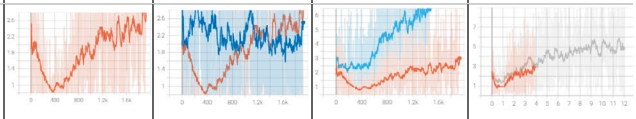
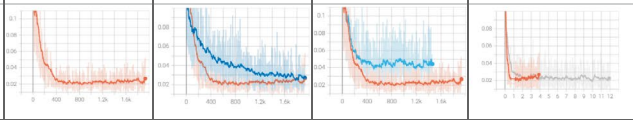
For such a library to properly function, we must classify floor plan components into encapsulated classes (Fig. 6): starting with wall components, Node, Door, and Window, then collecting walls and nodes in their respective containers, which contribute to building the whole floor plan. Each object includes relevant raster attributes related to colour masks, vector attributes containing geometric information, and meta-attributes used to describe its state.

On top of that, several helper methods were developed, built mainly on top of OpenCV and NumPy. Two types of helper methods were crucial for us: first, geometry-related methods (Fig. 7), which are the building blocks for geometry operations. Second, colour mask methods: they extract and process specific colour information from a segmentation map. They do so by creating colour masks based on an input colour threshold, as well as providing mask dilation and erosion functions.

Object Construction

In this part we will give an overview on how floor plan components are generated, the two key ones being nodes and walls. The principle behind is to generate two masks, one for nodes and the other for walls (Fig. 8). The latter is used to create separate contours for each wall, generating submasks. In parallel, by offsetting the node mask outwards by some additional pixels, we can get intersecting parts between each node and corresponding wall shapes. This is done in an

Table 2 Training results comparison (quantitative and qualitative)

Evaluation		(1) Reference	(2) Rotation	(3) Scaling Nearest	(4) Hybrid
Quantitative	Discriminator loss				
	Adversarial loss				
	L1 loss				
Qualitative (test set)					

elementwise way, using individual object sub-masks, which allows to create necessary associations to construct each wall.

However, applying those steps on deep learning model outputs is not as straightforward. This is mainly because those generated segmentation maps are by nature a prediction based on training, meaning they could provide ambiguous distinctions between objects, and they could generate false positives as well, as seen in Sect. 3.1.3. Therefore, we have to implement additional optimization steps to alleviate that limitation.

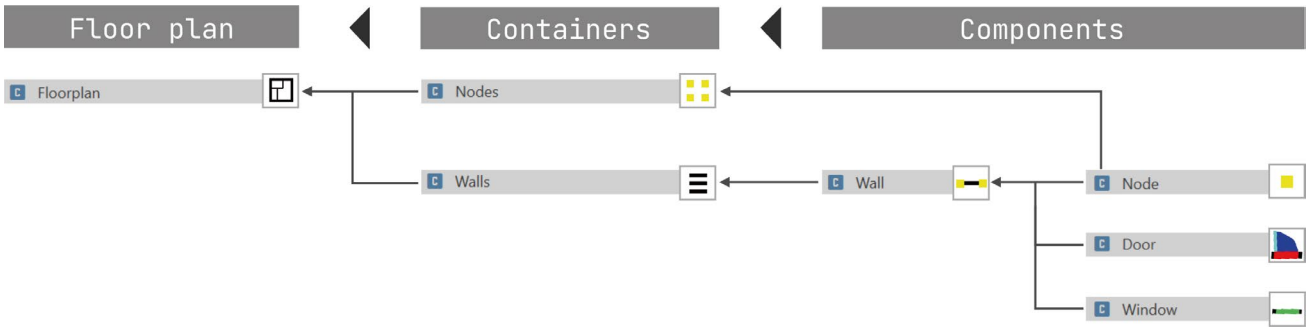


Fig. 6 Encapsulation of object classes of the library (*Source* author)

Fig. 7 Geometry-related helper methods (*Source* author)

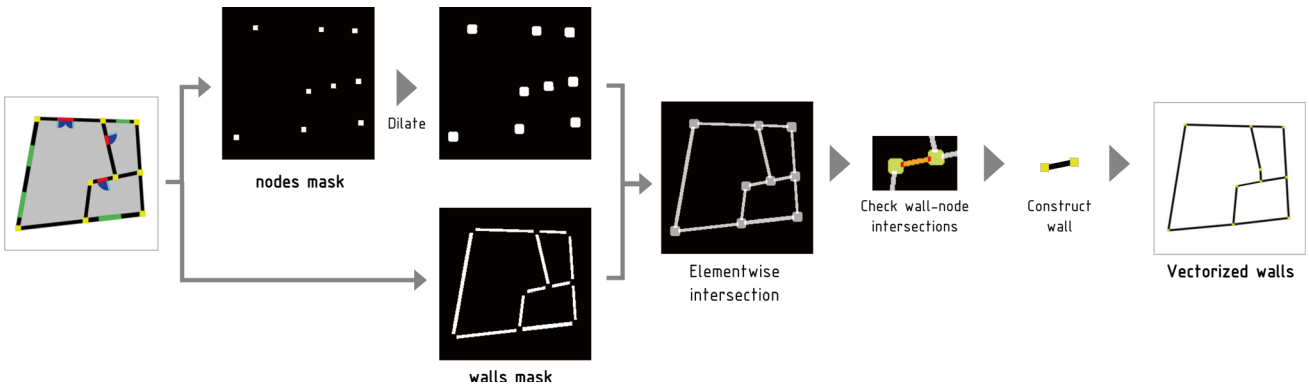


Fig. 8 Walls identification principle based on node and wall masks (*Source* author)

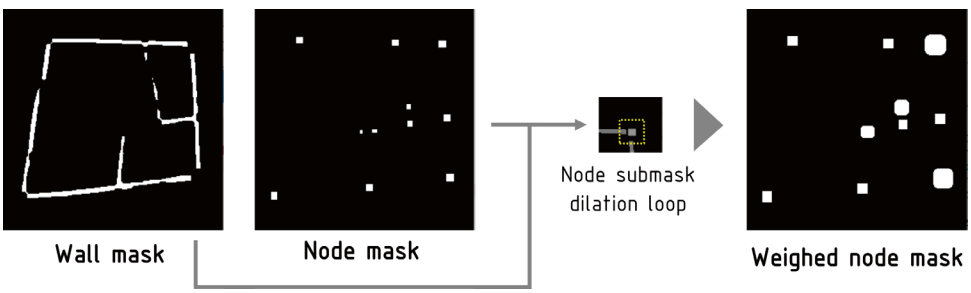


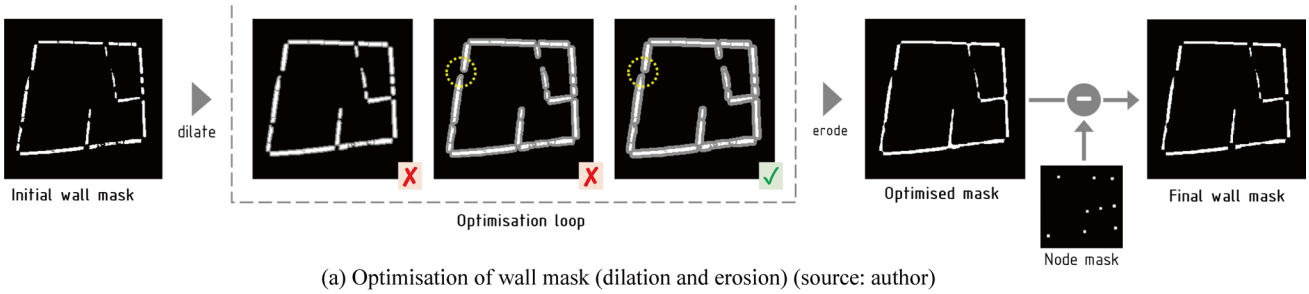
Fig. 9 Optimization of node mask (individual dilation values) (*Source* author)

– Nodes:

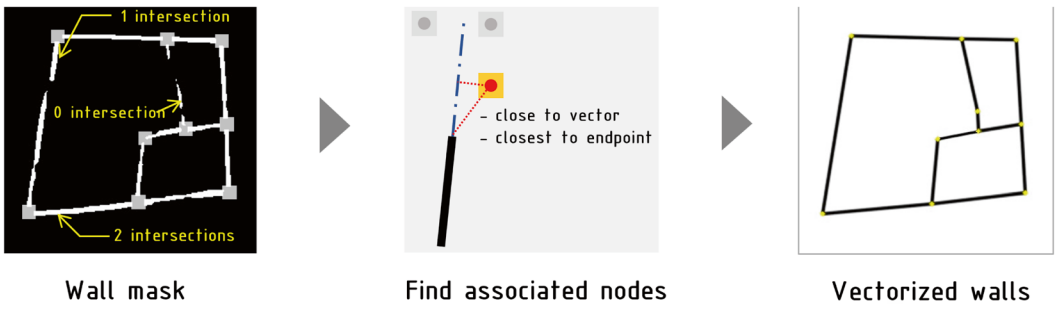
Before offsetting nodes, an optimization loop is incorporated to specify the best offset value for each node, based on the number of its intersections with walls (Fig. 9): the more intersections it has, the more likely it will produce better results. The outcome is a weighed node mask with individually dilated nodes.

– Walls:

The generated wall mask produces usually some unnecessary gaps within walls. In that sense, before intersecting it with node sub-masks, we must first eliminate those gaps as much as possible. To do so, offsetting the mask outwards and then inwards by the same value tackles for the most part this issue (Fig. 10a).



(a) Optimisation of wall mask (dilation and erosion) (source: author)



(b) Finding associated nodes for walls with no intersections (source: author)

Fig. 10 **a** Optimization of wall mask (dilation and erosion) (*Source* author) **b** Finding associated nodes for walls with no intersections (*Source* author)

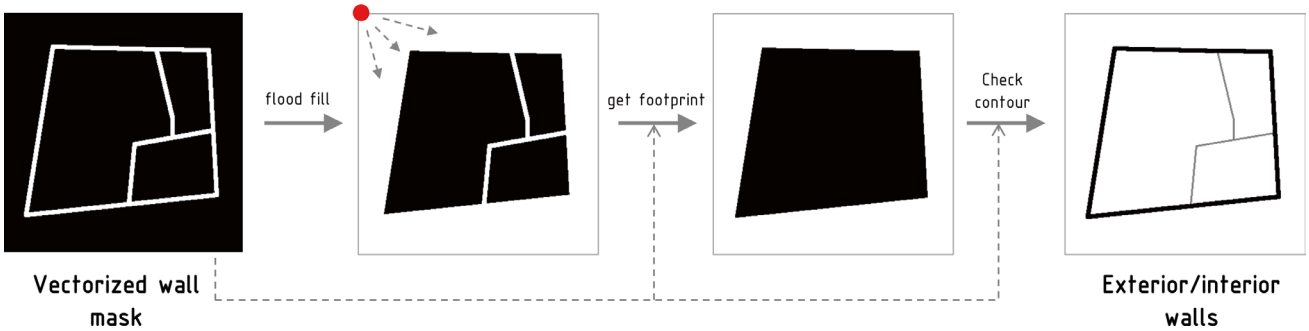


Fig. 11 Wall types (exterior/interior) based on footprint (*Source* author)

Even with that, after intersecting with node sub-masks we end up nevertheless with three possible cases per wall: two intersections, one intersection, or no intersection at all. The latter two cases require additional handling (Fig. 10b): inferring the wall shape's direction line and looking for its closest available nodes. Besides its effectiveness, this step may result in duplicate wall objects, which are purged later based on their matching endpoints.

Once all the walls are generated, one property can also be deduced pixel-wise, namely classifying walls as exterior or interior based on the floor plan footprint (Fig. 11). This is a useful information for attributing corresponding wall types later in the BIM model.

– Doors:

Identifying doors go through two stages (Fig. 12). The first is to associate doors with their host walls through sub-mask intersections, and the second consists of indicating the swing direction for each door object based on its hinge sub-mask.

– Windows:

Similarly to doors, windows are identified through element-wise intersections between window sub-masks and the vectorized wall mask (Fig. 13).

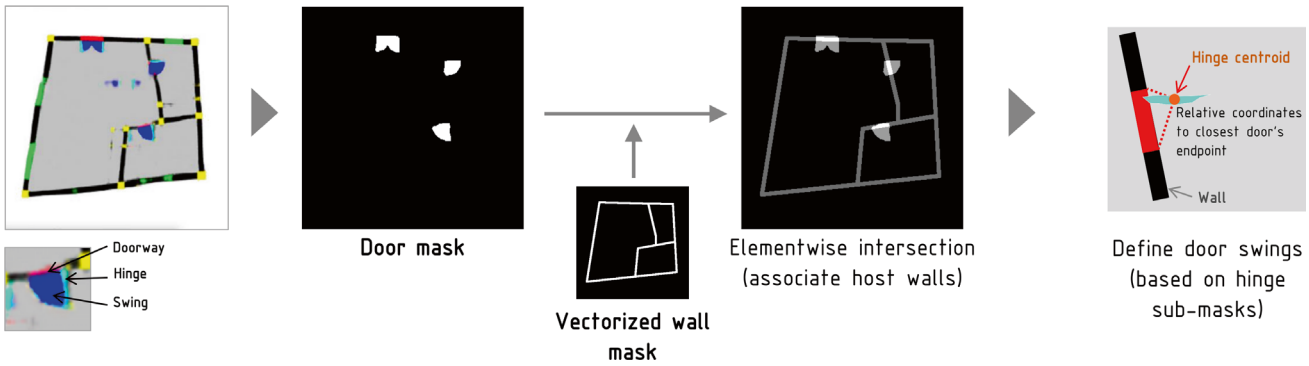


Fig. 12 Door generation process (mask intersection + swing direction) (*Source* author)

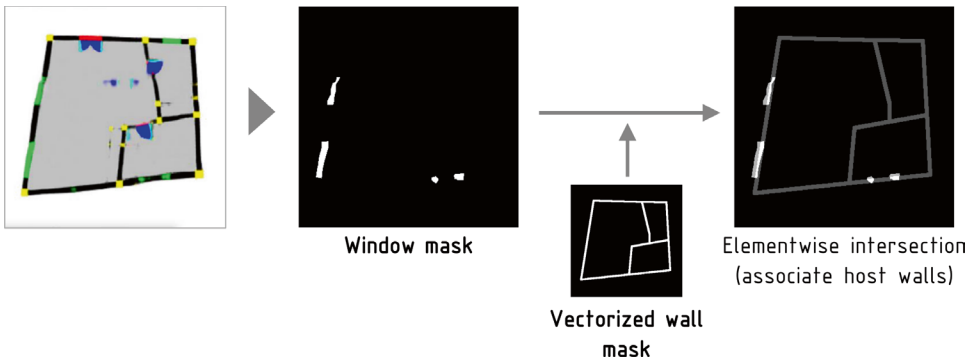


Fig. 13 Window generation process (*Source* author)

Angle Imperfections Correction

As we are dealing with floor plan sketches, they are prone to freehand drawing imperfections, particularly when it comes to wall orthogonality. To overcome that, the user specifies angle increments for walls (e.g., 45, 90°) and an approximation threshold. If a wall angle lays within that threshold, it will be realigned.

The imperfection correction algorithm goes through walls in a loop. Depending on the loop order, this could result in distorting already corrected walls, due to their shared nodes. To solve this issue, additional attributes were added to keep corrected nodes locked as illustrated in Fig. 14.

Main Function

The whole vectorization process, from colour mask extraction and object construction to angle correction, is incorporated in a "vectorize" method within the floor plan object class. Its main arguments are the input image and its segmentation map. The rest of the parameters are set to fine-tune the output for floor plan components.

3.3 Graphical User Interface

Shaping the solution into a user-friendly package is an important step. The developed user interface is bundled with a pre-trained version of our sketch segmentation model, and it has access to the vectorization library.

SketchPLAN's user interface has three sections (Fig. 15). The top one shows a preview of the loaded sketch image, its generated segmentation map, and a preview of the vectorized output. The section beneath contains buttons both to load input image and to add reference scale. As for the last section, it includes mainly sliders representing parameters to adjust the vectorization process and its outcome in terms of colour mask processing and imperfections correction. When the user starts to adjust those parameters, an additional preview of wall masks and node masks will appear on the right side of the window (Fig. 16a). This gives more insight into the impact of those parameter adjustments.

In order to attribute a metric value to the vectorized floor plan, as required for later transition to a BIM model, the user specifies two reference lengths in the X-axis and in the Y-axis in the "Define scale" window (Fig. 16b). That input will be used to calculate the corresponding scaling factors. As a preview for that, a representative coordinate system will be shown at the lower left corner of the floor plan.

Once reaching a satisfying result, clicking on "Run flask stream" button will launch an internal Flask server that streams the vectorized information to Grasshopper.

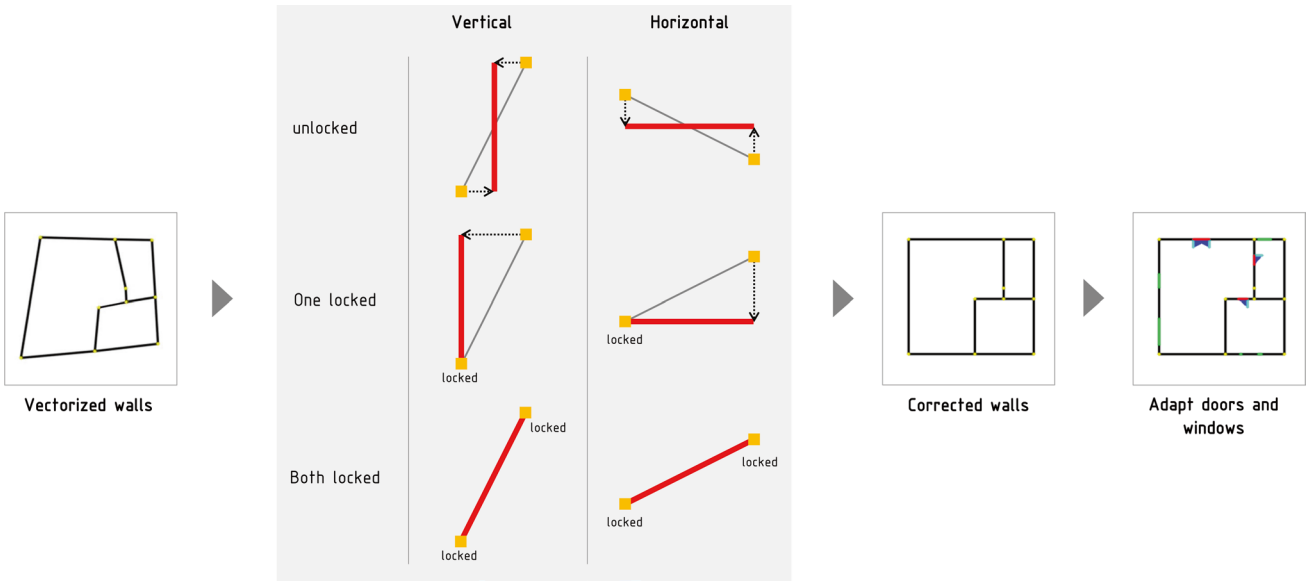


Fig. 14 Angle imperfection correction (*Source* author)

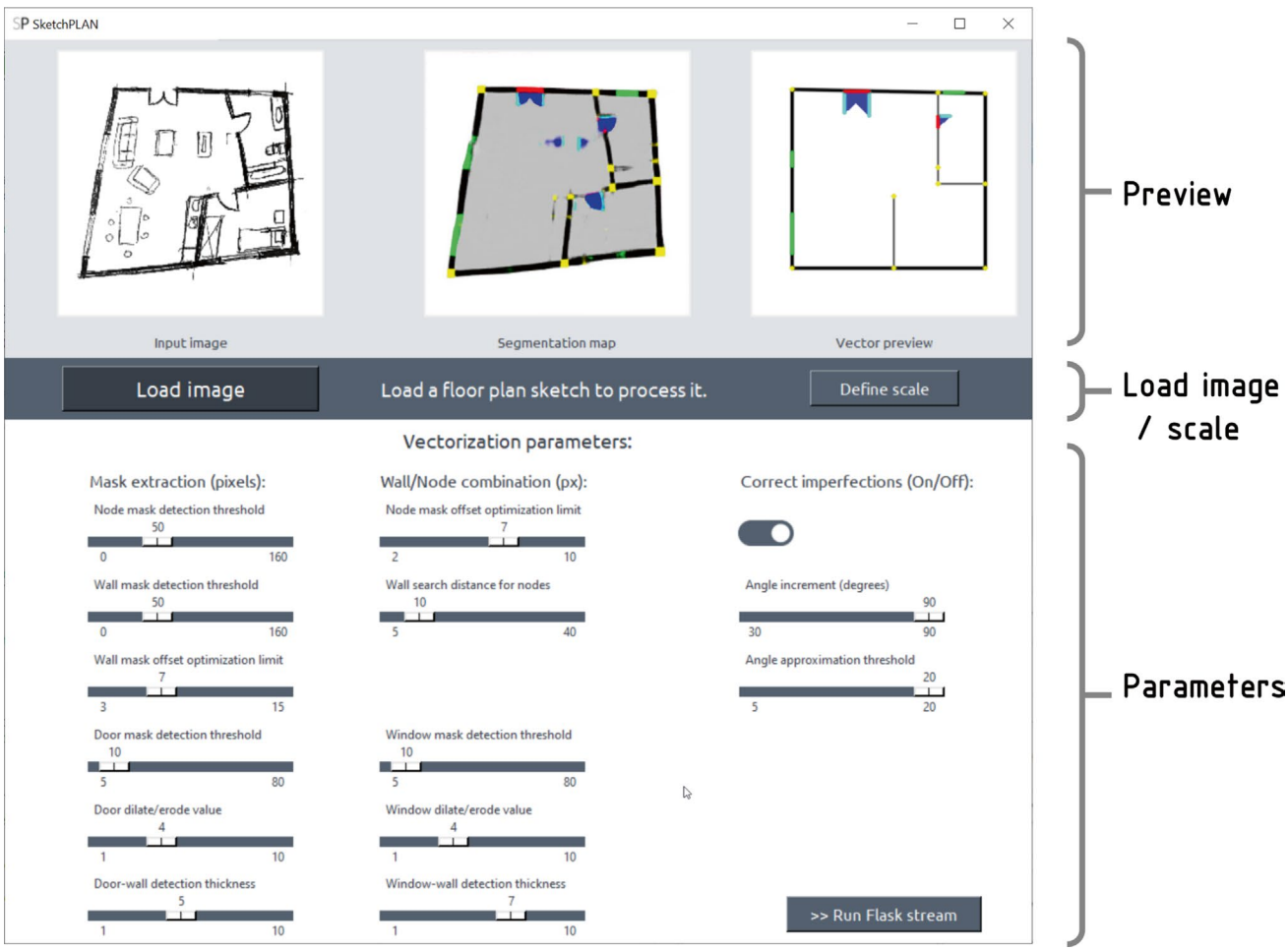
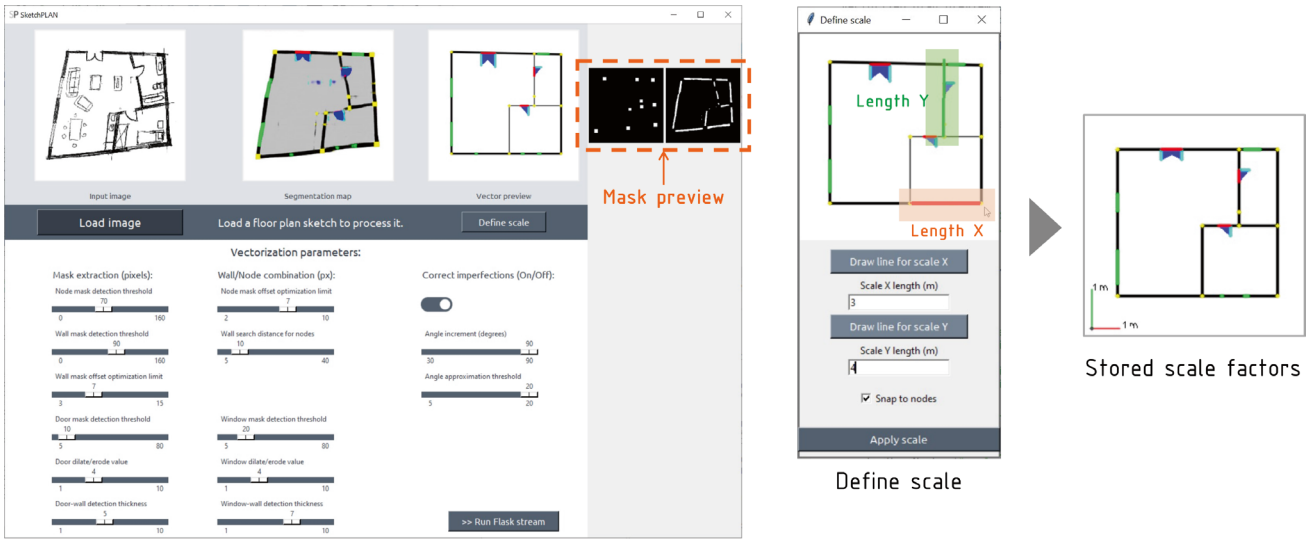


Fig. 15 SketchPLAN user interface (*Source* author)



(a) Mask preview while adjusting vectorisation parameters (source: author)

(b) Window for defining scale (source: author)

Fig. 16 **a** Mask preview while adjusting vectorization parameters (*Source* author) **b** Window for defining scale (*Source* author)

3.4 BIM Environment Integration

To bring vectorization output into Revit software, we have to fulfil three requirements: converting vector information to a datatype readable by Revit, creating a streamlined connection, and having access to Revit project file with all its families and types. Considering these three points, Grasshopper came as a suitable candidate, allowing us to build a functional pipeline for our purpose (Fig. 17).

The conversion of geometry is done using Rhino3dm Library, which facilitates creating Rhino Geometry objects in various programming environments (GitHub-McNeel, 2022). Therefore, we were able to convert all floor plan vector components to Rhino curves.

Once the geometry is converted, it will be communicated to Grasshopper via Hops, a communication component capable of launching Flask webservers to stream data. As per connection to Revit, Rhino.Inside allows Grasshopper to run directly within Revit, accessing its model data.

Matching Revit Family Types

Converting floor plan components into Revit type instances requires a matching step. First, the user has to choose the model types for walls, doors, and windows. Those chosen types will be used as a reference to match floor plan geometry. However, before matching the dimensions of vector objects to those selected types, their meta-attributes need to match (Fig. 18), e.g., doors are classified based on their host wall, exterior/interior, then based on their leaf count, single/double.

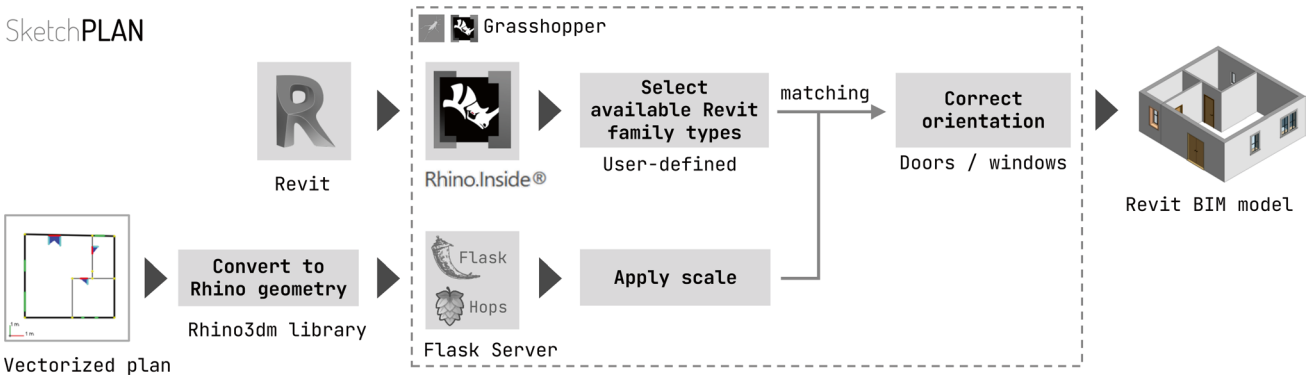


Fig. 17 Pipeline to integrate vectorized floor plan within Revit using Grasshopper (*Source* author)

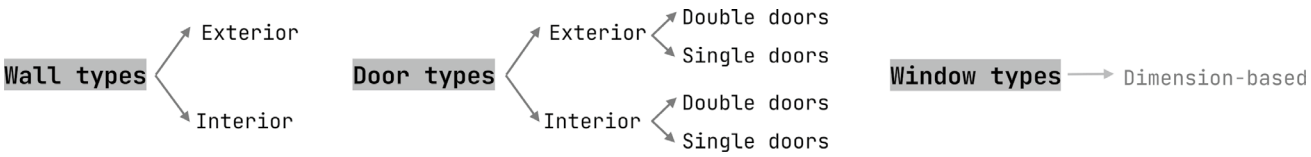


Fig. 18 Matching criteria for walls, doors, and windows (*Source* author)

Additionally, when it comes to placing doors and windows in Revit, two properties control their orientation: Facing and Handling. These parameters are manipulated by the algorithm as it places objects: it compares each door's position to its vectorized counterpart, and it flips the Revit object if necessary, in order to match the orientation.

BIM Model Output

By applying this process, we obtain a BIM model out of the vectorized floor plan geometry (Fig. 19). The dynamic link between generated objects and Grasshopper's script is maintainable. This allows for further manipulating or updating them using the same script. On the other hand, there is also the possibility to unpin those objects. This way, we can start to modify them directly inside Revit. In other words, that floor plan sketch becomes indeed an input for Revit software, as a starting point to develop the BIM project further.

4 Results and Discussion

In order to have a good impression on the overall performance of the process, it makes sense to do several tests on samples of floor plan sketches and document all their intermediate steps, i.e., starting from the segmentation map, vectorized output, until the generated BIM model.

Through Table 3, we can perceive the functionality of the whole process; however, taking a deeper look reveals some limitations. First, we see that the most impact on the overall accuracy is within the recognition step. Just by using different pre-trained deep learning models in (a) and (b), we get drastically varying results. This is observable by comparing Setup (1) and Setup (4) outputs. Each struggled/excelled in different respects than the other.

Furthermore, the deep learning model showed also mixed behaviour in terms of distinguishing relevant elements from background noise. While it did a good job in sample (e),

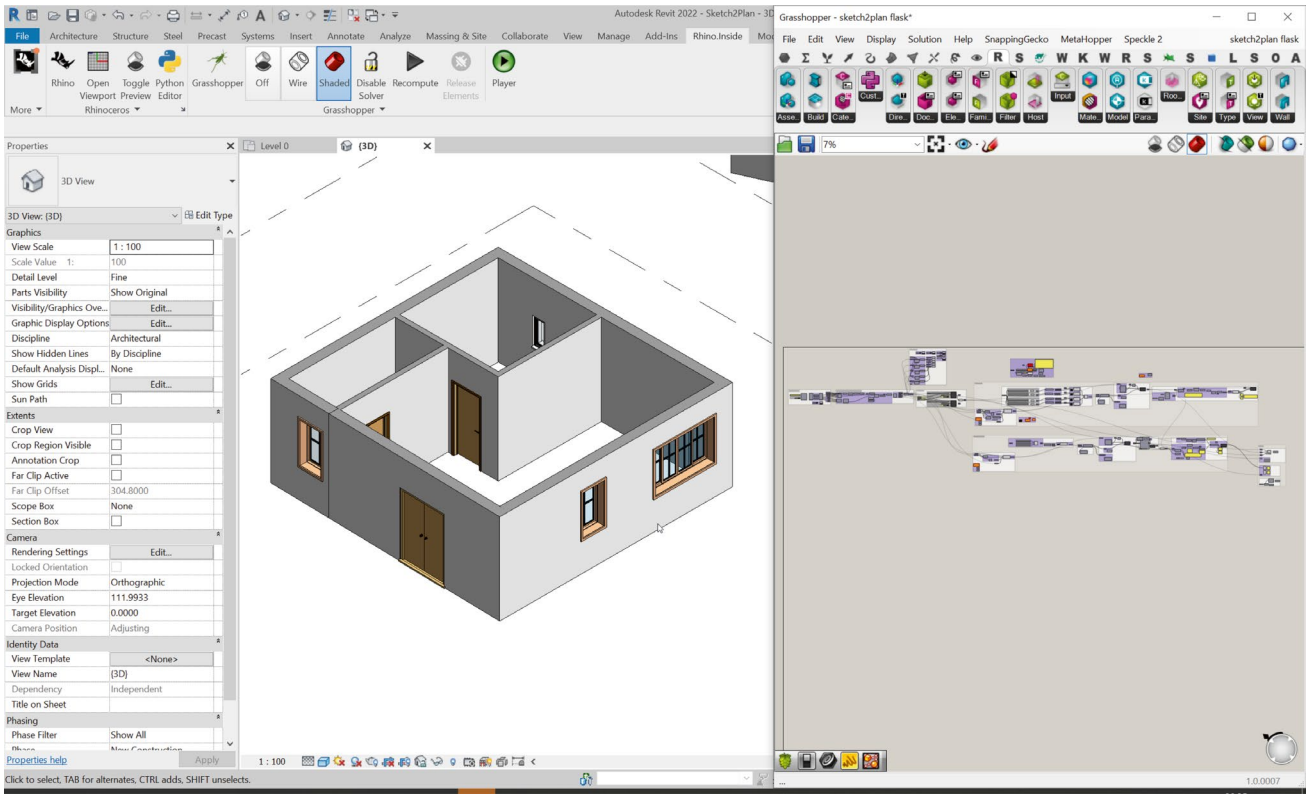


Fig. 19 Grasshopper script in action (*Source* author)

Table 3 Qualitative results overview for the whole process (recognition, vectorization, and BIM integration)

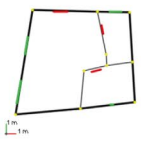
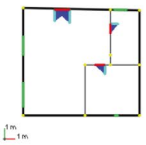
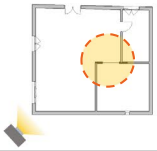
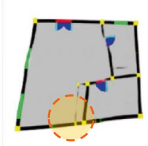
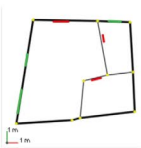
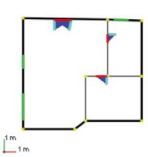
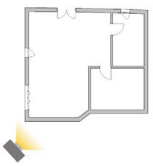
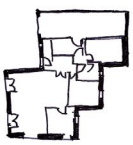
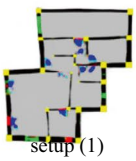
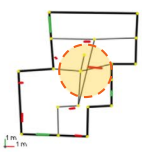
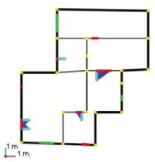
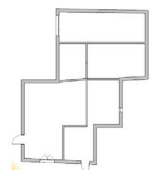
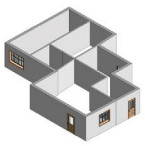
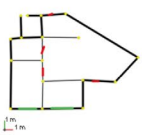
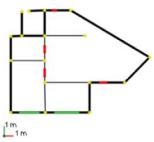
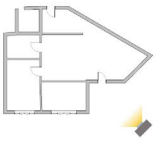
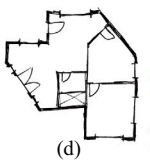
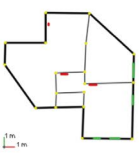
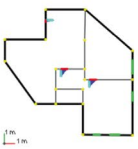
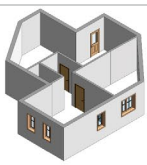
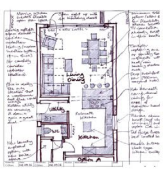
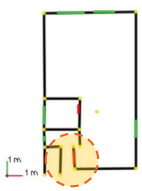
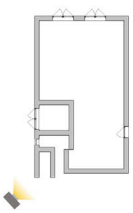
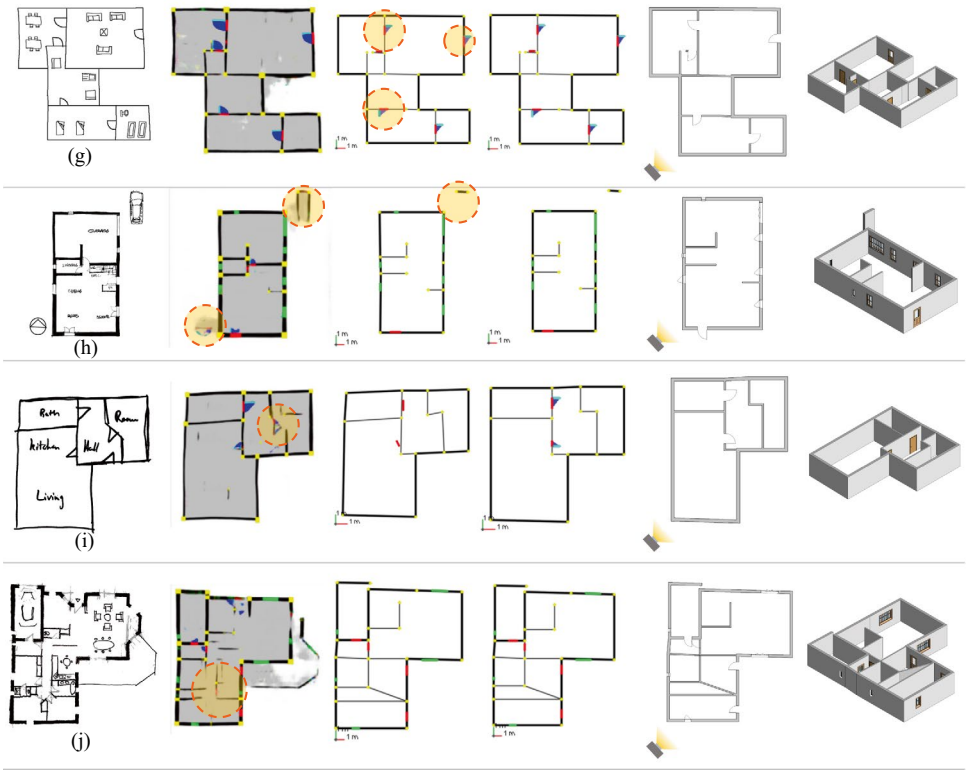
Sketch input	Generated segmentation map	Vectorized	Corrected angles	BIM model (plan)	BIM model (3D)
 (a)	 setup (1)				
	 Setup (4)				
 (b)	 setup (1)				
	 Setup (4)				
 (c)					
 (d)					
 (e)					
(f)					

Table 3 (continued)



ignoring hand-written annotations, it falsely interpreted outside equipment, namely car and entrance, in samples (f) and (h). This is mainly due to the limited number of annotated object classes. We assume that adding more object classes to segmentation maps would give more data to train a better model.

Moreover, detecting nodes for non-orthogonal walls was not fully successful in samples (c) and (d). Such an issue could be alleviated by applying more rotation augmentation and ultimately by adding more non-orthogonal floor plans to the training set.

As per vectorization, the algorithm delivered good output based on its fed segmentation maps. It has nevertheless shown an unexpected outcome in some cases. When correcting angles, it worked generally well. However, the node locking property results in some remaining noncorrected walls. Even after setting the correction to do several iterations, some walls just remain untouched. Generating doors has also suffered in some situations: telling too near doors apart (e) and ignoring a few of them (i), which is due to the fact that extracting colour masks is based on colour thresholds. Several doors also have wrong opening sides (g).

Finally, Rhino.Inside script did a decent job translating vectorized geometry into Revit type instances. The property matching algorithm managed generally to make good associations between geometry and family types, as long as the interpreted information in the two previous steps of recognition and vectorization was correct.

5 Conclusion and Outlook

Through this work, we have investigated how deep learning techniques can be leveraged to bring sketches to digital design tools for architecture. Our use case was the integration of freehand floor plans in Revit BIM environment, through the development of SketchPLAN recognition and vectorization process.

6 Contribution

We have proposed a framework for integrating sketch-based input into architectural design software. Through our demonstration with floor plans, we have collected and annotated an image dataset of floor plan sketches. The annotation system was thought of right from the beginning as a means to vectorize those images. Training pix2pix model on that dataset showed its capability to semantically segment hand-drawn floor plans, even with their disparate personal styles and imperfections.

Moreover, our floor plan vectorization library performed well to convert those segmentation maps into usable geometry. As a side feature, the same library can also be used for vectorizing regular floor plans, provided that a training dataset is made and annotated using the same colour coding logic. The whole software was packaged together into a graphical user interface.

We have also explored the integration with BIM through Rhino.Inside the Revit platform, which fulfilled our need of converting geometry into a proper Revit model.

7 Limitations

Our implementation shows indeed several drawbacks which point mainly to the recognition step. Indeed, the relatively small dataset does not contain enough occurrence of special cases such as exterior site components, balconies, annotation symbols, and tilted walls. Even with applying various augmentation techniques, the trained model was not capable to properly generalize on those cases. This suggests the need to gather additional sketch samples.

Further in the process, the developed vectorization library does not cover less common situations, like sliding doors and curved walls. Besides that, beyond adding scale and correcting imperfections, the graphical interface does not offer a more granular possibility to edit individual walls to move specific components.

8 Further Development

At each step of SketchPLAN, there is indeed room for improvement. In general, it evolves around enhancing the performance of the current system: mainly in terms of deep learning model, which would require bigger datasets to reach a more reliable stage, as well as adding more annotation classes to the dataset. The vectorization part needs more development to address special cases, as well as to improve angle correction by implementing transformation matrices as a more general-purpose technique. A more user-friendly interface could also be achieved by introducing two usage modes, a basic one with little to no adjustable parameters, and an advanced one with full access to vectorization parameters. Adding to that, the link to the BIM environment could also be more streamlined by building a native Revit add-in, directly linked to the backend.

As for the next investigation possibilities, rethinking each stage of the process could also lead to interesting results. For collecting data, an automated system could be helpful to generate synthetic sketches with different graphical styles, varied enough to better train the model. For recognition, other strategies could be investigated: trying other semantic segmentation architectures like DeepLab (Chen et al. 2018). Considering potentially vector-based sketches, RNN-powered recognition methods are interesting to explore in that regard. Other purposes could also be tried, such as inferring spatial configuration from sketches, in terms of rooms and relations.

9 Outlook

By taking an overall look, we can indeed see the utility of freehand input in digital architectural tools, together with the role of deep learning techniques as an intermediate. We noticed along the way the high potential of a direct pipeline between drawing a sketch and digitally manipulating it. Within a wider framework, more possibilities could emerge, especially if we think about other sketching types and digital drawing input devices.

Linking this work with other artificial intelligence tools could also open the door to other promising horizons for Human–Computer Interaction, such as design recommendation systems, real estate search websites, and collaborative platforms.

References

Ahti Kalervo, Juha Ylioinas, Markus Häikiö, Antti Karhu, Juho Kannala (2019) *CubiCasa5k*. Zenodo. https://zenodo.org/record/2613548https://doi.org/10.5281/zenodo.2613548

Chaillou S (2019) AI + Architecture. Towards a new approach. Master, Harvard University Graduate School of Design. www.academia.edu. https://www.academia.edu/39599650/AI_Architecture_Towards_a_New_Approach

Chaillou S (2020) Space Layouts & GANs. GAN-enabled Floor Plan Generation. Towar Data Sci. https://www.academia.edu/42289131/Space_Layouts_and_GANs_GAN_enabled_Floor_Plan_Generation

Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018). Encoder-Decoder with atrous separable convolution for semantic image segmentation. https://arxiv.org/pdf/1802.02611

Cho D, Kim J, Shin E, Choi J, Lee J-K (2020) Recognizing architectural objects in floor-plan drawings using Deep-learning Style-transfer algorithms. http://papers.cumincad.org/cgi-bin/works/paper/caadria2020_446

Creswell A (2019) Representation learning using generative models [Ph.D Thesis]

de, P., Mandal, S., Bhowmick, P., & Chanda, B. (2019) Robust vectorization method for electrical circuit drawings using component morphology. Pattern Anal Appl 22(4):1341–1359. https://doi.org/10.1007/s10044-018-0686-3

Dodge S, Xu J, Stenger B (2017) Parsing floor plan images. In 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA).

Dong S, Wang W, Li W, Zou K (2021) Vectorization of floor plans based on EdgeGAN. Info 12(5):206. https://doi.org/10.3390/info12050206

Gao C, Liu Q, Xu Q, Wang L [Limin], Liu J, Zou C (2020) SketchyCOCO: Image generation from freehand scene sketches. https://arxiv.org/pdf/2003.02683

GitHub—mcneel. (2022). GitHub—mcneel/rhino3dm: Libraries based on OpenNURBS with a RhinoCommon style. https://github.com/mcneel/rhino3dm

Gryaditskaya Y, Hähnlein F, Liu C, [Chenxi], Sheffer, A., & Bousseau, A. (2020) Lifting freehand concept sketches into 3D. ACM Trans Graph 39(6):1–16. https://doi.org/10.1145/3414685.3417851

Huang F, Canny JF, Nichols J (2019) Swire: Sketch-based user interface retrieval. In Brewster S, Fitzpatrick G, Cox A, Kostakos V (Eds), Proceedings of the 2019 CHI Conference on Human

Factors in Computing Systems (pp 1–10). ACM. https://doi.org/10.1145/3290605.3300334

Isola P, Zhu J-Y, Zhou T, Efros AA (2016) Image-to-Image translation with conditional adversarial networks. https://arxiv.org/pdf/1611.07004

Jiang J, Wang R, Lin S, Wang F (2019) SFSegNet: Parse Freehand Sketches using deep fully convolutional networks. https://arxiv.org/pdf/1908.05389 https://doi.org/10.1109/IJCNN.2019.8851974

Kalervo A, Ylioinas J, Häikiö M, Karhu A, Kannala J (2019) CubiCasa5K: A dataset and an improved Multi-Task model for floorplan image analysis. https://arxiv.org/pdf/1904.01920

Kensek KM, Noble D (2014) Building information modeling: BIM in current and future practice. Karen M Kensek, LEED BD+C, Assoc. AIA, Douglas Noble, FAIA, Ph.D, (ed) ; foreword by Chuck Eastman. Wiley

Kim GJ (2015) Human-computer interaction: Fundamentals and practice. Gerard Jounghyun Kim (1st). Auerbach

Kim S, Park S, Yu K (2018) Application of style transfer in the vectorization process of floorplans (Short Paper). In 10th International Conference on Geographic Information Science (p 6). https://doi.org/10.4230/LIPIcs.GISCIENCE.2018.39

las Heras, L.-P. de, Terrades, O. R., Robles, S., & Sánchez, G. (2015) Cvc-FP and SGT: A new database for structural floor plan analysis and its groundtruthing tool. Int J Doc Anal Recognit (IJDAR) 18(1):15–30. https://doi.org/10.1007/s10032-014-0236-5

Liu, C [Chen], Wu, J., Kohli P, Furukawa Y (2017) Raster-to-Vector: revisiting floorplan transformation. In 2017 IEEE International Conference on Computer Vision (ICCV)

Lv X, Zhao S, Yu X, Zhao B (2021) Residential floor plan recognition and reconstruction. Undefined. https://www.semanticscholar.org/paper/Residential-Floor-Plan-Recognition-and-Lv-Zhao/cdaa2f41b6e010e55dae03c75dbd27c2c27e34e1

Muneeb A, Jeff W (2014) Generation of slides from Hand-Drawn sketches. https://www.semanticscholar.org/paper/Generation-of-Slides-from-Hand-Drawn-Sketches-Ahmed-Wheeler/a80d625716185c65b96faff7f99b72140a57e6da

Nalbach G, Figa D (2003) Die erste Skizze: = The first sketch. Universität Dortmund Förderkreis Dortmunder ModellBauwesen

National Institute of Informatics. (2015). LIFULL HOME'S dataset. https://www.nii.ac.jp/dsc/idr/en/lifull/

Ouyang TY, Davis R (2011) ChemInk: A Natural Real-Time recognition system for chemical drawings. 4503–0419. https://dspace.mit.edu/handle/1721.1/78898

Pan Y, Qian J, Hu Y (2021) A preliminary study on the formation of the general layouts on the northern neighborhood community based on GauGAN diversity output generator. In Yuan PF, Yao J, Yan C, Wang X, Leach N (Eds), Proceedings of the 2020 DigitalFUTURES (pp 179–188). Springer Singapore

Perez-Gomez A, Pelletier L (1992) Architectural representation beyond perspectivism. Perspecta 27:20. https://doi.org/10.2307/1567174

Radne A, Forsberg E (2021) Vectorization of architectural floor plans: PixMax - a semi-supervised approach to domain adaptation through pseudolabelling [Master's thesis]. Chalmers Univ Technol, Gothenburg, Sweden

Reddy RR, Panicker MR (2021) Hand-Drawn electrical circuit recognition using object detection and node recognition. https://arxiv.org/pdf/2106.11559

Sharma D, Gupta N, Chattopadhyay C, Mehta S (2017) DANIEL: a deep architecture for automatic analysis and retrieval of building floor plans. In 14th IAPR International Conference on Document Analysis and Recognition: ICDAR 2017 : proceedings : 9–15 November 2017, Kyoto, Japan (pp 420–425). Conference Publishing Services, IEEE Computer Society. https://doi.org/10.1109/ICDAR.2017.76

Suleri S, Sermuga Pandian VP, Shishkovets S, Jarke M (2019) Eve: A Sketch-based software prototyping workbench. In Brewster S, Fitzpatrick G, Cox A, Kostakos V (Eds), Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp 1–6). ACM. https://doi.org/10.1145/3290607.3312994

Swahn E (2019) Latent spaces : final presentation—Studio 9: Architectural Notations. https://www.studio9.arch.kth.se/2019/06/10/latent-spaces-final-presentation/

TensorFlow (2021). pix2pix: Image-to-image translation with a conditional GAN. TensorFlow Core. https://www.tensorflow.org/tutorials/generative/pix2pix

Thoma F, Bayer J, Li Y (2021) A public Ground-Truth dataset for handwritten circuit diagram images. https://arxiv.org/pdf/2107.10373

Wang F, Lin SS, Wu H, Li H, Wang R, Luo X, He X (2019) SPFusionNet: Sketch segmentation using Multi-modal data fusion. In 2019 IEEE International Conference on Multimedia and Expo (ICME)

Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B (2017) High-Resolution image synthesis and semantic manipulation with conditional GANs. https://arxiv.org/pdf/1711.11585

Xu P, Hospedales TM, Yin Q, Song Y-Z, Xiang T, Wang L [Liang] (2020) Deep learning for Free-Hand sketch: a survey and a toolbox. https://arxiv.org/pdf/2001.02600

Xu P, Huang Y, Yuan T, Pang K, Song Y-Z, Xiang T, Hospedales TM, Ma Z, Guo J (2018). SketchMate: Deep hashing for Million-Scale human sketch retrieval. https://arxiv.org/pdf/1804.01401

Yu Q, Yang Y, Song Y-Z, Xiang T, Hospedales T (2015) Sketch-a-Net that beats humans. https://arxiv.org/pdf/1501.07873

Enhancing Architectural Plan Generation with Machine Learning and Space Syntax Analysis for Optimized Spatial Configuration

Mehmet Baraa Sabsabi and Hatice Kalfaoglu Hatipoglu

Abstract

Since recent years till now, a lot of research has been conducted on generative design to help architects in the design process to save time and cost, and produce multiple and better solutions for a given problem. Such solutions have been generating architectural plans based on users' preferences, generating architectural plans that reduce the dependence on energy for cooling by improving natural ventilation, and generating façade systems that reduce harmful sun light and increase natural light. However, there are a few number of research studies that consider the quality of spatial configuration in terms of privacy hierarchy and the spatial relationship of the automatically generated architectural plans. Spatial configuration is one of the most important features in the architectural design process, and using generative techniques in relation to Machine Learning (ML) in this process has become a requirement in recent years. Although a lot of studies have been carried out about shape grammar and ML relationships, there is not a study which combines spatial configuration using Space Syntax (SS) with ML, which can create a potential for this requirement. Therefore, in this paper, a computational framework has been developed to evaluate the spatial configuration of the generated architectural plans by training a supervised neural network on some spatial feature values of three Syrian houses (post-independence from the French colonization period). These values have been gained by analyzing these houses using the DepthmapX software, which is based on Space Syntax theory. The trained model has been tested on another Syrian housing plan from the same typology. The outcomes of the study demonstrate the potential of the trained model to predict the suitable space function with few errors caused by the strong similarity in spatial features of some spaces and the lack of training samples. The trained model can then be integrated into any plan-generating algorithm or used as a separate tool to enable architects to enhance their spatial configuration in the early design stage. Although the trained model is still under development without accomplishment, it creates a base for further investigations in terms of spatial conditions and ML.

Keywords

Space syntax · Spatial configuration · Machine learning · Generative design · Housing design

1 Introduction

Automation has revolutionized our traditional ways of doing things, especially in the industry sector, for many years by replacing human labor with robots that can increase production and reduce cost and time. With the advancement of technology and Artificial Intelligence (AI), we have been able to rely on machines not only in production but also in decision-making, from identifying junk e-mails from normal e-mails to recognizing cancer tumors based on images to self-driving cars. In the field of Architecture, Engineering, and Construction (AEC), a large number of computational generative design research studies have been conducted with the intention of elevating the use of computers from a basic drafting tool to an artificially

M. B. Sabsabi (✉)
Graduate School of Natural Science, Ankara Yildirim Beyazit University, Ankara, Turkey
e-mail: m.baraasabsabi@gmail.com

H. K. Hatipoglu
Faculty of Architecture and Fine Arts, Ankara Yildirim Beyazit University, Ankara, Turkey
e-mail: hhatipoglu@aybu.edu.tr

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_7

intelligent machine capable of assisting and helping architects and engineers in the design process to reduce cost, time, and human efforts, as well as generating many different alternatives to achieve more optimum and reliable results.

Some of the well-known generative design algorithms are L-systems (Lindenmayer 1968), Shape grammars (Stiny & Gips 1971), genetic algorithms (Holland 1975) and cellular automata (Wolfram 2002). Such computational generative design applications are generating multiple alternatives hosing plans according to users' preferences based on cellular automata algorithm (Dinçer et al. 2014), generate structural networks that could be suitable as architectural structural systems inspired by angiogenesis that simulate the vascular systems development to generate a structural network from two or three support point (Klemmt & Bollinger 2017), generating energy-efficient architecture solutions based on genetic algorithm (Caldas 2008), optimizing daylight and glare using genetic algorithm and machine learning to reduce time needed for simulation (Wortmann 2017), creation of architectural plans using shape grammar (Stiny & Mitchell 1978), "Finch" algorithms developed by BOX Bygg and Wallgren Arkitekter written in Grasshopper, which generate different spatial configuration based on the context boundary and total area (Wallgren 2020).

That could be the reason why most of the generative design techniques have not been able to generate architectural plans similar to the ones created by architects, as architecture in real life is more than just an arrangement of shapes and geometries. The main rule in architectural layout is the arrangement of the spatial configuration in such a way that each space has a functional relationship with the other spaces (Hillier & Hanson 1984). These spatial configurations are based on the understanding of the user's comfort, social relationships, habits, and needs, which can change according to peculiarities arising from differences in context and culture. These differences are clearly emphasized in Uyar & Griffiths (2017) research by comparing the spatial configuration of multiple traditional houses of three different cultures (Turkish, Japanese, and British) in terms of privacy, flexibility, and introvertedness using space syntax. The study showed that based on Convex Map analysis, the circulation spaces are the most integrated in British houses, while in Turkish and Japanese houses, the living rooms are the most integrated. Based on Axial Map analysis, the most connected space in Turkish houses are the living rooms, in Japanese houses are the open spaces, and in British houses are the circulation spaces. Based on Justified Graph analysis, the bedrooms in Japanese houses are the easiest to access and less private than the bedrooms in the other two cultures. Moreover, spatial configurations transform over time, even in the same culture. Therefore, considering these differences in the design phase using generative techniques is an important contribution to design quality, which keeps away from designing generic plans for different contexts and backgrounds.

Space syntax, though decades of research have proved its reliability to investigate the relationship between spatial design and people's behavior, Hillier et al. (1987) describe space syntax as "a set of techniques for the representation, quantification, and interpretation of spatial configuration in buildings and settlements". Since spatial configuration is one of the important indicators of architectural design, space syntax can be considered an important contribution to the investigation of the relationship between physical and social dimensions. Moreover, the representation of space with a topological network, a "justified graph," provides an understanding of the whole syntax's relation and motivation in the formation processes. Therefore, it is an important tool for reasonability and to quantitatively measure the spatial features that influence the users' behaviors and the activities held in the space. The ability of space syntax to numerically describe the spatial feature of each space in a configuration made it the best suitable tool to integrate with machine learning and therefore integrate it with generative design algorithms.

The aim of the paper is to bridge the gap between generative design and space quality in terms of privacy hierarchy and spatial relations through the development of a computational framework that evaluates the spatial features of each generated space to enhance design automation. The main concern of the study is not generating the spatial configuration or shape of the generated spaces, but the evaluation or classification of the suitable space function based on the spatial feature of the generated space based on specific contexts and cultures. Moreover, the strength of this study comes from the ability to integrate the trained neural network with any architectural plan-generating algorithm to achieve better spatial configuration quality and relationships. The limitation of this study is the lack of data needed to train the neural network model, which requires manual preparation and drawing of the architectural plans to perform Visibility Graph Analysis (VGA) and manual preparation of the Convex Map required to perform Convex Map analysis. Finally, when the neural network model is trained, it can be used as a separate tool to help architects in the early design stage to enhance their design in terms of spatial relations in the spatial configuration of the targeted culture, context, or typology, or it can be integrated with an architectural plan-generating algorithm that will be able to generate plans with consideration of spatial configuration quality.

2 Theoretical Framework

2.1 Space Syntax Method in Architectural Design

Space Syntax is a set of theories and techniques to analyze architectural space and urban form. Space syntax sees both buildings and cities as a configuration of parts and relationships as it is not affected by scale. It is worth mentioning that space syntax research and theories date to the 1970s and the 1980s, where it took different ramifications in different schools of thought. Some focus on shapes as the object of calculation, as is the case with "Shape Grammar", founded by the American mathematicians Stiny & Gips (1971). Others focus on the hierarchy and pattern of spatial arrangement, as Alexander et al. 1977. Space syntax as we know it today is a result of active development and research by Bill Hillier and Julienne Hansin at the UCL Bartlett School of Architecture. Bill Hillier Space Syntax differs from Stiny's and Gips's Shape Grammars by turning attention away from the strict mathematical rules related to shapes and geometry while emphasizing instead the topological relationship between the spaces in order to describe spatial relations independently from shape. While shape grammar focuses on shapes and geometry in order to analyze and generate architectural forms, space syntax focuses on analytical analysis of spatial configuration to understand social interaction and cultural meaning (Hillier & Hanson 1984).

Space Syntax analysis depends on a set of techniques to nalyse the spatial configuration of a city or architectural design to understand the social interaction and cultural meaning of its users. These techniques are Convex Map analysis, Visibility Graph Analysis, Axial analysis, and Segment analysis. Axial and Segment analysis are more suitable for urban-scale analysis. The Convex Map analysis is more suitable for building scale analysis, while VGA analyses are suitable for both urban and building scales (Yamu et al. 2021). Since the focus of the paper is on building scale, the concepts of both Convex Map analysis and VGA analysis are explained in detail.

Convex Map Analysis

In order to understand Convex Map analysis, we first have to understand the "Justified Graph". Justified Graph is based on the graph theory, which has a broad set of mathematical theories. However, space syntax is concerned with representing each space as a "node" and each connection between these spaces as a "line" connected to the corresponding nodes to analyze the spatial relationships and connectivity. In complex spatial relations, a drawing of the Justified Graph can make it easier to understand. By choosing a particular space or node from the spatial configuration as the root node (Depth 0) and then aligning the other nodes in levels based on the number of spaces one must pass through to reach the destination space (Hillier et al. 1987). As shown in Fig. 1. The Justified Graph could be drawn differently based on the root space selected. Convex Map is the first step of Justified Graph analysis by translating the architectural or urban plan into convex spaces. Convex space is a space where the placement of any two points within the space can always allow the connection between the points in a straight line without being intersected by a wall (Hillier & Hanson 1984, p 68). The preparation of Convex Map could sometimes be represented in different arrangements of the same spatial configuration. Therefore, none of the space syntax applications is able to automate the generation of Convex Map. However, a rule to follow is fewer and larger convex spaces in which the fewer prevail over the larger (Al_Sayed et al., 2014). The Justified Graph can help us retrieve two main spatial features: depth and choice, in addition to many other subsidiary features that can express cultural ideas and social relations.

- ***Depth Feature***

A space is at depth 1 if it is directly connected to the root or selected space; at depth 2 if it is necessary to pass through another space to reach the desired space; at depth 3 if we must pass through two spaces to reach the desired space; and so on (Hillier et al. 1987). As shown in Fig. 1.

- ***Choice Feature***

Simply it is the number of possible routes that one can chose to reach the desired space. As demonstrated in Fig. 2, where we have two different spatial relations of the same number of spaces with their Justified Graph drawn respectively. In Fig. 2a, we can see that to reach space *b* from space *c* we have two choices, either pass-through space *a* and then reach *b* or directly reach *b*, while in Fig. 2b in order to reach space *b* from space *c* we have only one choice which is pass-through space *a* and then *b*. when a spatial configuration has a number of spaces *K* and *K-1* links this means there is only one choice for all the spaces in the configuration and when a space lies on a ring, this means it has two or more choices based on the number of rings. This means spaces can be differentiated from each other based on whether they lay on a ring or not, the number of rings they lay on, and which ring they lay on (Hillier et al. 1987).

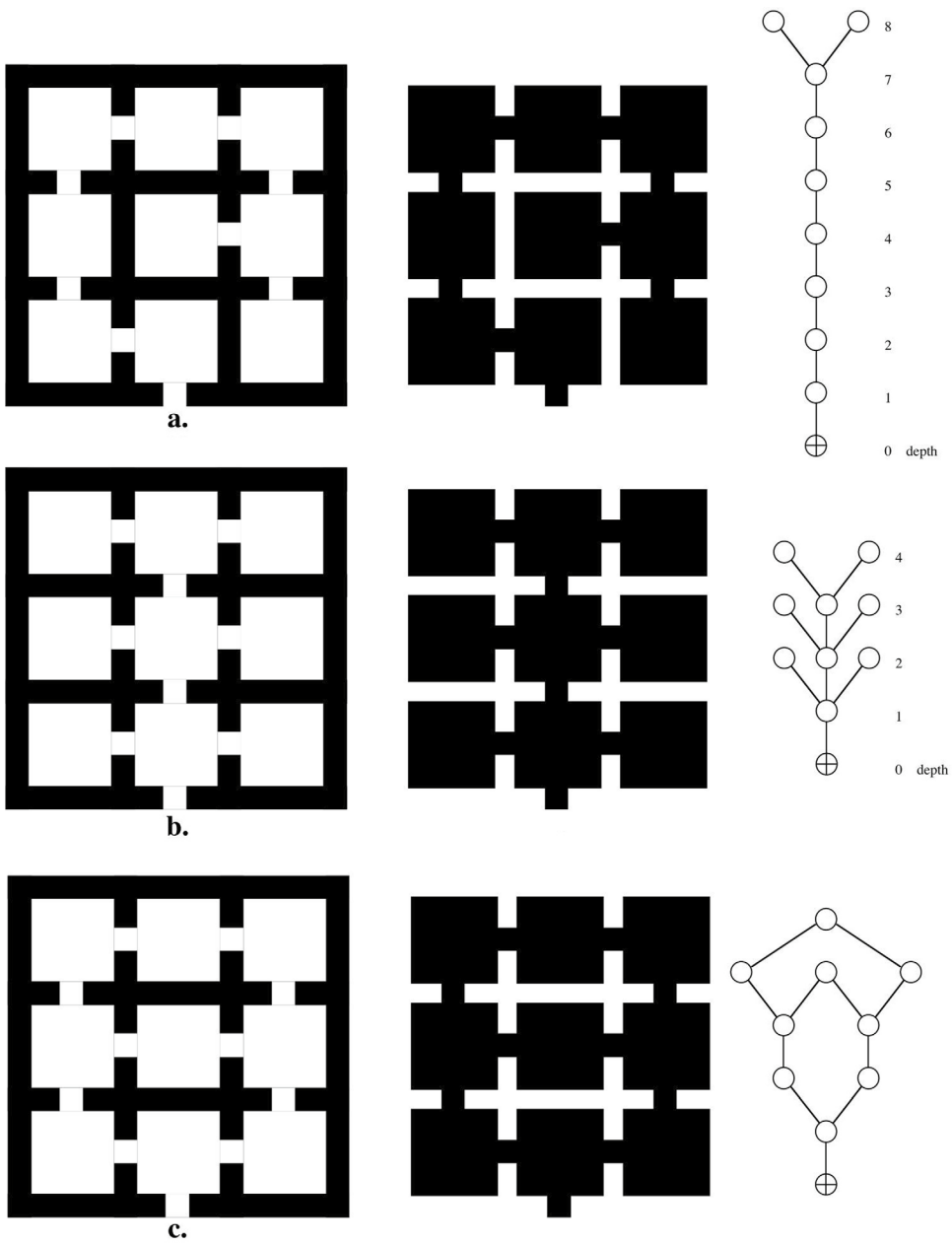


Fig. 1 A simple example to explain the hierarchical relations between spaces in space syntax *Source* (Hillier 1996)

- ***Integration [HH] Feature***

The integration [HH] feature is derived from relative symmetry *(RA),* which is derived from the depth feature as it represents the relative depth of one space from all other spaces in the configuration through the formula:

$$RA = \frac{2(MD - 1)}{k - 2} \tag{1}$$

where ***MD*** is the mean depth of the spaces from the selected space and ***k*** is the total number of spaces in the configuration. Then, by calculating the real relative asymmetry through the equation:

$$RRA = \frac{RA}{d - value} \tag{2}$$

where ***d-value*** is the ***RA*** value for the root space (the space at the bottom of the j-graph) of the system of ***k*** numbers of spaces. The higher the *RAA* value, the more segregated the space, and the lower the *RAA,* the higher the integration, then calculate the integration value through the equation (Hanson, 1998, p. 28):

$$Integration[HH] = \frac{1}{RAA} \tag{3}$$

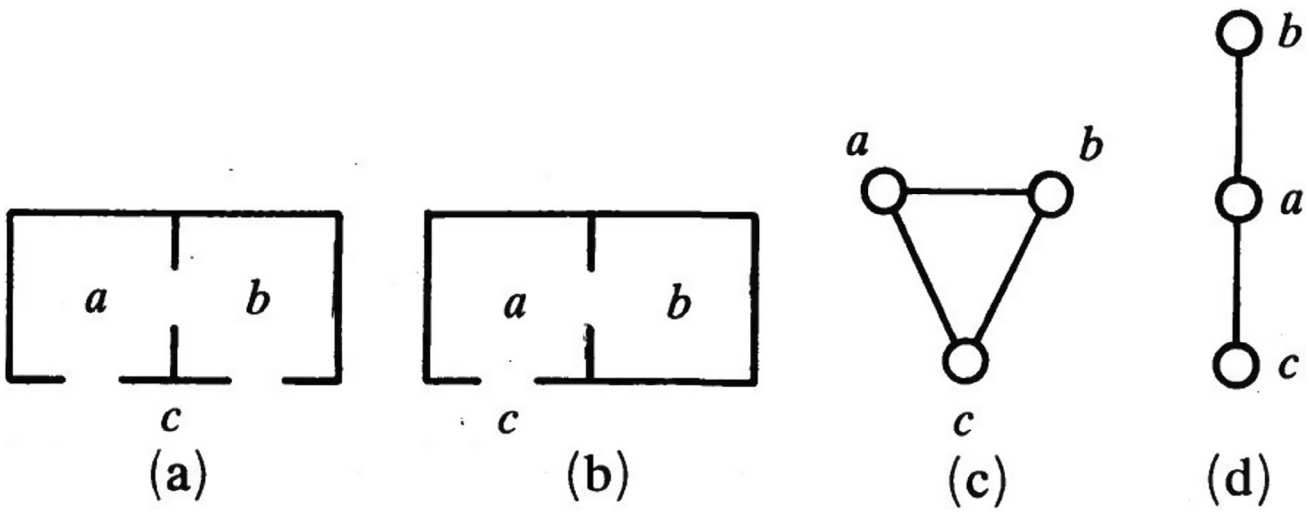


Fig. 2 **a**, **b** Two possible spatial relations of same number of spaces, spaces a and b to the outside, space c **c**, **d**, the corresponding justified graphs *Source* (Hillier et al. 1987)

The integration value will vary from space to space within a configuration, and that is one of the reasons different functions or activities are assigned to different rooms based on their integration or segregation. That is one element of how social relations and culture express themselves through space. Therefore, a numerical value can be assigned to different functions based on their spatial relations with other spaces, and if this numerical difference between different functions is in consistent order within a sample, then we can retrieve a cultural pattern from these spatial configurations, and that will lead us to the entropy feature (Hillier et al., 1987).

- ***Entropy Feature***

The entropy feature, or "different factor", is a calculation of the different values of integration of three or more spaces through the formula:

$$Entropy\ Value = H = -\sum \left[\frac{a}{t}\ln\left(\frac{a}{t}\right)\right] + \left[\frac{b}{t}\ln\left(\frac{b}{t}\right)\right] + \left[\frac{c}{t}\ln\left(\frac{c}{t}\right)\right] + \ldots \left[\frac{n}{t}\ln\left(\frac{n}{t}\right)\right] \quad (4)$$

where *a, b,* and *c* are the integration values of each space, and t is their sum. Then the value can be relativized between 0 and 2 through the equation:

$$Relativized\ Entropy = H^* = \frac{H - ln2}{ln3 - ln2} \quad (5)$$

The closer the value to 0, the more differentiated and structured the spaces; the closer the value to 1, the less differentiated and more homogenized the spaces to a point where all the spaces have equal integration (Hanson, 1998, p.30–31).

- ***Relative Asymmetry (RA)***

Relative asymmetry is a way to describe how deep or shallow a system is from a specific space (Hillier & Hanson, 1984, p.108–109). The calculation equation was mentioned earlier with the integration [HH] calculation.

- ***Step Depth***

Step depth is the number of spaces one has to pass through from a selected space (e.g., entrance) to reach the destination space (e.g., bedroom).

- ***Total Depth***

Total depth is the sum of the depths in the Justified Graph from any node or space to all others. A low value means the system is "Shallow" and the j-graph of it is bush-like, while a high value means the system is "deep" and the j-graph of it is tree-like. (Hillier & Hanson, 1984, p 108).

An explanation of the features used in this paper has been provided above. Although many other features and properties can be retrieved from Convex Map analysis or Justified Graph analysis.

Visibility Graph Analysis (VGA)

Visibility Graph Analysis is based on the concept of "Isovist", first introduced within architectural context by Benedikt (1979) as a way to analyze the spatial characteristics of an architectural environment. The Isovist works by placing a point in a plan that simulates the panoptic view of an observer with a horizontal viewing angle of 180 degree or 360 degree (Hillier 2001). It is a field of vision (Batty

2001). A viewshed polygon or Isovist field can be drawn, as shown in Fig. 3a, to retrieve various quantifiable properties of the physical environment, such as Isovist area, perimeter, length of open or closed edges, number of vertices, etc. These measures can be comparable, combined, and applied to different mathematical calculations to even get more integrative properties such as vertex density and roundness. Since Isovist describes the physical properties of space from only a single point (Turner et al. 2001) developed a technique to better describe the space from multiple observer points, which is Visibility Graph Analysis. By laying an equal-sized grid of points on the studied sample of architectural plan or urban then each point is connected to every point visible to it until we get a visibility graph as in Fig. 3b. Other calculation can then be applied on the visibility graph to retrieve more visual spatial features such as "Isovist Area" of all the points, "visual integration" which is a calculation of the number of visual steps required to get from one point to any other point in the studied sample, "Visual Step Depth" is the number of visual turns plus one required to get from one point to any other point, "Metric Step Shortest-Path Length" is a calculation of the shortest path from one point to any other point by calculating the metric distance instead of the number of visual turns, "Visual Mean Depth" is a very similar to the calculation of step depth by calculating the shortest path of fewest number of visual turns from one point to any other point in the visibility graph and then summing the number of visual turns and divide the value by the number of nodes/points minus the selected node, "Visual Integration [HH]" is simply a normalized version of mean depth introduced by (Hillier et al. 1993) by dividing with a number called the d-value (Turner 2004). And many other features. VGA can be applied at human eye level—what people can see—and at human knee level—where people can move (Turner et al. 2001).

To calculate all these features, many applications have been developed, such as Axwoman, Isovists and DepthmapX. Axworman is a plug-in for ArcView that focuses on the analysis of urban morphology using Axial Map analysis based on space syntax, with the intention of integrating ArcView GIS with space syntax (Jiang 2015). Isovists_App is free software focusing on Isovist and Visibility Graph analysis only. It can provide multiple spatial feature results, both numerically and visually, and can be applied on different scales, from a single room to an urban street (Isovists.Org). DepthmapX is open-source software founded by Alasdair Turner and further developed by Tasos Varoudis from the space syntax laboratory. DepthmapX includes all space syntax analysis techniques. Axial and Segment Map Analysis, Convex Map Analysis, Visibility Graph Analysis, and Agent-Based Analysis, it can be operated on both building and urban scale (DepthmapX development team 2017). DepthmapX has been chosen for both VGA and Convex Map analysis in this paper.

2.2 Machine Learning

Traditionally, in order to recognize patterns and relationships, researchers have had to conduct observation and surveys. When technologies improved and data started to increase, scientists and researchers tended to use explicit programming and analytical models. Recently, with the advancement in Artificial Intelligence, we were able to build machine learning models that can perform certain tasks without being explicitly programmed by humans or recognize implicit patterns and relationships when dealing with large and complex data where traditional analytical methods are unable to recognize these patterns. Machine learning training process differs from genetic algorithms and simulated annealing as it does not require a clearly defined objective function *f(x),* while it relies on examples

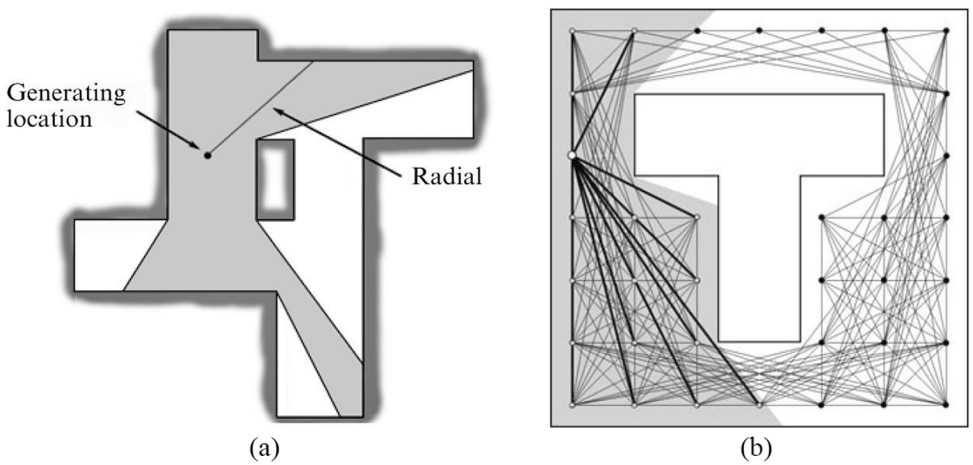


Fig. 3 **a** An Isovist polygon represents the visible area from a specific location/point. **b** An example of a first-order visibility graph, showing the pattern of connections for a simple configuration *Source*: (Turner et al. 2001)

of training input and output data for its learning algorithm in regression and classification tasks, as we will explain in detail in ML learning approaches later in this paper. However, genetic algorithms has many promising applications in the Architecture, Engineering, and Construction fields, for example, optimizing the environmental performance of buildings in terms of thermal and lighting (Caldas & Norford 2002), evaluating a semi-automatically generated design (Barczik & Kruse 2016), solving the distribution of facility layout in hospital interior design (Yeh 2006), optimal floor plan generation (Zheng & Ren 2020), and optimization of structure to achieve high performance with lightweight structure and low cost (Sonmez 2008).

Learning Approaches

In general, machine learning has four main learning approaches: **Supervised learning**, **Unsupervised learning, Semi-supervised learning**, and **Reinforcement learning (RL)**. All these approaches rely on data for the learning process; however, they differ based on what and how these data are being processed.

- ***Supervised Learning***

Supervised learning is a method where a model is trained with known data and then provided with new data to predict results. It can be divided into four methods: Decision Trees, Bayesian Networks, Support Vector Machines (SVMs), and Artificial Neural Networks. Artificial neural networks multiply input features with different weights to find relationships between inputs and outputs (McCulloch & Pitts 1943). When the input features and output labels are fed to the model the algorithm runs to train the model using Back Propagation (Werbos 1974) or other learning algorithms such as Resilient Back Propagation, Evolutionary, and Levenberg Marquardt. All methods can be used for regression and classification tasks, with SVMs and neural networks performing better with big data and nonlinear features (Kotsiantis 2007).

- ***Unsupervised Learning***

Unsupervised learning is self-learning, using algorithms to find patterns in abelling data through clustering (K-Mean Clustering, K-NN Clustering, and Hierarchical Clustering) or associating (Apriori algorithm, FP-Growth algorithm). This information is extracted without human effort and can be further processed for correlations, simplifying large datasets, and cleansing unwanted features.

- ***Semi-supervised Learning***

Semi-supervised learning is a combination of supervised and unsupervised learning used in problem-solving or predicting values when abellin data is scarce or expensive. It involves using abellin data to train a supervised model, clustering abelling data, and abelling clustered data. The process repeats multiple times until convergence is reached, proving useful in web, video mining, and document classification.

- ***Reinforcement Learning (RL)***

Reinforcement learning is a unique learning approach that uses sequential data based on environment changes to optimize agents' actions. The "Policy Network" transforms input into output actions, with agents randomly interacting with the environment. Performance is rewarded for good performance, and bad performance is bunched until the model optimizes tasks for more rewards. Some usage examples of RL are learning robots how to perform complex manipulation tasks that are challenging or impossible through explicit programming (Clark 2015) and beating world champions at the Go game (Silver & Hassabis 2016).

3 Materials and Methods

This study develops a computational framework to evaluate the convenience of the spatial configuration of the generated architectural plans by training a supervised neural network on some spatial feature values of three Syrian houses. These values have been gained by analyzing these houses using the DepthmapX software, which is based on the Space Syntax theory. The developed framework is represented in Fig. 4.

The spatial feature values are variables that change from one space to another based on the spatial configuration, as explained in Convex Map analysis; in other words, each space has its own spatial feature values, and these values determine its suitable function. By finding the spatial feature pattern of each space function of a specific culture, an evaluation can be performed to determine if the designed space is suitable for that function. Instead of relying only on the difference factor value as in Hillier et al. (1987) research, six spatial features of VGA—Isovist Area, Visual Integration [HH], Metric Step Shortest-Path Length, Visual Mean Depth, Visual Entropy, Visual Control—and five features of Convex Map analysis—Integration [HH], Entropy, RA, Step Depth, and Total Depth—has been selected for

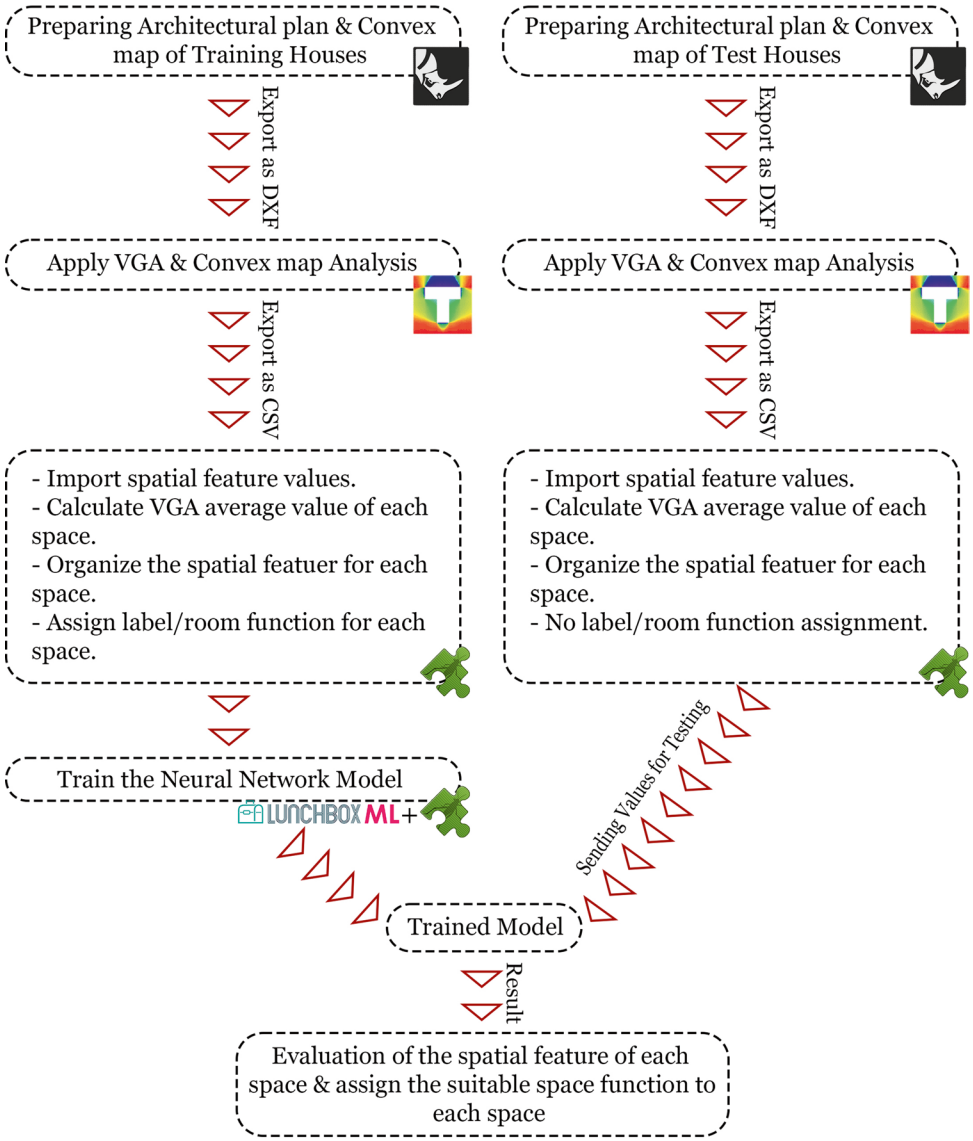


Fig. 4 The developed framework
Source (Author)

each space. The features and their values are represented in Table 1.

In order to find these spatial feature patterns, a supervised machine learning neural network model has been trained on the spatial features of three houses (post-independent from the French colonization period) in Aleppo, Syria. The houses have been manually drawn and prepared in Rhinoceros CAD software to apply both Visibility Graph Analysis, and Convex Map Analysis in DepthmapX software based on space syntax theory. The visual representation of the VGA is shown in Figs. 5, 6. The spatial feature has been exported from DepthmapX into a CSV file. Then the CSV file is imported into Grasshopper, a visual programming environment, to calculate the average value of the VGA features for each space, then organize and label the spatial features of each room function correspondingly in order to provide it to train the neural network as shown in Fig. 7. After the model has been trained, another home of the same culture and period has been prepared and analyzed for testing the trained model. The spatial features of the testing home are shown in Table 2.

The training and testing of the artificial neural network are accomplished in Grasshopper by using "Lunchbox ML" plug-in for Grasshopper developed by Proving Ground (Anaraki & Miller, 2017) using Accord.NET framework (Souza 2012). As shown in Fig. 8.

4 Results of the Analysis and Discussion

Homes 1, 2, and *4* are located within the "Ottoman Waqf" building in "Al Jamelaiah" area, which is considered one of the first residential expansion outside the old city wall, it has a few different design structures from traditional

Table 1 The selected spatial features and their values of Space Syntax Analysis, used for training the neural network model *Source* (Author)

Case	Room name	Isovist area	Visual integration [HH]	Metric step Shortest-Path length	Visual mean depth	Visual entropy	Visual control	Room area	Integration [HH]	Entropy	RA	Step depth	Total depth
Home 1	Entrance	15.817	7.821	1.685	2.043	1.052	0.617	5.380	0.885	1.303	0.333	0	25
	Hall	46.243	10.432	6.726	1.796	1.289	1.096	22.870	2.212	1.008	0.133	1	16
	Circulation (Lobby)	18.000	7.322	6.248	2.152	1.248	0.914	3.106	1.206	1.553	0.244	2	21
	Circulation (Corridor)	44.652	10.909	8.140	1.757	1.148	1.465	3.398	1.896	1.475	0.156	2	17
	Bath (W.C)	5.594	5.385	8.180	2.545	1.266	0.566	1.665	0.664	1.804	0.444	3	30
	Bath	12.004	7.156	9.918	2.170	1.048	0.757	5.260	0.829	1.804	0.356	3	26
	Kitchen	12.565	5.896	8.309	2.453	1.290	0.900	8.911	0.664	1.804	0.444	3	30
	Dining	41.932	8.857	10.420	1.956	1.429	0.858	21.003	0.737	1.726	0.400	3	28
	Living/Reception	58.320	11.411	8.604	1.720	1.365	1.175	33.199	1.474	1.475	0.200	2	19
	Bed (1)	26.388	7.598	11.884	2.091	1.428	0.975	22.344	0.829	1.804	0.356	3	26
	Bed (2)	23.334	7.372	11.323	2.130	1.351	0.969	18.925	0.829	1.804	0.356	3	26
Home 2	Entrance	32.931	8.660	1.765	1.963	1.321	1.129	9.619	1.474	1.553	0.200	0	19
	Hall	30.978	7.032	6.856	2.185	1.627	1.037	25.423	1.327	1.371	0.222	1	20
	Circulation (Lobby)	13.104	4.844	5.294	2.773	1.794	0.971	5.964	0.948	1.848	0.311	1	24
	Circulation (Corridor)	37.437	7.665	8.288	2.123	1.714	1.447	3.204	1.021	1.804	0.289	2	23
	Bath (W.C)	6.265	3.837	8.799	3.120	1.874	0.650	1.485	0.577	2.099	0.511	2	33
	Bath	10.763	5.784	10.076	2.446	1.566	0.709	5.115	0.603	2.099	0.489	3	32
	Kitchen	14.762	4.413	7.334	2.916	1.892	1.011	12.523	0.577	2.099	0.511	2	33
	Dining	24.024	5.631	5.695	2.449	1.714	0.967	20.168	0.737	1.848	0.400	1	28
	Living/Reception	27.777	6.690	4.851	2.300	1.514	1.004	22.184	0.737	1.848	0.400	1	28
	Bed (1)	24.181	5.171	12.322	2.628	2.045	0.989	19.946	0.603	2.099	0.489	3	32
	Bed (2)	23.916	5.591	11.157	2.467	1.941	0.973	19.444	0.603	2.099	0.489	3	32

(continued)

Table 1 (continued)

Case	Room name	Isovist area	Visual integration [HH]	Metric step Shortest-Path length	Visual mean depth	Visual entropy	Visual control	Room area	Integration [HH]	Entropy	RA	Step depth	Total depth
Home 3	Entrance	39.918	8.040	1.245	2.111	1.522	0.942	3.350	1.299	1.596	0.212	0	26
	Hall	55.341	7.310	5.870	2.218	1.826	1.155	40.020	0.866	2.003	0.318	1	33
	Circulation (Corridor)	46.028	8.968	4.394	1.952	1.389	1.309	10.200	2.021	1.491	0.136	1	21
	Circulation (Lobby)	19.275	6.211	7.979	2.380	1.589	0.882	2.580	1.299	1.703	0.212	2	26
	Bath (W.C)	10.099	5.361	3.101	2.600	1.630	0.575	3.848	0.909	1.791	0.303	2	32
	Bath (1)	12.293	5.344	6.798	2.599	1.599	0.818	8.168	0.909	1.791	0.303	2	32
	Bath (2)	10.260	3.564	11.712	3.389	1.838	0.825	5.120	0.535	2.215	0.515	4	46
	Kitchen	25.349	5.245	7.189	2.623	1.867	0.993	23.053	0.909	1.791	0.303	2	32
	Dining	35.869	7.658	8.707	2.141	1.499	0.813	14.400	0.568	2.215	0.485	2	44
	Living/Reception	35.502	4.804	9.901	2.828	2.139	0.899	22.160	0.568	2.215	0.485	2	44
	Bed (M)	17.934	4.814	11.707	2.832	1.853	1.038	13.785	0.791	2.061	0.348	3	35
	Bed (1)	22.608	4.366	12.211	2.987	1.954	1.004	21.533	0.727	1.915	0.379	3	37
	Bed (2)	30.406	6.377	10.217	2.385	1.584	0.978	23.213	0.909	1.791	0.303	2	32

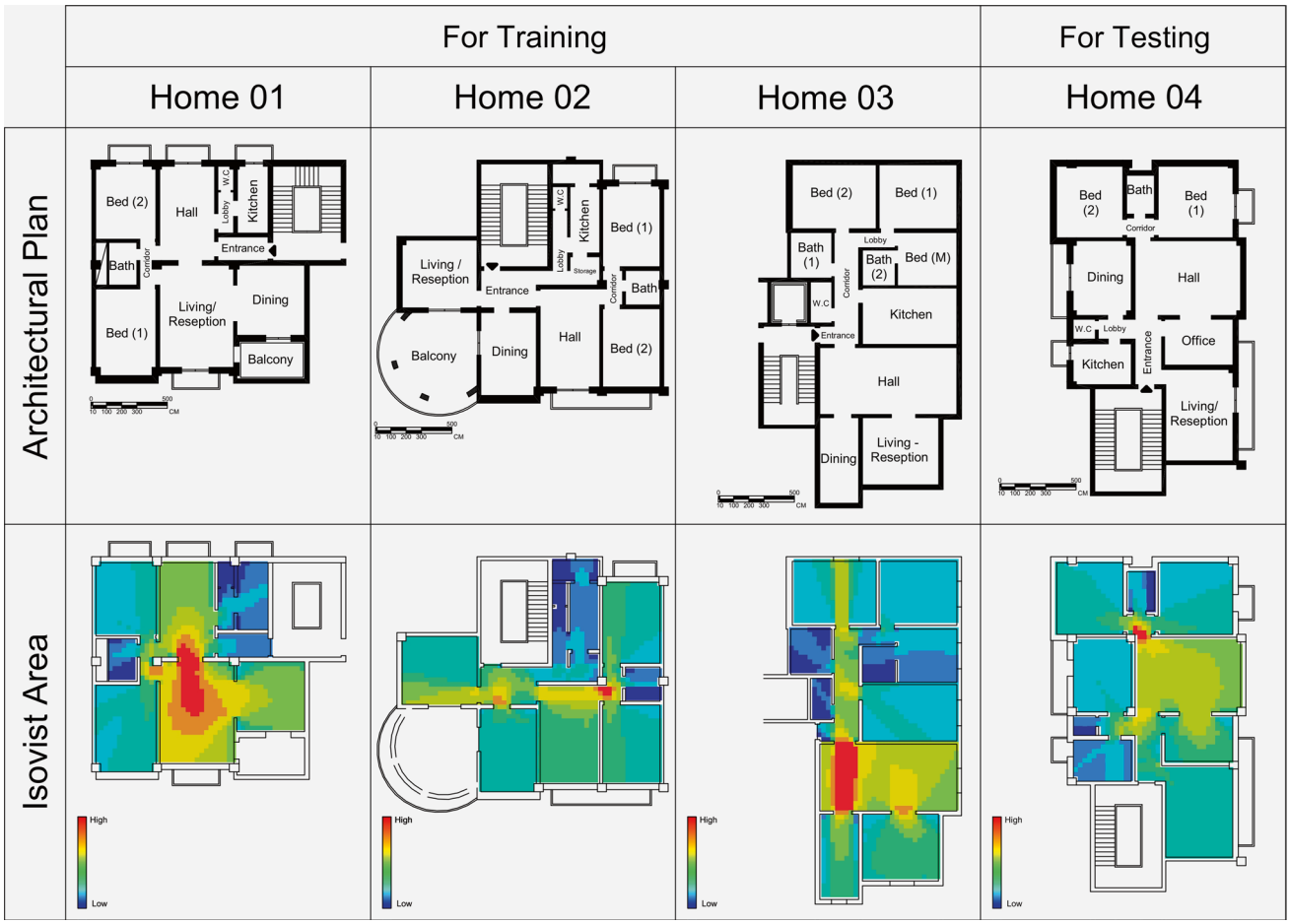


Fig. 5 Visual representation of VGA analysis for the home's cases used for training and testing the neural network *Source* (Author)—Continued in Fig. 6

housing as it represents an important stage of the evolution in design directly after the independence from French colonization to create a modern residential design without copying the western style. The façade design is inspired by Arabic Islamic architecture. The integration of commercial and residential functions in the same building has been represented for the first time in this building (Commercial in ground floor, Residential in typical floors), the ground floor included a theater of artists syndicate to hold cultural events, which was a rare case in Aleppo (Kandakji 2013). Figure 9.

While *home 3* is located in "Al Sabil (Faysal Street)" area built in the 1960s, characterized by a clear separation between day and night activity in addition to a comfortable and easy movement path to reach all the spaces within the home (Kandakji 2013). Figure 10.

From the space syntax analysis results shown in Table 1, a pattern can be recognized between the space function and spatial features. For instance, the spatial feature of the bedrooms in the three homes has a high value of step depth between 3 and 4, and a high metric step length from the entrance compared to the other spaces. Although the bathroom's spatial features are relatively similar to those of a bedroom in terms of step depth and metric step length, but the Isovist area is significantly lower than that of bedrooms. Such spatial patterns, and definitely more complex ones that are difficult to recognize by humans, are recognizable by deep neural networks. After the process of training the neural network model on the spatial features of the labeled rooms, another home has been prepared with its spatial features extracted and fed to the trained model for testing.

The testing results were fascinating, as the trained model was able to classify most of the spaces according to their suitable functions as they were used by the user. This means that the spatial features of these spaces are suitable to hold the corresponding activity or function based on the values used to train the model. The spaces that have been predicted successfully are the entrance, hall, circulation areas, bathrooms, kitchen, and bedrooms. However, the dining room, on the other hand, has been predicted to be either a hall, living room, or dining room. Similarly, the predicted function of the living room is either hall or living room. The reason

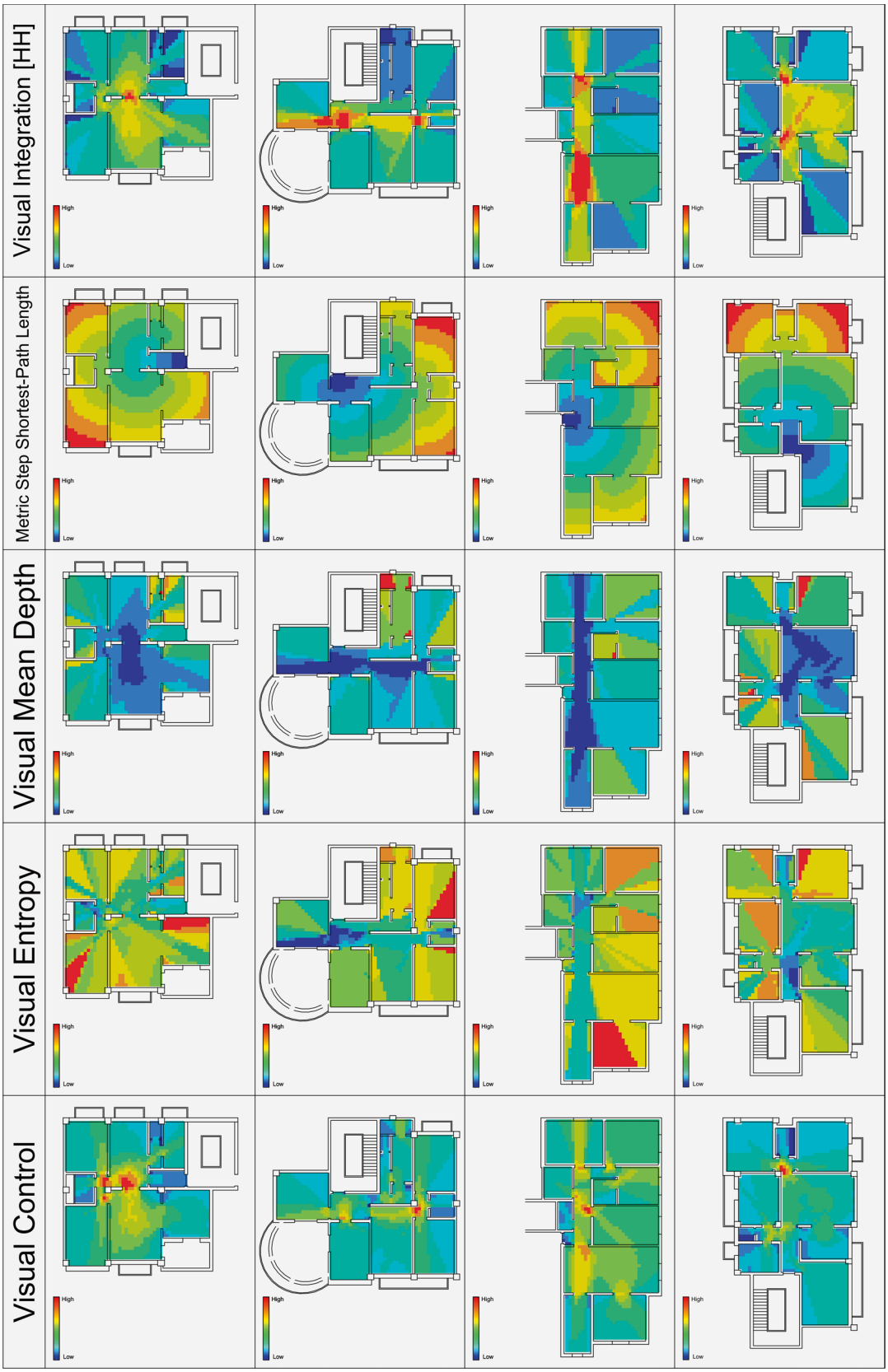


Fig. 6 Continue of Fig. 5 *Source* (Author)

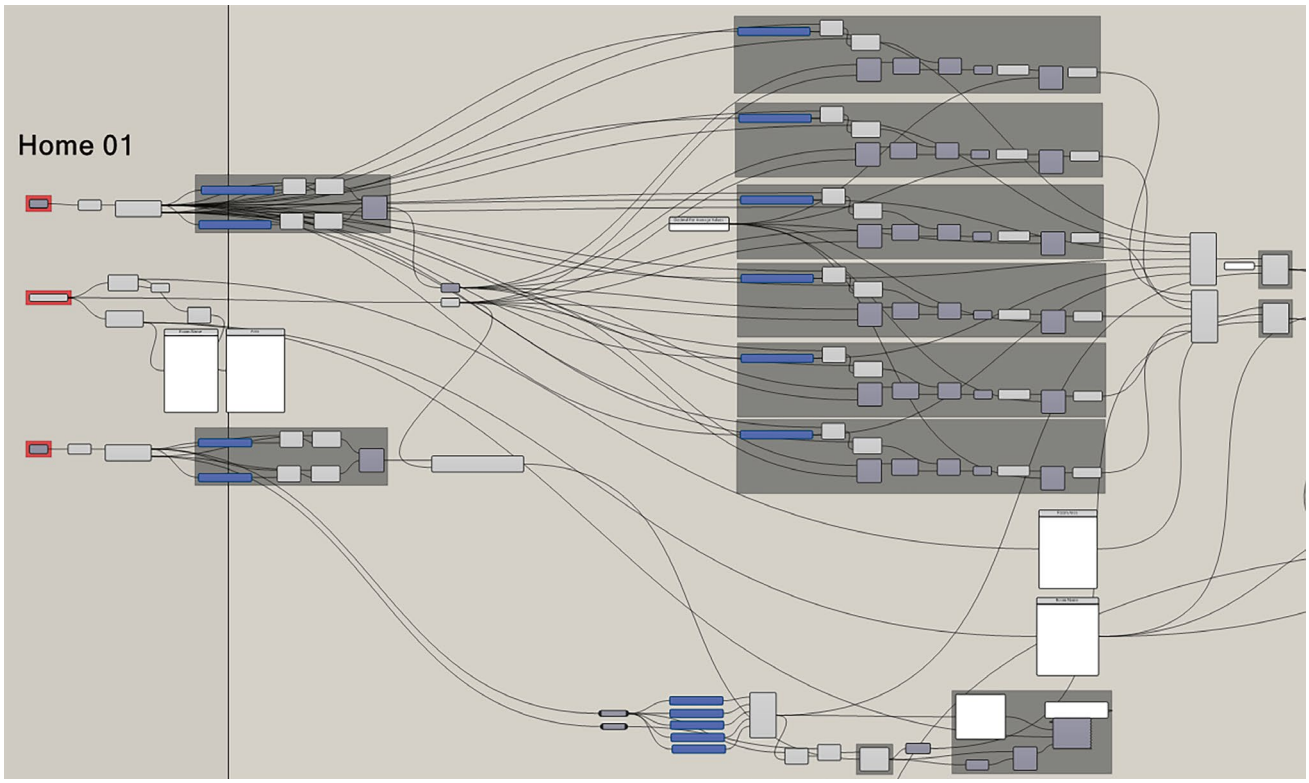


Fig. 7 Grasshopper definition for importing the spatial feature, calculate average VGA values, and organize the values for corresponding space. *Source* (Author)

Table 2 The selected spatial features and their values of Space Syntax Analysis, used for testing the training model (Room Name was not provided) *Source* (Author)

Room name	Isovist area	Visual integration [HH]	Metric step Shortest-Path length	Visual mean depth	Visual entropy	Visual control	Room area	Integration [HH]	Entropy	RA	Step depth	Total depth
Entrance	37.772	9.604	2.173	1.874	1.125	1.036	8.262	1.327	1.553	0.222	0	20
Hall	49.586	9.846	7.706	1.845	1.352	1.131	34.286	1.206	1.475	0.244	1	21
Circulation (Lobby)	36.308	7.645	4.463	2.092	1.552	1.301	3.269	1.021	1.804	0.289	1	23
Circulation (Corridor)	48.837	9.206	10.201	1.934	1.485	1.507	2.830	0.948	1.804	0.311	2	24
Bath (W.C)	8.814	5.566	6.491	2.561	1.423	0.581	1.945	0.603	2.099	0.489	2	32
Bath	15.764	6.938	12.358	2.244	1.285	0.677	4.720	0.577	2.099	0.511	3	33
Kitchen	15.691	5.481	6.366	2.556	1.718	0.882	10.340	0.603	2.099	0.489	2	32
Dining	22.466	5.706	7.571	2.476	1.804	0.978	18.442	0.603	2.099	0.489	2	32
Living/ Reception	28.427	5.741	4.035	2.481	1.632	0.994	26.475	0.698	1.804	0.422	1	29
Bed (1)	26.930	5.770	13.643	2.461	1.876	0.976	23.131	0.577	2.099	0.511	3	33
Bed (2)	26.958	6.425	13.599	2.317	1.707	0.948	20.509	0.577	2.099	0.511	3	33

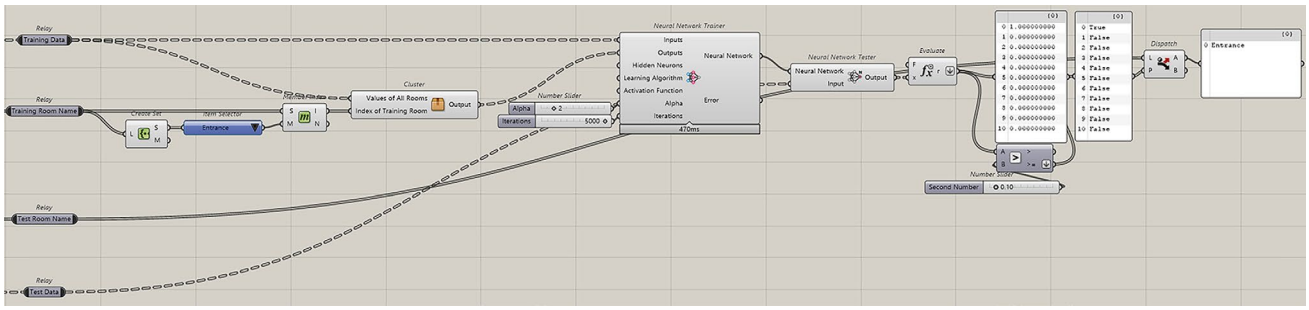


Fig. 8 Grasshopper definition for training and testing the neural network model for entrance space *Source* (Author)

Fig. 9 Ottoman Waqf building typical floor plan, represent homes 1, 2, and 4 *Source* (Kandakji 2013)

for this failure can be explained by the fact that in Syrian culture, the living room and hall functions are interchangeable, especially in the case of guest reception, where male and female guests are separated in either room.

Although the weight or influence of each spatial feature on the decision-making of the neural network to determine the space function is still ambiguous, which is called the "Black Box". Lots of research has been conducted to make an interpretable neural network model in order to understand the influence of each feature on determining the result (e.g., Chen et al. 2018; Ming et al. 2019; Zhang et al. 2019; Fan et al. 2020). Therefore, when an interpretable neural network model is developed, an understanding of the influence of each spatial feature on specifying the space function can be achieved.

5 Conclusion

The research study presented in this paper aims to improve the spatial quality of the living spaces generated by algorithms to ensure the comfort of the users when they interact with these spaces. Therefore, spatial feature data has been gathered by analyzing three houses belonging to the same context, culture, and period of time to train a deep neural network to find the spatial feature pattern that is correlated to room function. The developed algorithm can be integrated into any generative design algorithm to evaluate the spatial quality in terms of conformity and convenience of the generated plan to a specific culture or context, or it can be used as a separate tool to help architects enhance the spatial quality of their architectural design.

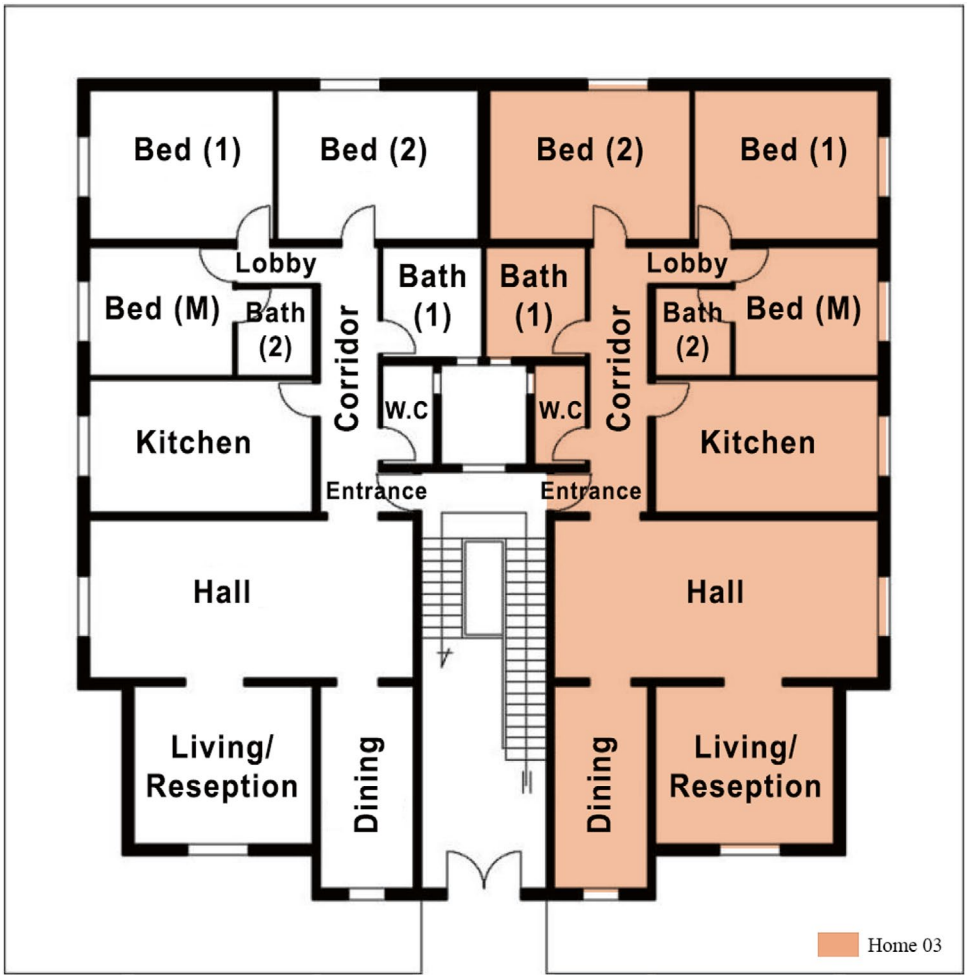


Fig. 10 Al Sabil (Faysal street) building, represent home 3 *Source* (Kandakji 2013)

Although the trained neural network model has been able to successfully recognize most of the suitable functions based on their spatial features, only by training the model on three houses, it can be indicated that there is a need to train the model on a greater number of houses to make it more reliable and accurate. A rule of thumb for the number of samples is ten times the number of weights, according to Miotto et al. (2017). Which means hundreds of houses need to be analyzed and fed to train the model. This is difficult and time-consuming considering that the houses must be drawn and prepared manually for VGA and Convex Map analysis. Until now, there has not been a reliable automation method to convert architectural plans from images to vector lines to be readable by CAD software, which has been a limitation for the study. Since there is a need to automate the process of generating Convex Maps for Convex Map analysis by space syntax applications in order to make the progress more reliable, the authors aim to develop an algorithm in future works that converts architectural plans from images to vector lines and an algorithm to generate Convex Maps automatically.

It is worth emphasizing that the strength of this study comes from the integration of space syntax and ML by developing an architectural plan evaluating algorithms to achieve better spatial configuration quality and relationships. This developed algorithm has pointed the way towards a number of areas for future investigations, such as how other spatial features such as Connectivity, Isovist Compactness, Visual Clustering Coefficient, Choice, Intensity, and many others spatial features could contribute to better evaluation or classification of space function, how the number of spatial features used to train the model could be decreased to only use the features that actually determine the space function, and finally, how the developed algorithm could be integrated in evaluating different architectural plans such as offices, malls, or even urban planning.

Acknowledgement This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alexander C, Ishikawa S, Silverstein M (1977) A pattern language: towns, buildings. Oxf Univ Press, Construction

Al_Sayed K, Turner A, Hillier B, Iida S, Penn A (2014) Space syntax methodology. Bartlett Sch Arch, UCL, London, 4th Edition.

Anaraki NAT, Miller N (2017) LunchBoxML. https://bitbucket.org/archinate/lunchboxml/src/master/

Barczik G, Kruse R (2016) Shifting Design Work from Production to Evaluation—An Evolutive Design Tool. Proceedings of the International Conference on Education and Research in Computer Aided Architectural Design in Europe, 2, 109–115. https://doi.org/10.52842/CONF.ECAADE.2016.2.109

Batty M (2001) Exploring Isovist Fields: Space and Shape in Architectural and Urban Morphology. Environment and Planning B 28(1):123–150. https://doi.org/10.1068/B2725

Benedikt ML (1979) To take hold of space: Isovists and Isovist fields. Environ Plann B Plann Des 6(1):47–65. https://doi.org/10.1068/B060047

Caldas L (2008) Generation of energy-efficient architecture solutions applying GENE_ARCH: An evolution-based generative design system. Adv Eng Inform 22(1):59–70. https://doi.org/10.1016/J.AEI.2007.08.012

Caldas LG, Norford LK (2002) A design optimization tool based on a genetic algorithm. Autom Constr 11(2):173–184. https://doi.org/10.1016/S0926-5805(00)00096-0

Chen C, Li O, Tao C, Barnett AJ, Su J, Rudin C (2018) This looks like that: deep learning for interpretable image recognition. Advances in Neural Information Processing Systems, 32. https://doi.org/10.48550/arxiv.1806.10574

Clark J (2015). Berkeley's Preschool for Robots. Bloomberg. https://www.bloomberg.com/features/2015-preschool-for-robots/

DepthmapX development team (2017) DepthmapX (0.6.0). pklampros. https://github.com/SpaceGroupUCL/depthmapX/

Dinçer AE, Çağdaş G, Tong H (2014) A digital tool for customized mass housing design. Proceedings of the 32nd International Conference on Education and Research in Computer Aided Architectural Design in Europe (ECAADe), 1, 201–211. https://doi.org/10.52842/CONF.ECAADE.2014.1.201

Fan F, Xiong J, Li M, Wang, G (2020) On interpretability of artificial neural networks: A Survey. https://doi.org/10.48550/arxiv.2001.02522

Hanson J (1998) Decoding Homes and Houses. Cambridge University Press, In Decoding Homes and Houses. https://doi.org/10.1017/CBO9780511518294

Hillier B (1996) Space Is the machine. Cambridge University Press. https://discovery.ucl.ac.uk/id/eprint/3881

Hillier B (2001) A theory of the city as object: or how spatial laws mediate the social construction of urban space. 3rd International Space Syntax Symposium. https://doi.org/10.1057/PALGRAVE.UDI.9000082

Hillier B, Hanson J (1984) The Social Logic of Space. Cambridge University Press. https://doi.org/10.1017/CBO9780511597237

Hillier B, Hanson J, Graham H (1987) Ideas are in things: an application of the space syntax method to discovering house genotypes. Environ Plann B Plann Des 14(4):363–385. https://doi.org/10.1068/B140363

Hillier B, Penn A, Hanson J, Grajewski T, Xu J (1993) Natural movement: or configuration and attraction in urban pedestrian movement. Environ Plann B Plann Des 20(1):29–66. https://doi.org/10.1068/B200029

Holland JH (1975) Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Michigan University Press

Isovists.org. (n.d.). Retrieved July 23, 2022, from https://isovists.org/

Jiang B (2015) Axwoman 6.3: An ArcGIS extension for urban morphological analysis. http://giscience.hig.se/binjiang/Axwoman/

Kandakji L (2013) Design transformations of residential architecture In the syrian cities since independence till now [Doctoral Dissertation, University of Aleppo]. http://mohe.gov.sy/Masters/Message/PH/lila%20kantakji.pdf

Klemmt C, Bollinger K (2017) Angiogenesis as a model for the generation of load-bearing networks. Int J Archit Comput 15(1):18–37. https://doi.org/10.1177/1478077117691599

Kotsiantis S (2007) Supervised machine learning: a review of classification techniques. Informatica, 31(3), 249–268. https://informatica.si/index.php/informatica/article/viewFile/148/140

Lindenmayer A (1968) Mathematical models for cellular interactions in development I. Filaments with one-sided inputs. J Theor Biol, *18*(3), 280–299. https://doi.org/10.1016/0022-5193(68)90079-9

McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133. https://doi.org/10.1007/BF02478259

Ming Y, Qu H, Xu P, Ren L (2019) Interpretable and steerable sequence learning via prototypes. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 903–913. https://doi.org/10.1145/3292500.3330908

Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2017) Deep learning for healthcare: Review, Opportunities, and Challenges. Brief Bioinform 19(6):1236–1246. https://doi.org/10.1093/bib/bbx044

Silver D, Hassabis D (2016) AlphaGo: Mastering the ancient game of Go with Machine Learning. Google AI Blog. https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html

Sonmez FO (2008) Structural optimization using simulated annealing. Simulated Annealing, 281–306. https://doi.org/10.5772/5567

Souza CR (2012) Accord.Net framework. http://accord-framework.net/index.html

Stiny G, Gips J (1971) Shape grammars and the generative specification of painting and sculpture. https://www.researchgate.net/publication/221329330

Stiny G, Mitchell WJ (1978) The palladian grammar. Environ Plann B Plann Des 5(1):5–18. https://doi.org/10.1068/B050005

Turner A (2004) Depthmap 4: a Researcher's Handbook. Bartlett Sch Grad Stud, UCL. https://discovery.ucl.ac.uk/id/eprint/2651/

Turner A, Doxa M, O'Sullivan D, Penn A (2001) From isovists to visibility graphs: A methodology for the analysis of architectural space. Environ Plann B Plann Des 28(1):103–121. https://doi.org/10.1068/B2684

Uyar G, Griffiths S (2017) A configurational approach to vernacular domestic architecture: "Traditional" Houses in Turkey, Japan, and Britain. In: Proceedings of the 11th Space Syntax Symposium, 32.1–32.16. https://discovery.ucl.ac.uk/id/eprint/10033842

Wallgren P (2020) Finch. https://finch3d.com/

Werbos P (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences (Ph. D. dissertation). Harv Univ. https://www.researchgate.net/publication/279233597_Beyond_Regression_New_Tools_for_Prediction_and_Analysis_in_the_Behavioral_Science_Thesis_Ph_D_Appl_Math_Harvard_University?utm_source=twitter&rgutm_meta1=eHNsLTUxY-WdKdzZYcno5aFZuTTQzUTBnZlBzclhyemNWOGhyeGtxcTlIK2VoNGFLZ2pJcmZITVpaa1dmWTVSQ0FaVkZoejBXQXRkbWR0MVg1YXJTYzJKZzAzdUk%3D

Wolfram S (2002) A New Kind of Science. In arXiv preprint quant-ph/0206089 (Issue 0). Wolfram Media. https://www.wolframscience.com/nks/

Wortmann T (2017) Model-based optimization for architectural design: Optimizing daylight and glare in grasshopper. Technology|Architecture + Design, 1(2), 176–185. https://doi.org/10.1080/24751448.2017.1354615

Yamu C, van Nes A, Garau C (2021) Bill hillier's legacy: Space syntax—a synopsis of basic concepts, measures, and empirical application. Sustainability (Switzerland), 13(6). https://doi.org/10.3390/su13063394

Yeh IC (2006) Architectural layout optimization using annealed neural network. Autom Constr 15(4):531–539. https://doi.org/10.1016/J.AUTCON.2005.07.002

Zhang Y, Liu N, Ji S, Caverlee J, Hu X (2019) An interpretable neural model with interactive stepwise influence. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11441 LNAI, 528–540. https://doi.org/10.1007/978-3-030-16142-2_41/COVER

Zheng H, Ren Y (2020) Architectural layout design through simulated annealing algorithm. RE: Anthropocene, Design in the Age of Humans—Proceedings of the 25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA, *1*, 275–284. https://doi.org/10.52842/CONF.CAADRIA.2020.1.275

Machine Learning-Based QSAR Classifications for PIM Kinases Inhibition Prediction: Towards the Neoplastic in Silico Drug Design

Mohamed Oussama Mousser, Khairedine Kraim, Fouad Chafaa, and Mohamed Brahimi

Abstract

Promoting the use of strong AI tools in computational drug designing is a promising way to avoid early-stage failures of cancer drug discovery process. We build an inhibition targeted machine learning classifications, aiming to model the structure/activity relationships for PIM 1/2/3 protein kinases inhibitors, using different decision trees-based algorithms, starting from the data curation and analysis of previous experimental measurements. The therapeutic targets being studied are a family of serine/threonine protein kinases directly involved in various cellular processes, they have been implicated in cancer progression and identified as highly oncogenic. The constructed models showed Random Forest (RF) performances slightly better than XGBoost for the PIM 1 (+1% of difference in the accuracy scores), and XGBoost significant robustness for the PIM 2 and 3 datasets (+2% and + 4%, respectively), whereas the SVM algorithms were found to present a poor predictive ability from our datasets, either with a linear or a radial basis functional kernel. The benchmarking led to the selection of the strongest models: 85% of prediction accuracy for PIM 1 and PIM 2 datasets and 82% for the PIM 3 dataset. Data modeling along with technical methodology are discussed in details and the predictive strength of both RF and XGBoost algorithms on these data types is examined.

1 Introduction

PIM kinases are newly identified serine/threonine proteins playing an extremely important role in the regulation of cell proliferation, migration, and apoptosis. The Proviral Integration site for Moloney murine leukemia virus (PIM) is expressed as three different polypeptides, catalyzing the same chemical reaction, but having different amino acid sequencing (isozymes), PIM 1/2/3, which were found to take part in immunomodulation system and to be overexpressed, especially in hematologic malignancies and also in many other tumors. The three isozymes share a high degree of amino acid sequence homology (61% and 71%, respectively, between PIM 1 and 2, and between 1 and 3) (Liu et al. 2020). Figure 1 gives an overall view of the PIM 1 secondary structure complexed with a known inhibitor (shown in Fig. 2).

Various potent and selective chemotypes were developed for the druggability of the PIM kinases, but some cross-reactivities have been observed, with other tyrosine kinases such as Casein kinase 2 and Phosphoinositide 3-kinase family (Arrouchi et al. 2019). The three isozymes do not have an internal mechanism that regulates their biological activity, thus, their expression levels govern their constitutive activity (Qian et al. 2005). The activity of PIM 1, for example, is regulated through tight controls in transcription, translation, and proteasomal degradation, it's binding to Hsp90 reinforces its post-translational stability, whereas binding to Hsp70 causes the degradation of the PIM 1 activity (Le et al. 2015). However, the druggability

M. O. Mousser (✉) · K. Kraim · F. Chafaa · M. Brahimi
Laboratory of Physical Chemistry and Biology of Materials, Department of Physics and Chemistry, Higher Normal School of Technological Education of Skikda, 21300 Azzaba, Skikda, Algeria
e-mail: mohamedoussama.mousser@enset-skikda.dz

F. Chafaa
e-mail: f.chafaa@univ-batna2.dz

M. Brahimi
e-mail: brahimi@esi.dz

F. Chafaa
Faculty of Natural, Life Sciences University of Batna 2, 05078 Batna, Algeria

M. Brahimi
The National Higher School of Artificial Intelligence (ENSIA), Mahelma, Algeria

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_8

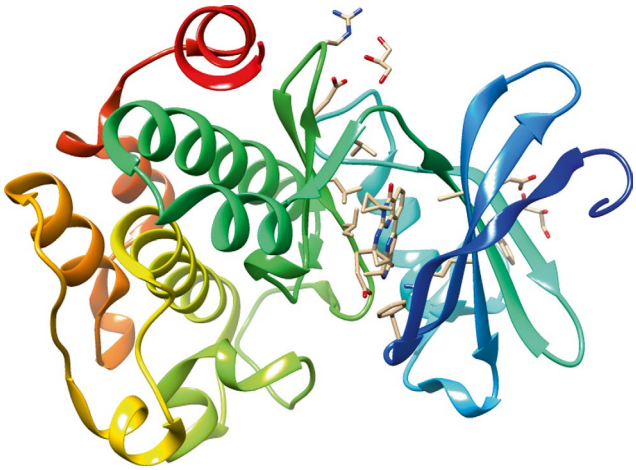


Fig. 1 Crystal structure of human PIM-1 kinase in complex with a quinazolinone-pyrrolodihydropyrrolone inhibitor (PDB: 6mt0) (Wang et al. 2019). The figure was generated using UCSF Chimera (Pettersen et al. 2004)

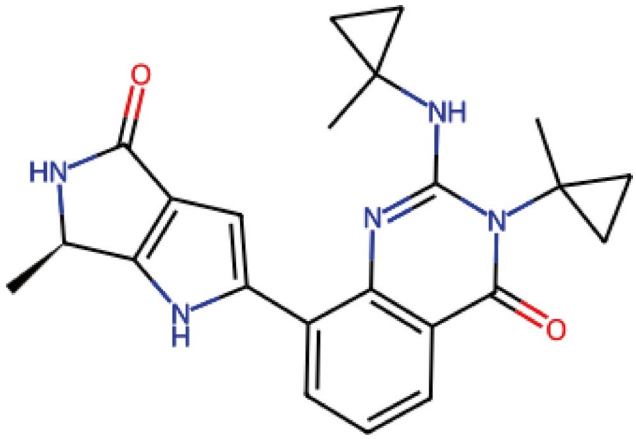


Fig. 2 Structure of the 6mt0 pdb complexed ligand: quinazolinone-pyrrolodihydropyrrolone, $IC_{50} = 0.05$ nM for both PIM 1 and PIM 2 (Wang et al. 2019). The chemical structure was generated using the RDKit suite (Bento et al. 2020)

of these enzymes is a time-consuming process. The goal of the oncologic drug research community is to find chemical compounds that inhibit these protein's activities when observed to be abnormal (over-expression). Only few PIM kinase inhibitors reached clinical trials and are not yet assured to surpass this stage. Figure 3 shows the structure of CXR1002, which reached phase 1 clinical trials in 2011, as a PIM kinases inhibitor and thus as an anticancer agent (Le et al. 2015).

One can note that from a structural point of view, inhibitors of a single protein could be completely different, yet presenting no pronounced difference in the pharmacodynamic properties, but entirely different at the pharmacokinetic level. The structures of the two active compounds (Figs. 1, 2) show the great difference that could be between two chemical compounds having a same given property, the first is a heterocyclic compound, and the second, an ammonium perfluorooctanoate, therefore, this calls the use of technological techniques for elucidating key factors influencing the phenomena. In addition, with the time-consuming aspect and early-stage failures of the drug development procedure, besides the bio-activity data increasing availability and the growth of computational power, nowadays, the use of in silico drug designing tools has become extremely demanded. One of the most important elements of the Computer-Aided Drug Design (CADD) tools is the Quantitative Structure Activity Relationships (QSAR) modeling (Cherkasov et al. 2014), in which cheminformaticians analyze experimental data generated by other scientists, and try to extract information from it in a prediction pursuing sense. The development of artificial intelligence (AI) techniques, especially machine learning (ML) and deep learning (DL), yielded a major step forward for drug designers in the medication research and development sector (Dudek et al. 2006).

Despite the power of DL and its ability to learn from any sort of dataset's complexity, avoiding under and over-fitting

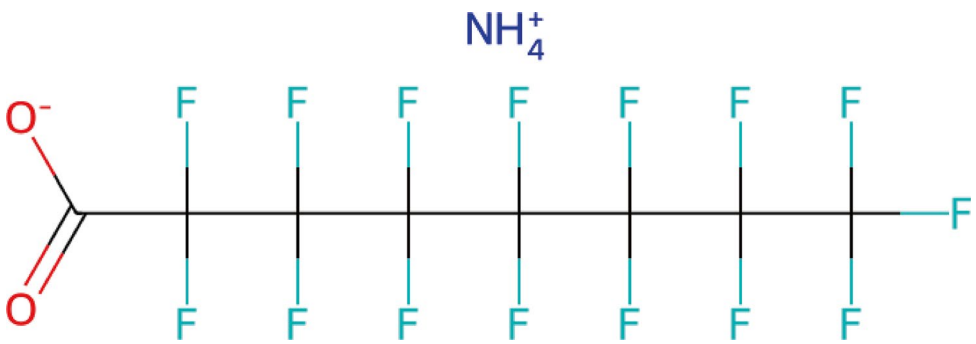


Fig. 3 Structure of CXR1002 PIM kinases inhibitor, IC_{50}: 40 µM PIM 1, 170 µM PIM 2, and 240 µM PIM 3 (Barnett et al. 2010), Cytotoxic to a wide range of human tumor cells, including pancreatic and ovarian carcinoma and sarcoma. The chemical structure was generated using the RDKit suite (Bento et al. 2020)

when working on neural network-based QSAR modeling, is an expert task. These are somehow negative aspects of DL that makes users more inclined to the use of other simple machine learning algorithms such as decision trees (DT) (Wu et al. 2021).

DT differs from most of the classification and regression statistical algorithms, in fact, each decision tree can be illustrated as an ensemble of predictive boolean rules. The classifiers are built on a tree structure system, including nodes and links. Nodes are linked in a hierarchical manner with multiple children nodes spread out from an actual parent node, in addition to leafs that are nodes with no ramifications. In each node is made a test using one feature, and based on the result of the test, the algorithm is directed to one of the child nodes branching from the parent, another test is performed there, and further traversal of the tree towards the leafs is carried out. Finally a decision is made based on the class associated with the leaf and on the probabilistic aspect of the two classes (Dudek et al. 2006). Figure 4 shows a typical DT-based classification principle using three descriptors.

In this paper, a bio-activity data science and information extracting logic is presented, along with a comparative machine learning algorithms study for modeling the structure–activity relationships of PIM 1/2/3 kinases inhibitors, by the use of three efficient and known DT-based algorithms, namely: Random forest (Singh et al. 2015), Support vector machine (Czerminski et al. 2001), and Extreme gradient boosting (Sheridan et al. 2016). The aim is to check the usability of these methods over the ChEMBL PIM kinases data for both avoiding the computational cost of DL and giving a serious alternative to the classical QSAR restrictive procedure (small datasets). In addition, a comparison between their strengths is presented through the disscussion of the resulting predictive ability of the models. We present a related works section followed by materials and methods in which we provide computational details, an experiments section and a brief results discussion section in which our findings are discussed. Finally, a conclusion and perspectives for the current work are presented.

2 Related Works

Deep learning (DL) as a subfield of machine learning in which artificial neural networks automatically learn and adapt from a large amount of data to represent the distribution characteristics of raw data. Deep networks are the connection of artificial neurons organized as input, hidden, and output layers.

Deep models rely on back propagation methods which propagate backward the signal from the output layer to the input layer for checking the importance score of different parts of the input (Li et al. 2019). The backpropagation is

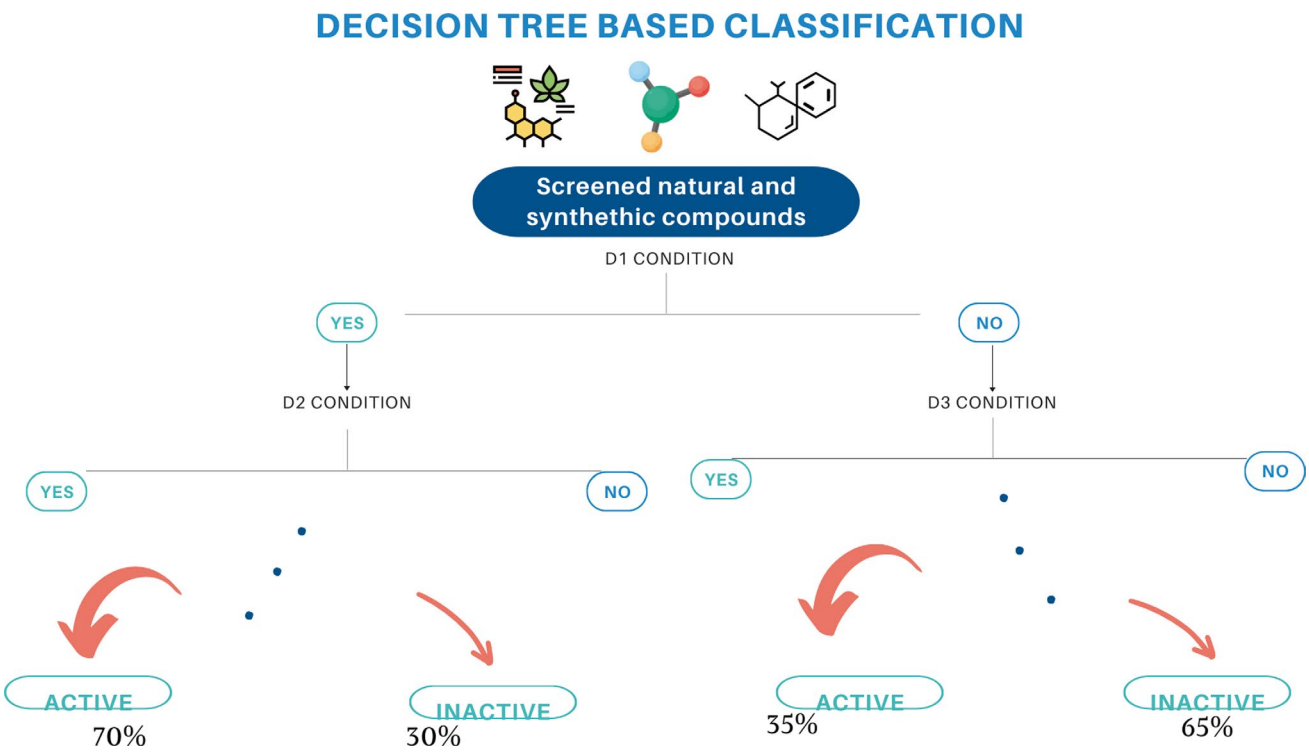


Fig. 4 Classification principle illustration using a decision tree based on three molecular descriptors conditions (*Source* the author)

thus used to tune the weights of the input and hidden layer's neurons (Dudek et al. 2006).

Training DL models usually require a large amount of data, which increases the computational cost. For this reason, Graphical Processors (GPUs) combined with larger memory are frequently used to speed up calculations on the great number of DL parameters (Chen et al. 2020).

Compared to DL, DT-based algorithms are less computationally costly (Sheridan et al. 2016). The boosting method with decision tree as base model has been used for QSAR modeling, and has shown significant performances within the COX-2 inhibition, estrogen, and dopamine receptors binding, multi-drug resistance reversal, CDK-2 antagonist activity, BBB permeability, logD, and P-glycoprotein transport activity (Svetnik et al. 2005). Random Forest (RF) and Support Vector Machine (SVM) are among the most popular decision tree methods, with distinct characteristics and applications (Sheykhmousa et al., 2020). These algorithms are frequently used in QSAR modeling with a relatively satisfactory results (Wu et al., 2021). In addition, XGBoost is also a very powerful DT-based algorithm that was poorly used in the drug discovery field. The biggest strength of XGBoost is its speed (Sheridan et al. 2016). We can define a set of parameters at which these DT-based algorithms present a good predictive ability by the use of new coding techniques.

The currently used QSAR methods include simple modeling techniques (i.e., multiple linear regressions) based on a small amount of chemicals (25 to 70) having generally the same base scaffold and a similar activity profile (Aouidate et al. 2017; Walhekar et al. 2022; Asati et al. 2020; Salman et al. 2019). This choice allows predicting efficient drug-like compounds with better pharmacological profile than the existing ones, but with the advent of AI, and the evolution of computational power, these methods are becoming less interesting to promote drug discovery. In particular, such models can only be applied for the activity prediction of same type of compounds (the activity cannot be predicted for any chemical structure that is outside the defined applicability domain of the model). In addition, notwithstanding the increasing attention in ML applied to the study of kinase therapeutic targets (Salimi et al. 2022; Vignaux et al. 2020), very few works were carried out to model the PIM kinases inhibition, for which there is a strong demand, especially since there are no approved drugs targeting the over-expression of these enzymes in the market. This is an important issue, considering that their implication in human autoimmune diseases, and potentially early rheumatoid arthritis, have recently been shown (Maney et al. 2021). The strength of the current methodologies is that all of them are based on the "same properties from same base scaffolds" principle (Abdulrahman et al. 2023). In drug design, one can frequently observe that some pharmacophoric groups are responsible of the larger portion of the binding affinity with the protein targets. However, this is not a general rule, since ligand–receptor interactions consist mainly of hydrogen bonds and electrostatic contacts, and it is always possible to find a better protein–ligand affinity with less toxic effects when virtually screening a different type of compounds than the ones known for their inhibition power (the activity of compounds with different chemical fragments can be predicted using ML). A highly relevant PIM 1 kinase investigation was recently performed based on ML classification (Almukadi et al. 2023). In this study, a pronounced reduction of the descriptors number by either genetic algorithms or the Boruta method (Kursa et al. 2010) was observed. Although it is known that only mathematically relevant descriptors are governing the structure–activity relationship models, the biological activity involves complex phenomena that could also be governed by details that seem not too relevant mathematically at a first basic view. Keeping a great number of descriptors will somehow allow the machine to learn from all the structural aspects of the bio-active compounds in the smaller details. Another important conclusion from that work was that, when working with large datasets, the chemical structures on which ML is built are of various types, and there will possibly be no more need to define an applicability domain, especially if we investigate that diversity by similarity measures between the compounds in the train set, this is possible by using Morgan or Tanimoto similarity measuring methods (Medina-Franco & Maggiora 2013). In addition, when labeling compounds as actives and inactives based on their IC_{50}, the limit for considering a chemical as active is generally set to $IC_{50} \leq 1000$ nM (Almukadi et al. 2023), and this is also a point that could be reconsidered, because we can expect a better tracking of the active-like compounds if this value is lower.

3 Materials and Methods

In this section, we shall describe the database used for the research scheme, the pre-processing methodology as well as the mathematical pre-treatment, and the training procedure.

3.1 Databases

Experimental bio-activity raw data was collected from the ChEMBL database (Mendez et al. 2019) for the three PIM kinases therapeutic targets. ChEMBL database is a large and open-access platform for bio-active chemicals. It includes important improvements with respect to the previous versions of the platform, such as a robust recording and

representation of experimental assay details with a new data deposition system allowing continuous update of datasets and the addition of supplementary biological assay results. In addition, it contains 2D structures, calculated properties (e.g., logP, Molecular Weight, Lipinski Parameters, etc.), and condensed bio-activity parameters (e.g., binding constants, pharmacology, and ADMET data) (Mendez et al. 2019). The data has been extracted and compiled from the original scientific literature, and then carefully selected and organized for presentation, along with a redesigned friendly web interface as well as enhanced search and filtering capabilities.

The ChEMBL database is formatted as a relational database (Mendez et al. 2019). It consists of several tables, including the compound table (chemical structures of the compounds and their properties), the assay table (assays used to measure the bio-activity of the compounds), and the target table (information about the protein targets) (Mendez et al. 2019).

Data collected consists of comma-separated value (CSV) files, containing detailed information for each compound assay. Including an identifier in the platform to facilitate the traceability, a molecule appellation if known (e.g., commercial name of a drug), molecular weight, number of Lipinski's rules violations, partition coefficient for lipophilicity, smiles code, standard relation between the reported and the actual value of the biological property, the reported value itself, it's unit, and much more relevant data as ligand efficiency and assay description, type, organism, cell type, parameters, etc. Bibliography referencing for each substance is also provided as the name of publication journal, publishing year, and the paper's title. The database contains information on more than 2.2 million compounds and over 18 million records of their effects on biological systems (Mendez et al., 2019).

3.2 Data Pre-Processing and Mathematical Pre-Treatment

Data pre-processing and mathematical pre-treatment were performed using the Pandas (McKinney et al., 2010) and NumPy (Harris et al. 2020) libraries and chemical features were generated using the paDEL software (Yap 2011). One-dimensional and topological (1D and 2D) molecular features were generated, and the classical QSAR mathematical pre-treatment was applied. Columns containing constant values were eliminated through the objective procedure, while for the subjective elimination, a Pearson's product-moment correlation coefficient (PCC) (Benesty et al. 2008) was computed to measure the similarity of the information expressed between each two descriptors (Eq. 1), a correlation matrix was then generated and only one column is retained if the PCC of a pair exceeds the threshold of 0.9 (columns correlated with a percentage greater than 90% were considered to give the same information).

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

where
x_i = values of the x-variable in a sample.
$\bar{x}$ = mean of the values of the x-variable.
y_i = values of the y-variable in a sample.
$\bar{y}$ = mean of the values of the y-variable.

After operating the descriptors elimination based on non-variating (constant) columns and highly correlated pairs of descriptors, a train/test splitting was made randomly with 75/25% proportions, and a standardization procedure was held to give all features the same influence on the distance metric. In fact, if one feature has very large range of values, it will govern other features when calculating the distance. For this reason, we make a scaling (transforming) to make an equal contribution of each descriptor and therefore improve the potential accuracy of the classifiers we intend to construct (Singh & Singh 2020). For every column, we take the maximal and minimal value, and replace each cell value (X) with a normalized X' value calculated by Eq. 2. Scaling is only based on the train set, minimal and maximal values of the feature column were defined from the train set, based on which the transforming was done over both the training and testing datasets.

$$X' = \frac{X - \min}{\max - \min} \quad (2)$$

where
X' = normalized value of the feature.
X = actual value of the feature.
min = minimal value in the descriptor column.
max = maximal value in the descriptor column.

3.3 Training and Evaluation Metrics of the Models

Extreme gradient boosting (Chen & Guestrin, 2016), Random Forest (Ho, 1995), and Support Vector Machine (Cortes & Vapnik, 1995) algorithms were used to fit ML models as implemented in the scikit-learn library (Pedregosa et al., 2011). These DT-based algorithms have many adjustable hyperparameters. In order to prepare them for specific needs, tuning these parameters is important to avoid neither over nor undertrained QSAR models. The scikit-learn RandomizedsearchCV function was used to find the optimal combinations of hyperparameters for each performed model, this is considered as an effective method for

finding an optimal alignment of the ML algorithm hyperparameters (Bisong & Bisong 2019; Paper & Paper 2020).

In order to evaluate the classification models, we calculate accuracy (Eq. 3), sensitivity (Eq. 4), specificity (Eq. 5), and the area under the curve (AUC). In these parameters, as inactives and actives, represent, respectively, the positive class (1) and the negative class (0), true positive (TP) is the number of compounds that are correctly determined as inactives, the true negative (TN) is the number of chemicals correctly predicted actives, while false positive (FP) is the number of substances that are classified in an incorrect way as inactives, and the number of compounds that are also incorrectly predicted actives was defined as false negative (FN). The AUC measures the true positive rate (TPR) versus the false positive rate (FPR) to evaluate the classifiers.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

Accuracy is a measure of overall correctness and is calculated as the ratio of correctly predicted instances to the total instances (*Numberof Correct P redictions/T otalNumberof P redictions*). Sensitivity measures the proportion of positive instances that are correctly identified by the model (also called true positive rate or recall), whereas the specificity (or true negative rate), measures the proportion of negative instances that are correctly identified by the model. These metrics are essential for evaluating the performance of a classification model. High accuracy is generally desirable, but sensitivity and specificity provide insights into how well a model performs on specific classes and whether it tends to have more false positives or false negatives.

Figure 5 presents the workflow through which the technical research methodology was undertaken.

In summary, the data was collected, primarily cleaned by the elimination of inorganic compounds and molecules for which the activity or the SMILES code isn't reported. A ligands preparation was then performed, by the removal of duplicates, conformers and compounds with chemical valence issues. The chemical descriptors were generated and the objective/subjective eliminations were made, followed by the ML fitting using the up-cited algorithms, and the hyperparameters optimization procedure, to finally analyse the models obtained and discuss their power, based on which, the best ones were selected. We present the details of each step in the experiments section.

4 Experiments

4.1 Data Analysis

We present in this section the organized data discussion and the results of our study. Datasets retrieved from the EMBL-ChEMBL database are used for chemical information extraction. The first step aims to prevent any potential lack of information or traceability during the next steps. Rows with missing values in: substance ID, IC_{50}, smiles, journal document and the year of publication were dropped. In addition, we eliminate all those compounds for which the standard relation between the reported IC_{50} value and the measured one isn't equal (=) or much more greater (>>). As to classify ligands as actives and inactives, the activity criteria is exclusively based on the value of IC_{50}, thereby, active compounds must have the exact experimental half-maximal inhibitory value. Whereas, inactive compounds would have values greater than a defined threshold. Regarding the fact that literature lacks exact rules for an IC_{50} value under which the inhibition of the biological activity is assured, ligands were labeled based on a statistical reasoning as follows: those having a half-maximal inhibitory concentration under 200 nM are considered as actives, while those with beyond 1000 nM are considered as inactives. Note that intermediate values are not considered to ensure precision. In fact, the machine is supposed to learn from the descriptors with highly and poorly bio-active compounds, thus, it will be able to define a well separated activity and non-activity perimeters, and consequently reinforce precision. Hence, the trained model is ready to screen a huge chemical library. The plot in the Fig. 6 indicates data availability distribution by half-maximal inhibition measured value, and by activity class in the PIM 1 training set (orange dots refer to active compounds, while blue ones are about inactive compounds), as labeled by the above procedure, in addition to the density of variation of both classes.

4.2 Machine Learning Models

For each studied model, the tuned parameters and experimental values are presented in Table 1. The parameters were fixed based on commonly used combinations (Mantovani et al. 2018). We used RandomizedSearchCV for our experimentations with values presented in the range column.

After convergence, Table 2 presents the accuracy, sensitivity, specificity and AUC of the three decision trees. Next we evaluate our classifiers by the analysis of confusion matrices, within which we find in each line a real class, and in each column an estimated class (true labels in the rows and predicted labels in the columns) (Zeng 2020).

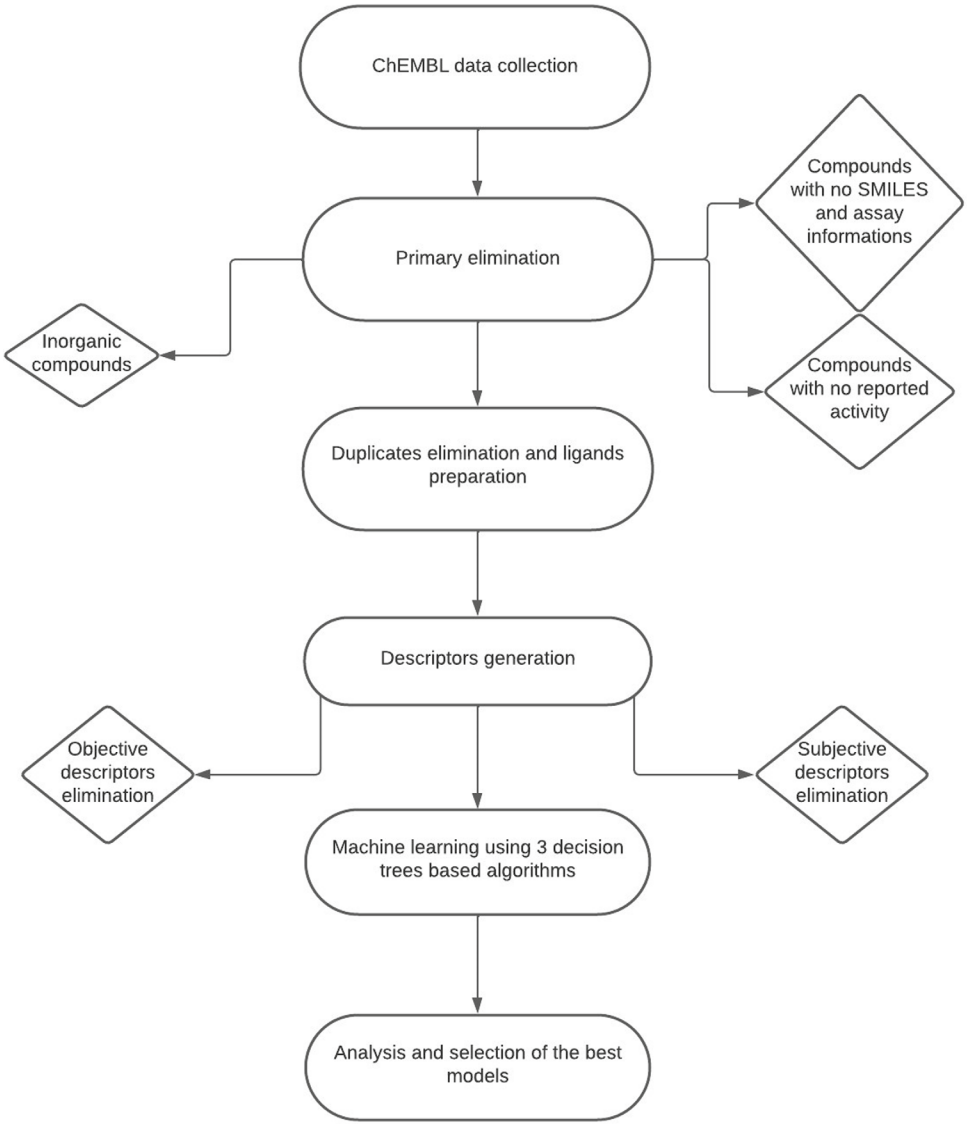


Fig. 5 Workflow of the research methodology (*Source* the author)

Figure 7 shows the confusion matrix of PIM 1, in which active compounds are labeled negative, while the non-active chemicals are labeled positive. We notice that TP and FN contain, respectively, 115 and 30 points, indicating a $\frac{115}{145}$ ratio for the sensitive power (79%). Whereas TN and FP include 172 and 19 points, respectively. This leads to a specificity ratio of 90% $\left(\frac{172}{191}\right)$.

The PIM 2 confusion matrix (Fig. 8) shows 69, 18, 92, and 11 compounds in the TP, FN, TN, and FP squares, respectively. This represents a ratio of $\frac{69}{87}$ and $\frac{92}{103}$ for the sensitive and specific character of the classification algorithm, respectively. In the PIM 3 model, the ratios are of $\frac{15}{26}$ for the sensitivity, and $\frac{60}{65}$ for the specificity (see Fig. 9). This could provide in theory a percentage of approximately 92% of accurately predicted active compounds and 58% of accurately supposed inactive compounds when using the model to screen a given chemical library. Hence, the obtained model has a very weak predictive power, accentuating the need of a deeper algorithm's optimization.

The RF and XGBoost classifications concerning the PIM 1 and 2, respectively, could be considered as excellent from a probabilistic point of view. In addition, if we consider Tropsha and Golbraikh conditions for a classical predictive QSAR model (Tropsha 2010), the evaluation metrics are very satisfying.

The ROC curve as an evaluation metric for binary classification problems is used to plot the TPR against FPR at various threshold values, showing the performance of the classifier at each threshold (Zeng 2020). The AUC as a summary of the ROC plots is also used to evaluate the model. When AUC = 1, the classifier can correctly distinguish between all the positive and the negative class points. However, if it is near the zero, then the classifier would invert all classes and thus predict all wrong. Therefore, we

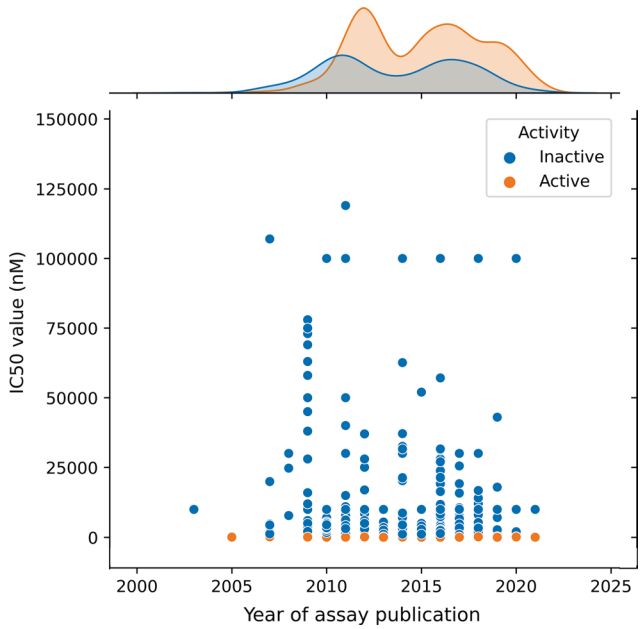


Fig. 6 Plot of the IC_{50} values over the year of the assay publication in our PIM 1 kinase cleaned dataset. The figure was created using seaborn (Waskom 2021)

Table 1 The values of experimented hyperparameters for the three decision trees

Algorithm	Parameters	Range
RF	n estimators	[100, 1000]
	max features	499
	max depth	[10, 50]
	min samples split	[2, 10]
	min samples leaf	[1, 10]
XGBoost	n estimators	[10, 2000]
	max depth	[3, 15]
	learning rate	[0.01, 0.3]
	colsample bytree	[0.2, 1]
	gamma	[0, 2]
	min child weight	[1, 10]
SVM	Penalty parameter C	[0, 10]
	Kernel type	['linear','rbf']
	Kernel coefficient	['scale','auto']
	Degree of the polynomial kernel function	[2, 3, 4]

can define a range between 0.5 and 1, in which the classifier could be moderately able to distinguish the active compounds from the inactive ones. Due to its ability to detect more numbers of TN and TP than FN and FP which is the case of our selected models: 0.85, 0.84 and 0.75 for the PIM 1 (RF), PIM 2 and PIM 3 (XGBoost), respectively.

When using ROC curves to evaluate our prediction tool, we expect the biggest area under the TPR(FPR) curve. It is directly proportional to a maximal TPR value corresponding to a minimal FPR value in the point where the curve starts sloping up considerably slower (inflexion). As expected from the AUC value results, the PIM 1 model is having a slightly better TPR(FPR) inflection point than the PIM 2 model (Fig. 10 and Fig. 11). Both are noticeably better than the PIM 3 classifier, which has a lower AUC value with almost 0.10 (Fig. 12). The relatively small AUC of the PIM 3 model can be clearly observed, due to a weak TPR in the inflection point (≈0.6), resulting in a low sensitive character (only 58% of class distinguishing ability).

5 Results and Discussion

Based on the confusion matrices and ROC curves (shown in Figs. 7 – Fig. 12), we can notice that at high AUC values, the selected models are globally considered able to distinguish between positive and negative classes. We also underline that, when it equals 0.5, then it has no ability to distinguish between positive and negative class points. In this case, the classifier either predicts a random class or a constant class for all the data points. As for PIM 3 SVM model, the AUC indicates exactly 0.5. The SVM model is unable to differentiate between active and non-active compounds, resulting in a zero sensitive power and an accuracy of 71% of accidentally correct predictions in the test set. Further, the constructed models are built based on a satisfactory diverse base scaffold chemicals, which allows for the screening of extremely large chemical databases. The Tropsha and Golbraikh criteria for QSAR models validation (Tropsha, 2010) are also satisfied when superposing the classification's metrics to the classical regression ones.

The feature elimination can only be based on a low variance filter and a correlation filter. Thus, dimensionality reduction won't be too pronounced, and the machine could learn the structure–activity relationship within the smallest details, particularly knowing that machine and deep learning are used for predicting in complex data, and there is no more mandatory need for modeling with basic linear nor polynomial QSAR equations.

The RF classifier was found to be suitable for our datasets and provides a noteworthy robustness against overfitting, while XGBoost excels in predictive performances. Despite the fact that SVM is powerful for both classification and regression tasks, we supposed after searching a good hyperparameters combination, that it has no notable strength regarding our datasets type. However, the used algorithms are considered to give an excellent quality/speed consensus, depending on the dataset's size, complexity

Table 2 Summary table of the constructed decision trees relevant information

Models	Algorithm	Train shape	Test shape	Sensitivity	Specificity	AUC	Accuracy
PIM 1	RF	(1011, 501)	(336, 501)	0.79	0.9	0.85	0.85
	SVM			0.02	1	0.51	0.57
	XGBoost			0.75	0.9	0.83	0.84
PIM 2	RF	(570, 478)	(190, 478)	0.78	0.86	0.82	0.83
	SVM			0.68	0.81	0.75	0.75
	XGBoost			0.79	0.89	0.84	0.85
PIM 3	RF	(274, 461)	(91, 461)	0.38	0.93	0.66	0.78
	SVM			0	1	0.5	0.71
	XGBoost			0.58	0.92	0.75	0.82

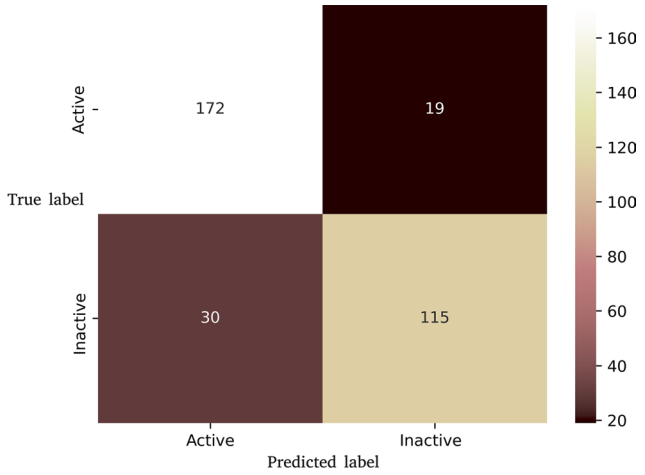


Fig. 7 Confusion matrix of the PIM 1 classifier (*Source* the author)

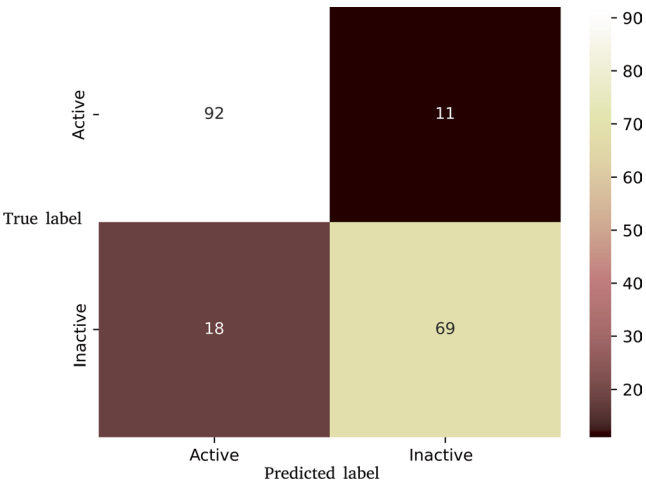


Fig. 8 Confusion matrix of the PIM 2 classifier (*Source* the author)

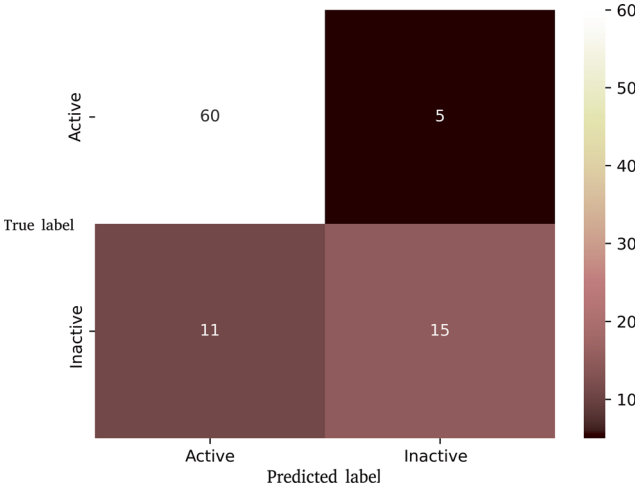


Fig. 9 Confusion matrix of the PIM 3 classifier (*Source* the author)

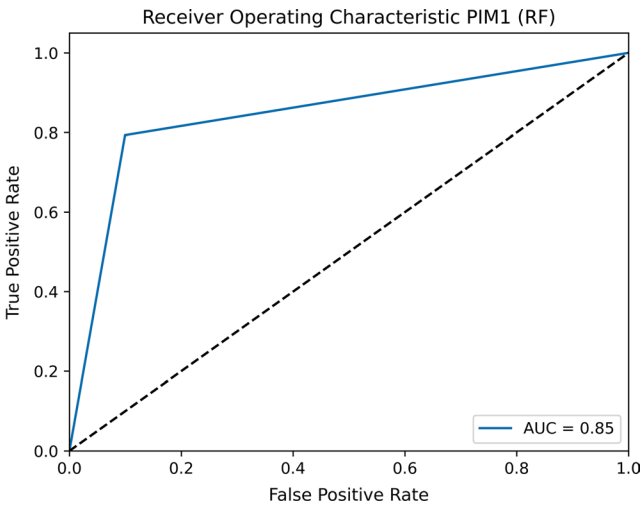


Fig. 10 Plot of the receiver operating curve for the random forest classifier (PIM 1) (*Source* the author)

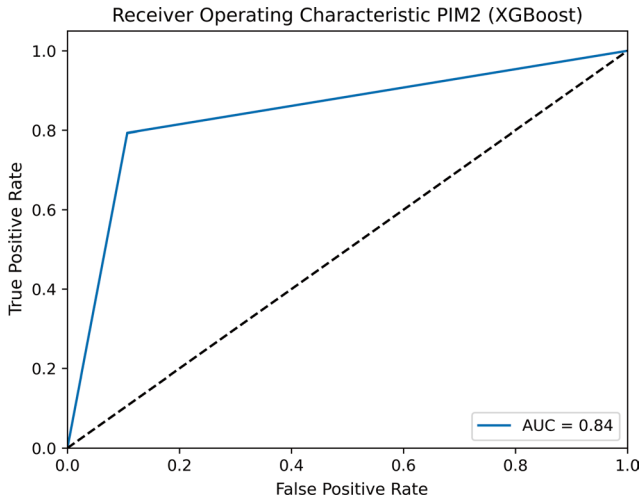


Fig. 11 Plot of the receiver operating curve for the XGBoost classifier (PIM 2) (*Source* the author)

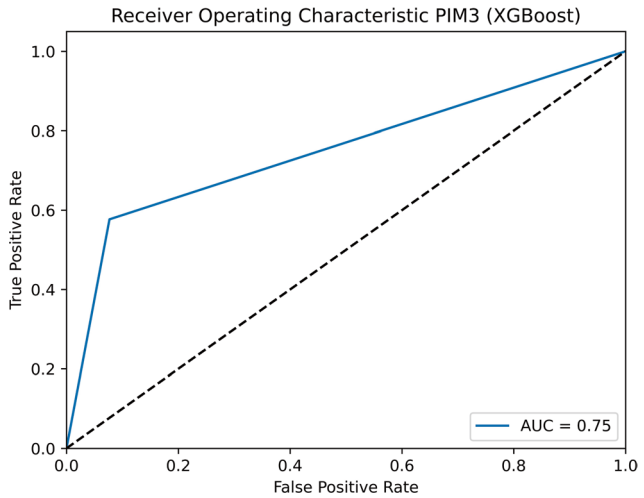


Fig. 12 Plot of the receiver operating curve for the XGBoost classifier (PIM 3) (*Source* the author)

and on the user's machine power, which isn't the case of deep learning, that generally requires more computational resources (Justus et al., 2018), and in which, it is more difficult to avoid under and over-fitting issues.

6 Conclusion

In this study, we collected bio-activity raw data concerning three Proviral Integration Site for Moloney murine leukemia virus (PIM protein kinases) from the ChEMBL platform, we cleaned these data tables to obtain a pre-processed datasets. Those compounds were prepared and molecular features were generated in order to train ML QSAR classifications, before which, a mathematical pre-treatment was applied. Three learning algorithms were used to evaluate predictive performances and a rational comparison between them was presented. In summary, we conclude that dataset size, non-linearity, interpretability, and computational power of the user's machines play crucial roles for the designation of the most appropriate algorithm in contemporary QSAR studies. Therefore, we emphasize the RF and XGBoost algorithms as highly efficient for binary classification tasks in drug discovery, for their predictive power, speed, and possible usage within modest-performance machines. Thus, we expect an improvement of the presented classifiers with deeper hyperparameter optimizations, and their use to virtually screen large chemical databases in order to identify novel hits against the PIM kinases, and consequently help the neoplastic targeted therapy rapid progress.

References

Abdulrahman FG, Sabour R, Abd El-Gilil SM, Mehany AB, Taha EA (2023) Design, synthesis, biological evaluation and molecular docking study of new pyrazolo [1, 5-a] pyrimidines as pim kinase inhibitors and apoptosis inducers. J Mol Struct, 136811

Almukadi H, Jadkarim GA, Mohammed A, Almansouri M, Sultana N, Shaik NA, Banaganapalli B (2023) Combining machine learning and structure-based approaches to develop oncogene pim kinase inhibitors. Front Chem 11:1137444

Aouidate A, Ghaleb A, Ghamali M, Chtita S, Choukrad M, Sbai A, Lakhlifi T (2017) Qsar studies on pim1 and pim2 inhibitors using statistical methods: a rustic strategy to screen for 5-(1h-indol-5-yl)-1, 3, 4-thiadiazol analogues and predict their pim inhibitory activity. Chem Cent J 11:1–10

Arrouchi H, Lakhlili W, Ibrahimi A (2019) A review on pim kinases in tumors. Bioinformation 15(1):40

Asati V, Agarwal S, Mishra M, Das R, Kashaw SK (2020) Structural prediction of novel pyrazolopyrimidine derivatives against pim-1 kinase: In-silico drug design studies. J Mol Struct 1217:128375

Barnett A, Ding S, Murray C, Chamberlain M, Plummer S, Evans T, Wolf C (2010) Anti-tumor activity of cxr1002, a novel anti-cancer clinical phase compound that induces er stress and inhibits pim kinases: human tumor xenograft efficacy and in vitro mode of action. EJC Suppl 8(7):45–46

Benesty J, Chen J, Huang Y (2008) On the importance of the pearson correlation coefficient in noise reduction. IEEE Trans Audio Speech Lang Process 16(4):757–765

Bento AP, Hersey A, Felix E, Landrum G, Gaulton A, Atkinson F, Leach AR (2020) An open´ source chemical structure curation pipeline using rdkit. J Cheminformatics 12:1–16

Bisong E, Bisong E (2019) Introduction to scikit-learn. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, 215–229

Chen C, Zhang P, Zhang H, Dai J, Yi Y, Zhang H, Zhang Y (2020) Deep learning on computationalresource-limited platforms: a survey. Mob Inf Syst 2020:1–19

Chen T, Guestrin C (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM Retrieved from http://doi.acm.org/https://doi.org/10.1145/2939672.2939785 https://doi.org/10.1145/2939672.2939785

Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M et al (2014) Qsar modeling: where have you been? where are you going to? J Med Chem 57(12):4977–5010

Cortes C, Vapnik V (1995) Support-Vector Networks. Mach Learn 20(3):273–297

Czerminski R, Yasri A, Hartsough D (2001) Use of support vector machine in pattern classification:´ Application to qsar studies. Quant Struct-Act Relat 20(3):227–240

Dudek AZ, Arodz T, Galvez J (2006) Computational methods in developing quantitative structure-´ activity relationships (qsar): a review. Comb Chem High Throughput Screening 9(3):213–228

Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Oliphant TE (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2

Ho TK (1995). Random decision forests. In Proceedings of 3rd International Conference On Document Analysis and Recognition 1, pp 278–282

Justus D, Brennan J, Bonner S, McGough AS (2018) Predicting the computational cost of deep learning models. In 2018 IEEE International Conference on Big Data (Big Data) pp 3873–3882

Kursa MB, Jankowski A, Rudnicki WR (2010) Boruta–a system for feature selection. Fund Inform 101(4):271–285

Le BT, Kumarasiri M, Adams JR, Yu M, Milne R, Sykes MJ, Wang S (2015) Targeting pim kinases for cancer treatment: opportunities and challenges. Future Med Chem 7(1):35–53

Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X (2019) Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. Methods 166:4–21

Liu Z, Han M, Ding K, Fu R (2020) The role of pim kinase in immunomodulation. Am J Cancer Res 10(12):4085

Maney NJ, Lemos H, Barron-Millar B, Carey C, Herron I, Anderson AE, Pratt AG (2021) Pim kinases as therapeutic targets in early rheumatoid arthritis. Arthritis & Rheumatology 73(10):1820–1830

Mantovani RG, Horvath T, Cerri R, Junior SB, Vanschoren J, de Carvalho ACP. d. LF (2018) An´ empirical study on hyperparameter tuning of decision trees. *arXiv preprint* arXiv:1812.02207

McKinney W (2010) Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference 445, pp 51–56

Medina-Franco JL, Maggiora GM (2013) Molecular similarity analysis. Chemoinformatics for drug discovery, 343–399

Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Felix E et al (2019) Chembl: towards´ direct deposition of bioassay data. Nucleic Acids Res 47(D1):D930–D940

Paper D, Paper D (2020) Scikit-learn classifier tuning from simple training sets. Hands-on Scikit-Learn for machine learning applications: data science fundamentals with python, 137–163

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay E (2011) Scikitlearn: machine learning in python. J Mach Learn Res 12:2825–2830

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) Ucsf chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612

Qian KC, Wang L, Hickey ER, Studts J, Barringer K, Peng C et al (2005) Structural basis of constitutive activity and a unique nucleotide binding mode of human pim-1 kinase. J Biol Chem 280(7):6130–6137

Salimi A, Lim JH, Jang JH, Lee JY (2022) The use of machine learning modeling, virtual screening, molecular docking, and molecular dynamics simulations to identify potential vegfr2 kinase inhibitors. Sci Rep 12(1):18825

Salman M, Ahmed S, Nandi S (2019) Qsar and anticancer drug design on benzothienopyrimidinones as promising pim kinase inhibitors utilizing structural descriptors. Int J Quant Struct-Prop RelatShips (IJQSPR) 4(2):82–99

Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM (2016) Extreme gradient boosting as a method for quantitative structure–activity relationships. J Chem Inf Model 56(12):2353–2360

Sheykhmousa M, Mahdianpari M, Ghanbari H, Mohammadimanesh F, Ghamisi P, Homayouni, S (2020). Support vector machine versus random forest for remote sensing image classification: A metaanalysis and systematic review. IEEE J Sel Top Appl Earth Obs Remote Sens, 13, 6308–6325

Singh D, Singh B (2020) Investigating the impact of data normalization on classification performance. Appl Soft Comput 97:105524

Singh H, Singh S, Singla D, Agarwal SM, Raghava GP (2015) Qsar based model for discriminating egfr inhibitors and non-inhibitors using random forest. Biol Direct 10(1):1–12

Svetnik V, Wang T, Tong C, Liaw A, Sheridan RP, Song Q (2005) Boosting: An ensemble learning tool for compound classification and qsar modeling. J Chem Inf Model 45(3):786–799

Tropsha A (2010) Best practices for qsar model development, validation, and exploitation. Mol Inf 29(6–7):476–488

Vignaux PA, Minerali E, Foil DH, Puhl AC, Ekins S (2020) Machine learning for discovery of gsk3β inhibitors. ACS Omega 5(41):26551–26561

Walhekar V, Bagul C, Kumar D, Achaiah G, Muthal A, Kulkarni R, Basavarju M (2022) Computational modelling strategies in exploring triazolopyridazine pim1 kinase inhibitors as anticancer agents. Anti-Cancer Agents Med Chem

Wang H-L, Andrews KL, Booker SK, Canon J, Cee VJ, Chavez F Jr et al (2019) Discovery of (r)-8-(6-methyl-4-oxo-1, 4, 5, 6-tetrahydropyrrolo [3, 4-b] pyrrol-2-yl)-3-(1-methylcyclopropyl)-2-((1methylcyclopropyl) amino) quinazolin-4 (3 h)-one, a potent and selective pim-1/2 kinase inhibitor for hematological malignancies. J Med Chem 62(3):1523–1540

Waskom ML (2021) seaborn: statistical data visualization. J Open Source Softw, 6(60), 3021. Retrieved from https://doi.org/10.21105/joss.03021 https://doi.org/10.21105/joss.03021

Wu Z, Zhu M, Kang Y, Leung EL-H, Lei T, Shen C, Hou T (2021) Do we need different machine learning algorithms for qsar modeling? a comprehensive assessment of 16 machine learning algorithms on 14 qsar data sets. Briefings in bioinformatics, 22(4), bbaa321

Yap CW (2011) Padel-descriptor: An open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32(7):1466–1474

Zeng G (2020) On the confusion matrix in credit scoring and its analytical properties. Commun Stat-Theory Methods 49(9):2080–2093

Parallelism in Performance Assessment: Energetic and Structural Analysis Aided by Computing Techniques

Calculating Cost-Optimal Energy Efficiency Levels for Opening Elements on an Exemplar Residential Building

Egemen Kaymaz and Filiz Senkal Sezer

Abstract

This study presents a comprehensive approach to enhance the energy performance of a multi-story residential apartment in a temperate region of Turkey. The strategy focuses on cost optimization in compliance with EPBD Directives and CEN EN 15459-1:2017 standard. It addresses energy consumption and environmental impact associated with glazed building components and openings, considering key factors like primary energy use, energy costs, CO_2 emissions, and annual heating, cooling, and lighting energy consumption throughout the year. The building energy simulation (BES)-based optimization process involves the design variables that exert a significant impact on energy efficiency. These variables encompass window size, window-to-wall ratio, solar shading systems, glazing material properties, window frame composition, and profile dimensions. EnergyPlus and DesignBuilder with genetic algorithms facilitate this process. The study explores trade-offs among design alternatives and identifies building envelope configurations that improve energy performance and reduce costs compared to the reference building. Compared to the reference case, the energy-efficient cost-optimum solution achieved savings of 23.7% in primary energy (29.1 kWh/m^2), 21.4% in global costs (44.4 €/m^2), and a 24.4% reduction in carbon emissions (above 8 kg/m^2). These results emphasize the benefits of using simulation-driven optimization in early design stages to meet energy efficiency and environmental goals.

Keywords

Energy efficiency · Cost-optimality · Optimization · Building envelope · Openings

1 Introduction

As indicated by International Energy Agency (IEA 2021), the buildings account for nearly one-third of total global energy consumption and contribute to 15% of CO_2 emissions. This trend is accompanied by a continuous surge in energy demand from buildings, particularly in developing nations. Amid Turkey's energy perspective, the IEA's 2021 energy policy review highlights that the residential sector ranks as the third largest consumer of energy. It encompasses 20% of total final consumption, 9% CO_2 emissions, 21.1% of the total electricity consumption, and 25.5% of total gas consumption in the year 2018.

According to statistical data from Turkstat (2022), Turkey boosted around 9.5 million buildings as of October 2019, with over 100,000 new constructions annually. Remarkably, approximately 90% of this building stock consists of residences, equating to 24 million dwelling units. Over the past decade (2008–2018), energy consumption in residential sector has escalated by 12%. This upswing can be attributed to the rapidly expansion and transformation of the building stock driven by population growth and urbanization. Consequently, enhancing energy efficiency within residences holds substantial significance for Turkey's economy, considering its heavy reliance on imported fossil fuels (93% oil and 99% gas) as indicated by the IEA (2021).

The European Union (EU) introduced a series of directives (EU 2009, EU 2010, and EU 2012) aimed at enhancing energy efficiency in buildings. The Energy Efficiency Directive (EU 2018a), the Directive on the Promotion of the Use of Energy from Renewable Sources (EU 2018b), and the Energy Performance of Buildings-recast Directive (EU 2018c) were published with the overarching objectives

E. Kaymaz (✉) · F. S. Sezer
Department of Architecture, Faculty of Architecture, Bursa Uludag University, Bursa, Türkiye
e-mail: ekaymaz@uludag.edu.tr

F. S. Sezer
e-mail: filizs@uludag.edu.tr

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_9

of energy savings, reducing greenhouse gas emissions, and fostering the adoption of renewable energy technologies within the buildings sector. These goals were subsequently revised from reaching the nearly zero-energy (NZEBs) level for new constructions by 2020 to align with the goal of achieving net-zero emissions (climate neutrality) in both new and renovated buildings by 2050 (EC 2019).

The recast of the EPBD in 2010 introduced a methodological framework for determining cost-optimal levels pertaining to minimum energy performance requirements criteria, focusing on achieving the most efficient balance between expenses and energy conservation on a global scale (EU 2010). Furthermore, the Energy Performance of Buildings Directive underwent a revision in 2021, incorporating several legislative proposals designed to meet the EU's broader objective of attaining a minimum 55% reduction in greenhouse gas emissions by 2030 relative to 1990 levels (EC 2021).

Since the developments within the EU are followed in Turkey, EPBD documents are the important components to direct the national building energy policy and shape the nation's building energy strategy. Substantial efforts have been dedicated to aligning with the EPBD's principles, resulting in the implementation of energy-focused regulations and laws. Examples include the enactment of the Energy Efficiency Law (EN-VER) in 2007 (Turkish Republic Ministry of Public Works and Settlement 2007) and the establishment of the Energy Performance of Buildings Regulation (BEP) in 2008 (Turkish Republic Ministry of Public Works and Settlement 2008). These measures were accompanied by the legal requirement for energy performance certification (EPCs) for both new and existing buildings (Turkish Republic Ministry of Public Works and Settlement 2010).

While the National Energy Efficiency Action Plan defines the country's specific energy performance targets to be attained by 2023 (Turkish Republic Ministry of Energy and Natural Resources 2016), it's apparent that more rigorous legal measures will be necessary to effectively achieve the ambitions of the EU climate and energy policy by 2050. This implies that Turkey is gearing up to adopt stricter legislative frameworks to ensure the attainment of the EU's long-term climate and energy objectives.

Building openings hold a pivotal role as key design elements that significantly impact a building's energy performance. The specific attributes including the geometry, optical and thermophysical properties, orientation and the use of solar shading attachments for windows determine various aspects such as daylight provision, energy conservation, solar gains, and the overall indoor thermal and visual comfort experienced by the occupants. Addressing this challenge, which involves effectively balancing trade-offs between objectives like energy efficiency, comfort, availability of natural light, and the associated costs of improvement measures has garnered the attention of numerous researchers. They have explored this complex issue utilizing optimization techniques for building performance assessment.

For instance, Tuhus-Dubrow and Krarti (2010) developed a simulation-based optimization tool to optimize building shape and envelope features for energy-efficient residential buildings. This tool, integrating a genetic algorithm and energy simulation, identifies optimal parameters across various shapes and envelope components. Their findings indicate that rectangular and trapezoidal shapes consistently exhibit better energy performance, with minimal variation among shapes when considering all envelope features, offering flexibility for architects in shape selection for different climates and building types.

Ochoa et al. (2012) conducted a study exploring window optimization, which involved evaluating both energy consumption and visual comfort criteria. Their research highlighted the importance of integrating various visual comfort measures, as optimizing solely for energy efficiency could lead to excessive consumption. Conversely, focusing solely on visual comfort neglects energy efficiency and visual acceptance standards, underscoring the significance of collaborative design that considers climatic factors and maintains a balanced approach.

In Ascione et al.'s (2016) research, design criteria for achieving nearly Zero-Energy Buildings (nZEB) in Mediterranean climates are discussed. The study employs dynamic energy simulations and multi-objective optimization to evaluate various passive strategies for building envelopes. The results indicate that achieving the best trade-off between summer and winter performance, while maintaining thermal comfort, presents challenges, but certain guidelines for selecting envelope solutions are suggested. The study emphasizes the complexity of optimizing high-performance buildings for energy efficiency and thermal comfort in Mediterranean regions.

In the work of Zhai et al. (2019), a multi-objective optimization method that combines NSGA-II and EnergyPlus was introduced to optimize window design by considering parameters like orientation, size, and material. This approach effectively addresses energy consumption, indoor thermal conditions, and visual performance, providing architects with valuable insights through Pareto-optimal solutions. This method assists designers in achieving an ideal window design that not only lowers energy usage but also enhances indoor comfort and visual quality.

Xue et al. (2019) centered their study on optimizing the Window-to-Wall Ratio (WWR) using sunshades while considering factors like daylighting performance and energy usage. Their research introduced a comprehensive workflow applied to an initial facade design for a hotel in China's

low-latitude region. The findings highlighted the superior energy performance of comprehensive sunshades, leading to the identification of optimal WWR ranges for various orientations and sunshade configurations. This approach provides an efficient technique for WWR optimization, effectively addressing visual and thermal requirements, and holds promise for adaptation to diverse regions and building types.

Liu et al. (2021) analyzed the energy efficiency of WWR designs in residential buildings within China's Hot-summer and Cold-winter zones. By assessing correlations between WWR and energy consumption for cooling, heating, and year-round needs, their research offers insights into achieving a balance between thermal comfort and energy efficiency in sustainable residential design. Similarly, Maleki and Dehghan (2021) focused on evaluating optimal window parameters to enhance energy efficiency and daylight quality in residential buildings, particularly within the hot and dry climate of Isfahan. Through simulations involving different window angles and orientations, the research identifies design choices that result in energy savings of up to 4.81%. This underscores the effectiveness of low-emissivity glass, façade orientation, shading techniques, and lighting strategies in achieving improved building performance.

Within the context of a temperate climate region in Turkey, the study focuses on optimizing the opening design of a multi-story residential apartment. In line with the methodological framework of EPBD 2010/31/EU and EN 15459-1 standard (CEN 2017), the main goals were to investigate transparent building envelope solutions that minimize primary energy consumption, reduce carbon emissions, and achieve optimal cost-effectiveness for a case study. In this regard, a building energy simulation-based multi-objective optimization approach is implemented during the façade design stage of the project. The research involves three main phases: i) building energy performance analysis of the case study; ii) sensitivity analysis to assess the impact of design variables including window size, WWR, external solar shadings, glazing's thermal and optical properties, window frame material and profile dimensions on the study objectives; and iii) an optimization analysis utilizing Genetic Algorithms (GAs) to determine optimal architectural design solutions that strike a balance between global cost and annual primary energy consumption.

2 Material and Method

In this study, we systematically investigate cost-optimal energy efficiency levels for opening elements in residential buildings using building energy simulation (BES) and BES-based optimization programs. Figure 1 provides an overview of the study, illustrating the data obtained from the respective computer programs.

2.1 Definition of the Case Study

The pilot study was conducted in Bursa Turkey (40°15′58.2″N 28°56′23.1″E) which falls within the tempered humid climate region. The study focused on multi-story residential buildings situated in a gated community built in 2016, primarily targeting high-income groups. This housing complex holds B-class building energy certification based on the Turkish Building Energy Performance Calculation Methodology (BEP-TR). The residential blocks, each spanning eight stories, were constructed on a land area of 70.638 sq. m. Each floor consists of three flats, each with a usable area of 212 sq. m. Notably, all living rooms and kitchens are oriented towards the inner courtyard. Refer to Fig. 2a for the site plan and Fig. 2b for an outdoor photograph depicting the residential project.

Site plan location, façade orientation, and obstruction angles are key factors that significantly influence both daylight availability and energy performance of buildings. The residential site comprises a total of nine apartment blocks. Among these, Block G exhibits the least favorable annual energy consumption in comparison to the remaining eight apartment blocks. Consequently, an assessment of energy efficiency measures was undertaken and simulations were run specifically for this building to improve its energy performance.

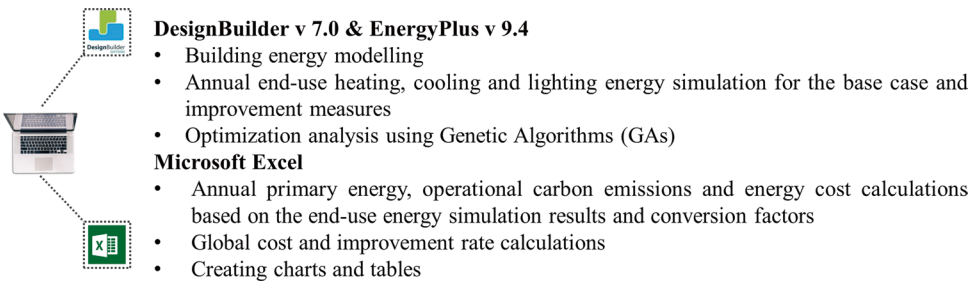


Fig. 1 Methodology of the study. (*Source* the author)

Fig. 2 **a** Site plan of the apartment blocks (Tatbikat Architecture Office archive, n.d.). **b** Outdoor image of the case study (photo taken by the author, 2019)

2.2 Reference Building Energy Performance Analysis

In the initial phase, the energy performance of the base case is determined through a detailed dynamic computational method. The geometric model of the reference building is configured using the Designbuilder Architectural Edition v7.0 software application and simulations are executed utilizing the EnergyPlus v9.4 engine. Key parameters such as thermal mass, insulation levels, HVAC (heating, ventilation, air conditioning), DHW (domestic hot water), and lighting system components are aligned with the design project which complies with the minimum requirements outlined in national BEP Regulations (Turkish Republic Ministry of Public Works and Settlement 2008) and rating of energy performance in buildings according to BEP-TR (Turkish Republic Ministry of Public Works and Settlement 2010) for the temperate and hot-dry climatic zones.

Weather data specific to of Bursa is obtained in the typical meteorological year (TMY) and EnergyPlus Weather format (EPW) through from an online database (Climateonebuilding, n.d.) Additionally, user preferences concerning indoor environmental conditions, operational schedules, and occupant profiles are analyzed through in-person surveys. The feedbacks obtained from these questionnaires, pertaining to boundary conditions (including building service system parameters, set point temperatures, household energy expenditure, etc.) is also integrated as reliable data within the simulation model.

The primary energy consumption is defined by multiplying the end-use energy with governmentally determined conversion factors, set at 1 for natural gas and 1.826 for electricity. The carbon emission conversion factors by are 0.234 and 0.555 for natural gas and electricity, respectively (Turkish Republic Ministry of Environment, Urbanization and Climate Change 2020). In Fig. 3, the energy performance of the case study building is presented, displaying the end-use energy, primary energy consumption, as well as operational CO_2 emissions and energy costs. The exchange rates for Euro/Turkish Lira are determined based on the prices of August 2022 (European Central Bank 2022).

2.3 Sensitivity Analysis

The energy efficiency of building openings is influenced by a range of parameters. To expedite the optimization process and reduce evaluation time for potential solutions, we conduct a sensitivity analysis to examine the impact of these parameters on the simulation results. In the subsequent phase, energy efficiency measures are defined for opening elements, and the performance of each measure is assessed through a series of parametric simulations. Additionally, the pay-back period of the initial construction costs is taken into account and operational carbon emissions are assessed. The research encompasses the following scenarios, summarized as follows:

Window Size and Configuration

Nine window-to-wall ratios (WWR) ranging from 0.20 to 0.60 in 5% increments and four window heights (1.00 m, 1.50 m, 2.00 m, 2.50 m) were taken into consideration for the residential spaces (see Table 1). The WWR of the reference building is 30%. The width of the windows in

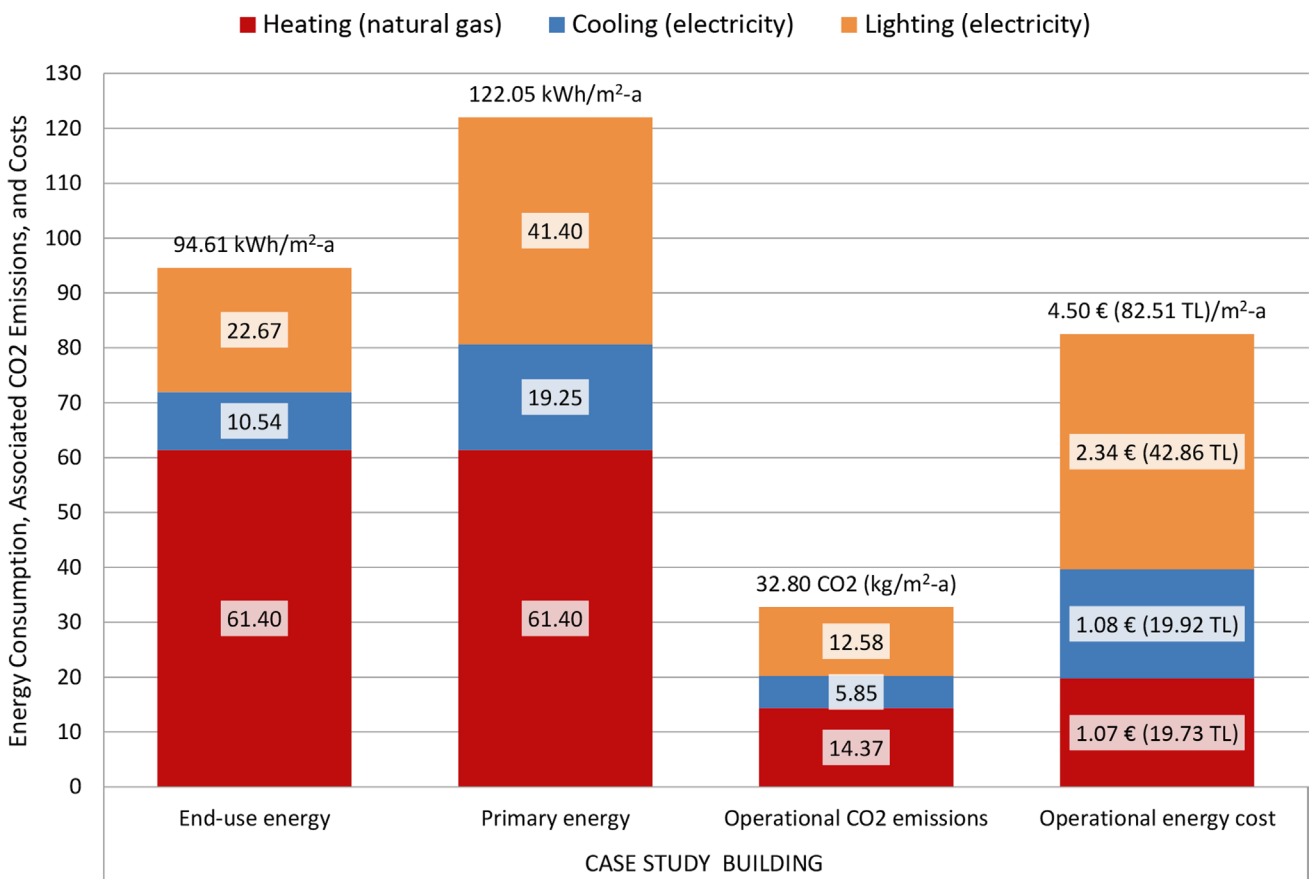


Fig. 3 Simulation results for the reference building with respect to the research objectives, including end-use energy, primary energy consumption, operational CO_2 emissions and energy costs. *Source* the author

Table 1 Façade transparency and window geometry scenarios

Window-to-wall ratio (%)	Window and parapet height alternatives			
	Window h: 1 m Parapet h: 1 m	Window h: 1.5 m Parapet h: 1 m	Window h: 2 m Parapet h: –	Window h: 2.5 m Parapet h: –
20	W1	W10	W19	W28
25	W2	W11	W20	W29
30	W3	W12	W21	W30
35	W4	W13	W22	W31
40	W5	W14	W23	W32
45	W6	W15	W24	W33
50	W7	W16	W25	W34
55	W8	W17	W26	W35
60	W9	W18	W27	W36

the reference building varies in each room to meet specific requirements of the living room, kitchen, and bedrooms. Nevertheless, the total height of windows and doors remains consistent at 2.15 m and 2.55 m for the rooms with access to daylight.

Glazing

As illustrated in Table 2, the sensitivity analysis encompasses 16 double-pane glazing options, which result from the combination of various variables. These variables include glazing type (clear, low emissivity), pane thickness

(4 mm, 6 mm), gap thickness (12 mm, 16 mm), and the filling of the gap between the panes (air, argon). Additionally, thermo-physical and optical properties of the glazing type, such as visible transmittance (T_{vis}), solar heat gain coefficient (SHGC), and thermal transmittance (U-value) are considered. The glazing scenarios are developed based on the published catalogue values of a leading glass manufacturer in Turkey (Şişecam 2021).

Window Frame

The impact of window frame material and profile dimensions on annual energy consumption is further examined through calculations in BES software applications. Simulations are conducted for both PVC and aluminum frames (with thermal break) across three different profile widths, as detailed in Table 3 (Çuhadaroğlu 2020; Saray 2020). It's important to note that the total U-value of the reference window system complies with threshold of 2.4 W/m^2K, as outlined in the TS 825 thermal insulation requirements for buildings standard (Turkish Standardization Institute 2013).

Solar Shading

Four external solar shading options were investigated for residential spaces with daylight access. These aluminum shading devices consist of overhangs and sidefins, with three different dimension possibilities. Notably, the reference building was simulated without external solar shading, but internal window shading is implemented in each scenario listed in Table 4.

Table 3 Window frame scenarios

Scenario No	Window profile width and material	U value (W/m^2K)
Reference	60 mm – PVC	1.64
F1	70 mm – PVC	1.30
F2	80 mm – PVC	0.95
F3	60 mm – Aluminum	3.09
F4	70 mm – Aluminum	2.10
F5	80 mm – Aluminum	1.70

2.4 Optimization Analysis

In the third phase, we examined the opening elements that demonstrated the most favorable performance in terms of annual primary energy consumption, including heating, cooling, and lighting loads, as well as carbon emissions and global cost. The optimization process was executed within the DesignBuilder software application tool. The analysis included a total of 11,520 potential design solutions for opening elements, achieved by combining 61 options across four design variables. Utilizing the genetic algorithm method, we conducted approximately 2500 simulations to pinpoint the optimal solutions within the design space, aligning with our research objectives.

Table 2 Glazing scenarios

Glazing Type	Scenario No	Double pane combination	Gap Filling	T_{vis} (%)	SHGC (%)	U value (W/m^2K)
Clear	Reference	4+12+4 mm	Air	80	75	2.9
	G1	4+16+4 mm	Air	80	75	2.7
	G2	4+16+4 mm	Argon	80	75	2.6
	G3	4+12+4 mm	Air	79	56	1.6
	G4	4+16+4 mm	Air	79	56	1.3
	G5	4+16+4 mm	Argon	79	56	1.1
	G6	4+12+4 mm	Air	71	44	1.6
	G7	4+16+4 mm	Air	71	44	1.3
	G8	4+16+4 mm	Argon	71	44	1.1
Low-e	G9	6+16+6 mm	Air	69	40	1.4
	G10	6+16+6 mm	Argon	69	40	1.1
	G11	6+16+6 mm	Air	63	43	1.4
	G12	6+16+6 mm	Argon	63	43	1.1
	G13	6+16+6 mm	Air	72	53	1.4
	G14	6+16+6 mm	Argon	72	53	1.1
	G15	6+16+6 mm	Air	80	64	1.4
	G16	6+16+6 mm	Argon	80	64	1.1

Table 4 Solar shading scenarios

Scenario No	External Shading	Internal Shading
Reference	–	Semi-open weave drapes, 3 mm Solar transmittance: 0.45 Solar reflectance: 0.25
S1	Overhang 0.5 m, 2 mm	
S2	Overhang 1 m, 2 mm	
S3	Overhang + sidefins 0.5 m, 2 mm	
S4	Overhang + sidefins 1 m, 2 mm	

2.5 Global Cost Calculations

Global cost calculations are conducted for the reference case and for the optimization solutions in line with the EN 15,459–1 standard (CEN 2017). To compute various categories of global cost, including initial investment cost, running costs, replacement costs, energy costs, and residual value, Eq. 1 is applied within the Excel program.

$$Cg(\tau) = CIn + \sum[\sum(Cy(\tau) + Cr(\tau)) - Vf, \tau(j)] \quad (1)$$

In the equation, Cg(τ) represents the global cost, *CIn* signifies initial investment costs, *Cy*(τ) denotes the yearly running costs, *C*r(τ) signifies replacement costs, and *Vf*, τ(*j*) represents the residual value for the combination of measures *j* at the end of the calculation period (τ). As evident in the formula, the financial calculations account for the time value of money by converting future incomes and expenditures into present value terms using a discount rate.

Based on the average value from the last 5 years (2018–2022) according to the Turkish Republic Central Bank statistics (2022), the following rates are employed: an inflation rate of 24.4%, a market interest rate of 17.5%, and a corresponding discount rate of 1.5%. Remarkably, Turkey's annual inflation rate reached its highest level in 2022 in the last two decades, which impacts the results of the global calculations.

The initial investment cost and product lifespans of were determined based on data collected from market research. Labor and transportation costs were sourced from the Construction and Installation Unit Prices book, published by the Turkish Republic Ministry of Environment, Urbanization and Climate Change (2022). The annual cost of energy consumption is calculated by multiplying the energy unit prices as of August 2022 (electricity: 2.17 TL/kWh; natural gas: 0.411 TL/kWh including TAX) by the corresponding end-use energy values (Uludagelektrik 2022; Bursagaz 2022). The increase in energy costs is assumed to be equal to the inflation rate.

3 Results

In the sensitivity analysis, a total of 61 solutions were simulated, derived from the matrix of design variables and options. Figure 4 presents the improvement ratio of the simulated parameters, including WWR, window size and configuration, glazing, and solar shading scenarios, in line with the research objectives. Solutions that result in a higher operational energy cost than the base case, such as options with aluminum frames or WWR of 45% and above, have been excluded. Additionally, solutions with a payback period longer than the building's expected lifetime (e.g., solar shading scenarios) are not represented in the graph. Solutions with a lower initial investment cost than the reference building are indicated by a negative value.

Following the sensitivity analysis, the optimization process is initiated. Figure 5 displays the reference building and the Pareto-front solutions in the design space. The X-axis represents the annual primary energy consumption (which includes heating, cooling, and lighting), while the Y-axis represents the global cost, considering initial investment, annual running, and replacement costs for transparent constructions and solar shading devices.

As part of the primary energy consumption and global cost analysis, we have identified 10 solutions that are considered the "most" optimal due to their advantages over the remaining solutions. Table 5 shows the selected combinations of the variables from the Pareto front. Rows highlighted correspond to the *energy-oriented, well-balanced,* and *cost-oriented* solutions. It is apparent that solutions in the design space often exhibit a trade-off, where efforts to reduce energy consumption tend to increase global cost. Conversely, solutions aimed at reducing global costs tend to lead to higher annual energy consumption and carbon emissions.

The results and notable changes regarding the reference building and optimal solutions are summarized as follows:

- The solutions demonstrate significantly improved performance compared to the reference building. For instance in the energy-oriented solution, there is a 24.8% reduction in energy consumption (30.4 kWh/m^2.a) and a 16.7% decrease in global cost (34.7 €/m^2.a) when compared to the reference building. Similarly, in the cost-oriented solution, a 20.7% reduction in energy consumption (25.4 kWh/m^2.a) and a 23.4% decrease in global cost (48.7 €/m^2.a) can be achieved compared to the reference building.
- The annual carbon emissions of the reference building are 24.7% higher (8.2 kg/m^2.a) compared to the energy-oriented solution and 22% higher (7.3 kg/m^2.a) compared to the cost-oriented solution.

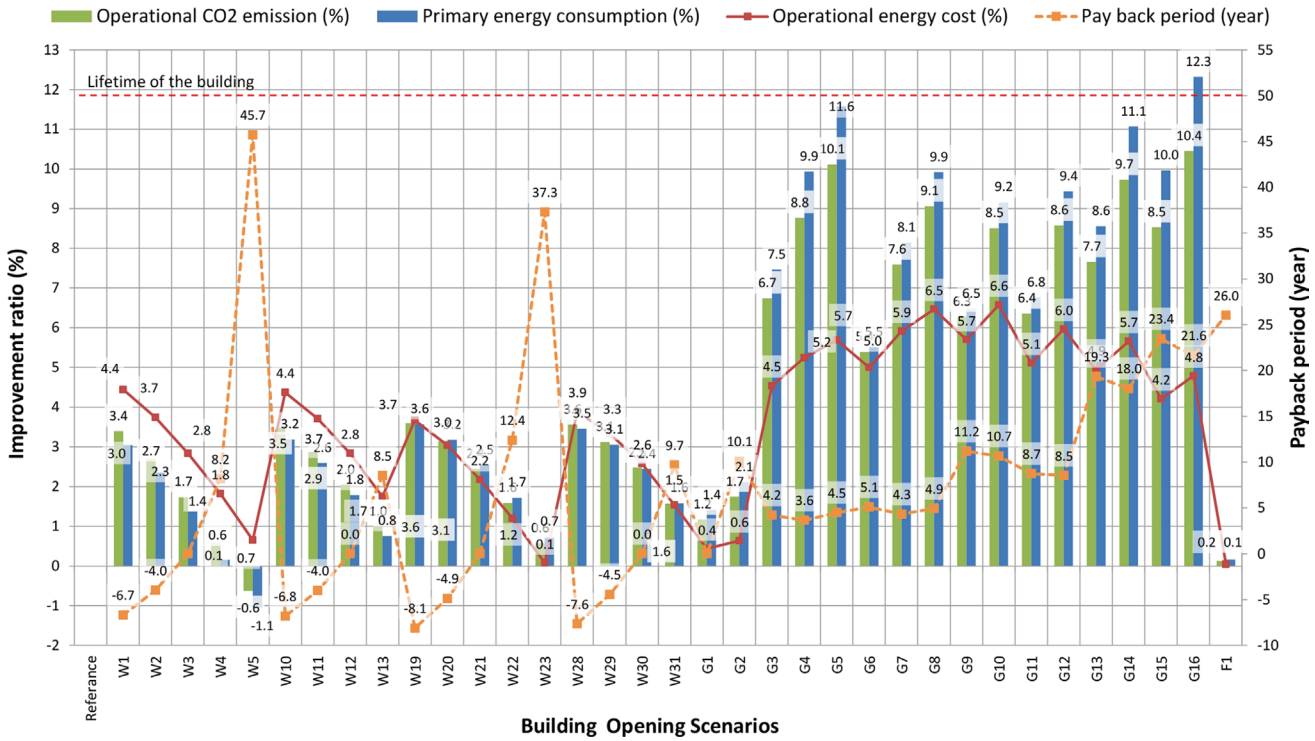


Fig. 4 Sensitivity analysis results: Improvement ratio and payback period for the scenarios

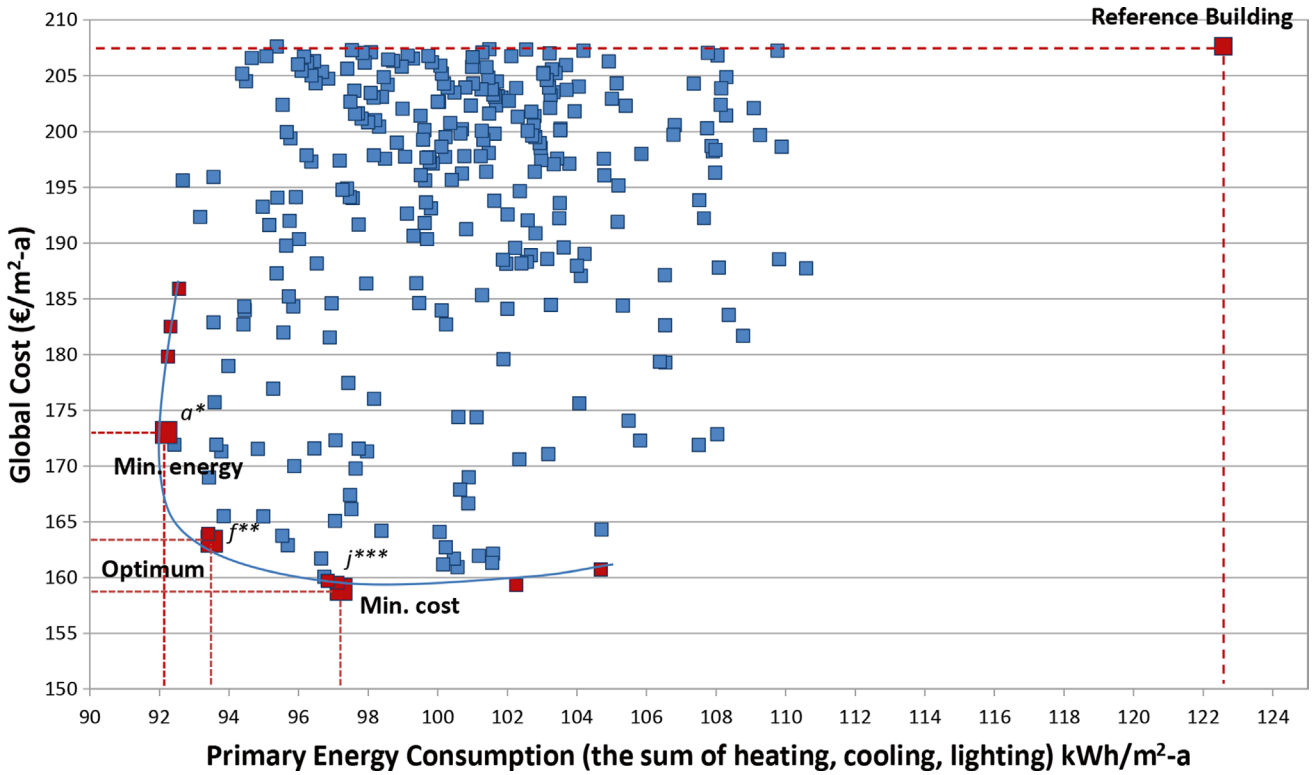


Fig. 5 Results of the optimization analysis and the selected solutions from the Pareto front

Table 5 The simulation results of the selected design combinations from the Pareto front

	Order	WWR (%)	Window h. Parapet h.	Frame size material	Glazing type	Primary energy (kWh/m².a)	Global cost € (TL)/m².a	CO_2 (kg/m².a)
< Lower Energy >	*a**	*40*	*1.5 m – 1 m*	*60 mm PVC*	*G16*	*92.2*	*173€ (3,170 TL)*	*24.9*
	b	45	2.5 m – 0 m	60 mm PVC	G16	92.2	180€ (3,294 TL)	25.0
	c	50	*1m – 1 m*	60 mm PVC	G16	92.3	183€ (3,344 TL)	25.0
	d	35	1.5 m – 1 m	80 mm PVC	G16	92.4	172€ (3,149 TL)	24.8
	e	30	1.5 m – 1 m	60 mm PVC	G16	93.4	164€ (3,003 TL)	25.0
	*f***	*30*	*1m – 1 m*	*60 mm PVC*	*G16*	*93.5*	*163€ (2,991 TL)*	*25.0*
	g	20	2m – 0 m	70 mm PVC	G16	96.7	160€ (2,933 TL)	25.7
	h	20	2m – 0 m	60 mm PVC	G16	96.8	160€ (2,926 TL)	25.8
	i	30	2.5m – 0 m	70 mm PVC	G8	97.1	160€ (2,923 TL)	25.7
< Higher Energy	*j****	*30*	*2.5m – 0 m*	*60 mm PVC*	*G8*	*97.2*	*159€ (2,912 TL)*	*25.8*
	a: energy-oriented solution, : f** well-balanced solution, j***: cost-oriented solution*							

- In comparison to the reference building, we achieved substantial savings: a 23.7% reduction in primary energy consumption (29.1 kWh/m2), a 21.4% decrease in global costs (44.4 €/m2), and a 24.4% cut in carbon emissions (above 8 kg/m^2) for the well-balanced solution.
- The energy-oriented solutions exhibit a relatively higher quantity of glazing compared to other solutions (see Table 5). However, the overall glazing ratio of the well-balanced and cost-oriented solutions remain the same as that of the reference building (30%) for the temperate climate region.
- Drawing precise conclusions regarding window configurations based on the selected design options in Table 5 proves challenging, primarily due to the presence of four different possibilities for glazing and parapet height within in the list.
- The thermal transmittance coefficient (U-value) of all glazing scenarios considered in the study is lower than that of the reference building, leading to savings in heating energy, primary energy consumption, and related energy costs and CO_2 emissions (see Fig. 4).
- The solar heat gain coefficient value of all glazing scenarios (except G1 and G2) is lower than that of the glazing option in the reference building (SHGC < 0.75). While this reduces the gain of unwanted solar radiation through the transparent component and lowers cooling energy consumption, it also leads to reduced utilization of passive solar energy by the residential spaces during winter.
- The daylight transmittance value of all glazing scenarios (except G15 and G16) is lower than that of the double glazing option in the reference building (Tvis < 0.80). This results in a slight increase in lighting energy consumption in the cost-oriented optimal solutions (i, j).
- The cost-oriented solution utilizes the glazing option of G8 (T_{vis}: 0.71, SHGC: 0.44, U: 1.1 W/m^2K, 4 + 16 + 4 mm glazing configuration with argon filling), while the energy-oriented and well-balanced solutions recommend G16 (T_{vis}: 0.80, SHGC: 0.64, U: 1.1 W/m^2K, 6 + 16 + 6 mm glazing configuration with argon filling).
- As indicated by the optimization analysis, it is recommended to maintain PVC frames with a 60 mm profile width for the optimal solutions (a, f, j).
- Window shading devices are not preferred in any of the selected design options since the primary living spaces of the case building are predominantly oriented to the northwest.

4 Conclusions

The study aimed to establish cost-optimal, energy-efficient solutions for opening elements in an exemplary residential building located in a temperate climate region in Turkey. This goal was achieved through the implementation of building energy simulations and BES-based optimization analysis within the design process. The optimization analysis yielded solutions within the design space that outperformed the reference case. These solutions were categorized as energy-oriented solution, characterized by the lowest primary energy consumption, "cost-oriented solution," aimed at achieving maximum energy efficiency at minimal cost, and a "well-balanced" or, more specifically, an energy-efficient cost-optimum solution for the temperate-humid climatic region. We established a model to depict the relationship between primary energy consumption and global cost, and the findings were visually presented using an XY scatter plot in MS Excel.

The results underscore the significance of modifying the transparency ratio, window size, configuration, and glazing properties of the building in reducing total annual primary energy consumption and long-term costs. These findings can provide guidance to architects during the detailed design stages, enabling them to focus on specific opening solutions. Façade opening design can further be refined based on the results of the optimization analysis, considering the primary objectives. It's also important to note that optimal solutions may differ depending on the building's orientation and climatic conditions.

For future studies, it is recommended to explore additional energy efficiency measures including the utilization of high-tech building envelope materials and renewable energy sources, which could be incorporated into the optimization process. In addition, there is potential for conducting annual climate-based daylight modeling, which could be a valuable focus for future research. Complementary daylight simulations could be run to evaluate the availability of daylight for proposed opening solutions and verify their compliance with the requirements outlined in daylight standards.

Acknowledgements The authors gratefully acknowledge the financial support provided by Bursa Uludag University Department of Scientific Research Projects, Project Development and Coordination Unit. Specifically, this research was conducted as part of the corresponding author's PhD project DDP (MM)-2019/6, entitled "A Design Support Model to Provide Thermal and Visual Comfort Conditions in Residential Buildings and Energy Optimization in the Building Envelope."

Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

References

Ascione F, De Masi RF, de Rossi F, Ruggiero S, Vanoli GP (2016) Optimization of building envelope design for nZEBs in Mediterranean climate: Performance analysis of residential case study. Appl Energy 183:938–957. https://doi.org/10.1016/j.apenergy.2016.09.027

Bursagaz (2022) Natural gas unit price in Bursa Turkey. Retrieved August 10, 2022, from https://www.bursagaz.com/konut-fiyat-bilgileri

Climate Onebuilding (n.d.) *Repository of free climate data for building performance simulation*. [Dataset: Turkey province-based climate files]. Retrieved August 2, 2021, from https://climate.onebuilding.org/WMO_Region_6_Europe/TUR_Turkey/index.html#IDBU_Bursa-

Çuhadaroğlu (2020) Aluminium door and window systems product catalogue. Retrieved May 7, 2022, fromhttps://cms.cuhadaroglu.com/uploads/Aluminium%20Door%20and%20Window%20Systems%202020.pdf

European Central Bank (2022) Euro foreign exchange rates. Retrieved August 8, 2022, fromhttps://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-try.en.html

European Commission (EC) (2019) European Green Deal (COM/2019/640) final report. Communication from the Commission to The European Parliament, The European Council, The Council, The European Economic and Social Committee and The Committee of The Regions. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN

European Commission (EC) (2021) Fit for 55: delivering the EU's 2030 Climate Target on the way to climate neutrality. COM(2021) 550 final. Brussels.https://eur-lex.europa.eu/legal content/EN/TXT/?uri=CELEX%3A52021DC0550

CEN (European Committee for Standardization) (2017) CEN EN 15459-1:2017 Energy performance of buildings - Economic evaluation procedure for energy systems in buildings - Part 1: Calculation procedures, Module M1–14.https://standards.iteh.ai/catalog/standards/cen/ff0c5a0e-d363-40ab-80b3-7d7cc28b19a5/en-15459-1-2017.

European Union (EU) (2009) Directive 2009/28/EU. European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC.

European Union (EU) (2010) Directive 2010/31/EU. European parliament and of the council of 19 may 2010 on the energy performance of buildings (recast). Off J Eur Union 2010

European Union (EU) (2012) Directive 2012/27/EU. European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC

European Union (EU) (2018a) Energy Efficiency Directive (EED) 2018/2002. European Parliament and The Council of 11 December 2018 amending Directive 2012/27/EU on energy efficiency. Off J Eur Union

European Union (EU) (2018b) Directive 2018/2001. European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources (recast)

European Union (EU) (2018c) EPBD recast Directive 2018/844/EU. European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency.

International Energy Agency (IEA) (2021) Turkey 2021 energy policy review - IEA world energy statistics and balances (database), Paris. https://www.oecd-ilibrary.org/energy/turkey-2021-energy-policy-review_0633467f-en

Liu H, Zhang Z, Ma X, Lu W, Li D, Kojima S (2021) Optimization analysis of the residential window-to-wall ratio based on numerical calculation of energy consumption in the hot-summer and cold-winter zone of China. Sustainability 13:6138. https://doi.org/10.3390/su13116138

Maleki A, Dehghan N (2021) Optimization of energy consumption and daylight performance in residential building regarding windows design in hot and dry climate of Isfahan. Sci Technol Built Environ 27(3):351–366. https://www.tandfonline.com/doi/full, https://doi.org/10.1080/23744731.2020.1812294

Ochoa CE, Aries MBC, Loenen JE, Hensen JLM (2012) Considerations on design optimization criteria for windows providing low energy consumption and high visual comfort. Appl Energy 95:238–245. https://doi.org/10.1016/j.apenergy.2012.02.042

Saray. (2020). PVC Window Systems Product Catalogue Retrieved May 7, 2022, from https://saray.com/en/media/catalogs-and-brochures/saray-selective-pvc-windows/

Şişecam (2021) Architectural glass product catalogue. Retrieved May 6, 2022, fromhttps://duzcam.sisecam.com/en/architectural-glass/professional-product-catalog

Tatbikat Architecture Office archive (n.d.) Site plan of the apartment blocks

Tuhus-Dubrow D, Krarti M (2010) Genetic-algorithm based approach to optimize building envelope design for residential buildings. Build Environ 45(7):1574–1581. https://doi.org/10.1016/j.buildenv.2010.01.005

Turkish Republic Central Bank statistics (2022) Inflation rate and market interest rate in Turkey between 2018 and 2022. Retrieved July 30, 2022, from http://www.tcmb.gov.tr

Turkish Republic Ministry of Energy and Natural Resources (2016). National Energy Efficiency Action Plan. Ankara. Retrieved April 20, 2020, fromhttps://enerji.gov.tr/

Turkish Republic Ministry of Environment, Urbanization and Climate Change (2022) Unit prices of construction and installations. Retrieved April 14, 2022, from https://yfk.csb.gov.tr/birim-fiyatlar-i-100468

Turkish Republic Ministry of Environment, Urbanisation and Climate Change (2020) Primary energy and carbon emissions conversion factors for natural gas and electricity. Retrieved February 13, 2022, from https://meslekihizmetler.csb.gov.tr/elektrik-enerjisinin-birincil-enerji-ve-sera-gazi-salimi-katsayilari-2021-yilindan-itibaren-kullanilmak-uzere-guncellenmistir-duyuru-411795

Turkish Republic Ministry of Public Works and Settlement (2007) Energy Efficiency Law (EN-VER). Number: 5627, 18.04.2007. Republic of Turkey Official Gazette, 26510, Ankara

Turkish Republic Ministry of Public Works and Settlement (2008) Building Energy Performance Regulation (BEP), Republic of Turkey Official Gazette, 27075, Ankara

Turkish Republic Ministry of Public Works and Settlement (2010) Building Energy Performance National Calculation Method (BEP-TR). Republic of Turkey Official Gazette, 27778, Ankara.

Turkish Standardization Institute (TSE) (2013) TS 825 Thermal insulation requirements for buildings

Turkish Statistical Institute (TurkStat) (2022) Building Permit Statistics - IV. Quarter: October – December.https://data.tuik.gov.tr/Bulten/Index?p=Building-Permits-Quarter-II:-April---June,-2022-45830&dil=2

Uludagelektrik (2022) Electricity unit price in Bursa Turkey. Retrieved August 10, 2022, from https://www.uludagelektrik.com.tr/

Xue P, Li Q, Xie J, Zhao M, Liu J (2019) Optimization of window-to-wall ratio with sunshades in China low latitude region considering daylighting and energy saving requirements. Appl Energy 233–234:62–70. https://doi.org/10.1016/j.apenergy.2018.10.027

Zhai Y, Wang Y, Huang Y, Meng X (2019) A multi-objective optimization methodology for window design considering energy consumption, thermal environment and visual performance. Renewable Energy 134:1190–1199. https://doi.org/10.1016/j.renene.2018.09.024

Effect of Urban Design for Residential Complexes on the Efficiency of Environmental Performance and Carbon Emissions

Ola Samy Ali Alhinawy

Abstract

Carbon emissions have greatly affected global warming and global climate change, which made environmental protection institutions repeatedly call for the importance of reducing carbon emissions. Urbanization is considered one of the most important causes of the increase in these carbon emissions because it consumes non-renewable energy to reach thermal comfort rates inside and outside buildings, in addition to the carbon emissions resulting from the building and finishing materials used in it. Thus, urban design is considered one of the most important factors affecting the quality of life in residential complexes, not only on the extent of road planning and open and closed spaces but also on the scope of the impact of urban design of residential complexes on environmental performance and the proportion of carbon dioxide emissions, which is what the research deals with to study. The research aims to study the effect of urban design on thermal performance and the proportion of carbon dioxide emissions into the air within residential complexes. The importance of the research is to shed light on the importance of urban design for residential complexes in influencing the efficiency of environmental performance within those communities, which will be directly reflected on human health, and the surrounding environment, and its repercussions on global warming to achieve more sustainable urban communities. The research uses the experimental and analytical approach through an urban design for three models of urban communities, in which the same variables are combined (heights, percentage of built areas, percentage of open areas, percentage of green areas, type of plants used in open spaces, finishing materials for the outer cover of the building as well as roads and corridors with open spaces), with one variable differs in it, which is the urban design for the distribution and shape for (three urban design models), with the aim of knowing the extent of the impact of the urban design of buildings on environmental performance and the proportion of carbon dioxide, as previously mentioned Envi-Met program was used to simulate this relationship. The results of the research have been concluded that the model (3) (the courtyard design) has the best thermal and environmental performance among the three models (from achieving thermal comfort rates and carbon dioxide emissions), followed by model (2) (Stripe and vertical design).

Keywords

Sustainability · Environmental design · Thermal comfort · Climate change · Carbon Emissions · Envi-Met · Sustainable communities · Courtyard

1 Introduction

Urban communities predominantly rely on non-environmentally friendly energy sources within their buildings to attain the desired thermal comfort conditions. This substantial reliance significantly amplifies the volume of carbon emissions released into the atmosphere. Consequently, when urban communities can achieve thermal comfort naturally, it results in a reduction in the reliance on unsustainable energy sources (Rosenzweig et al. 2018). In view of the urban design within numerous cities across the globe, particularly those characterized by high population density, a prevailing observation emerges: a deficiency in incorporating comprehensive environmental design standards aimed at enhancing the thermal performance of these urban communities. While urban planners and designers often exhibit a significant interest in

O. S. A. Alhinawy (✉)
Department of Architectural Engineering, Faculty of Engineering, Menoufia University, Al Menoufia, Egypt
e-mail: Olaelhenawy20@gmail.com; OLA.Alhenawy@sh-eng.menofia.edu.eg

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_10

crafting urban designs that accentuate aesthetic considerations and adhere to established design standards, there exists a notable gap in systematically evaluating the consequential effects on the thermal and environmental aspects of these urban settings. But most urban planners and designers prioritize aesthetic and design standards when formulating urban spaces, sometimes overlooking the imperative to holistically assess the implications of their design choices on the thermal efficiency and environmental resilience of the communities they shape. This divergence of focus can result in missed opportunities to harness urban design as a mechanism for fostering optimal thermal performance and sustainable environmental outcomes within densely populated cities (Wheeler 2013).

It is essential for urban planning and design practices to evolve towards a more integrated approach that synthesizes aesthetics, design standards, and the vital considerations of thermal comfort and environmental sustainability. By bridging this gap and incorporating environmental design standards into the urban design process, planners and designers can proactively mitigate the adverse impacts of urban heat islands, energy consumption, and other environmental challenges that can emerge from inadequately considered design choices. This integrated approach not only ensures the aesthetic appeal of urban communities but also elevates their overall quality of life, resilience, and long-term sustainability.

By leveraging simulation programs, the research endeavors to concretely demonstrate the intricate relationship between urban design and carbon emissions reduction. These programs offer a unique platform to model and assess diverse urban design scenarios, allowing researchers to quantitatively showcase the tangible impact of varying design choices on carbon footprints. The research strives to exemplify that meticulous urban design interventions have the potential to tangibly curb carbon emissions, thus directly contributing to the broader aim of realizing carbon–neutral urban landscapes.

In summary, the research seeks to emphasize the importance of urban design in mitigating carbon emissions, to reach cities with better environmental efficiency. Through the use of environmental simulation softwares for urban design, which seek to find urban design with better environmental efficiency and lower carbon emissions, in addition to enhancing the discourse on the pivotal role of urban design in shaping sustainable and low-carbon urban communities.

1.1 Literature Review

Several previous studies have concentrated on the relationships among urban design, population growth, carbon emissions, and climate impacts. The following summarizes essential findings from some of these studies.

Urban design is considered the link between urban planning and architectural design, as it deals with (spaces, paths, intersections, squares, buildings, green spaces, etc.), and through it, the character of the area and the personality of the place, and the connection between the building and the surrounding space can be achieved. Urban design begins from where urban planning ends, which is concerned with the formation of the city in only two dimensions, and it can be expressed with simplified maps, that is, it is concerned with the physical composition of the city's functions without much attention to the aesthetic and sensory aspects. And urban design ends where the architectural design begins, which is concerned with the building itself and how it performs its function efficiently through functional analysis and attention to proportions. Thus, urban design seeks to achieve maximum efficiency and quality for open and closed spaces, paths, and roads for urban communities. (Urban d. 2022).

The unstudied urban design is considered one of the problems that contribute to increasing the problem of pollution and global warming, especially with the large increase in the population around the world, which leads to unplanned urban areas. Within the context of the world's unplanned urban communities, as highlighted by the United Nations Human Settlements Program, a significant projection indicates that approximately 90% of urban growth is poised to transpire in regions characterized as less developed. The expansion of urban areas in these regions predominantly lacks comprehensive planning, thereby contributing to the persistent proliferation of informal settlements that are primarily inhabited by poor populations. These informal settlements, presently providing shelter for approximately one billion individuals, are notably susceptible to the exacerbating impacts of climate change (United Nation 2020). When considering the broader trajectory of global urbanization trends, it is noteworthy that urban populations have experienced a tripling effect within a relatively short time frame. With this momentum in mind, it is projected that in the imminent future, over half of the global populace will be residing within urban environs. This trajectory underscores the critical significance of addressing the challenges associated with unplanned urban expansion and the subsequent proliferation of informal settlements, particularly in the less developed regions (United Nations 2004).

Many studies have debated surrounding the influence of urban planning on climate dynamics and carbon emissions have been the subject of numerous studies, generating considerable debate. Among the pivotal factors identified are the spatial configurations of land use and the built environment, the intricate patterns of street networks, and the dimensions of constructed blocks. Notably, these aspects function as carbon locks, presenting formidable challenges to rapid reversal or alteration. The enduring nature of these elements underscores the critical importance of examining their impact on carbon emissions (Seto et al. 2016). Urban

design occupies a pivotal role in the overarching pursuit of establishing carbon–neutral cities—a paramount objective that has garnered global attention and engagement. As cities across the world endeavor to realize this aspiration, the significance of urban design as a cornerstone emerges vividly. A crucial outcome of adept urban design is the facilitation of urban communities characterized by diminished carbon emissions (Karen et al. 2021).

The planning and urban design have the potential to mitigate the adverse impacts of global climate change. Alterations to outdoor environments can enhance thermal comfort. Methods for integrating microclimate data into design should be simplified and made readily accessible to designers. Additionally, disseminating knowledge concerning the influence of landscape planning and design on climate should effectively reach decision-makers at various administrative levels (Robert 2011).

Sustainability has become an integral element within the realm of urban planning and urban design, providing a solid foundation for research on climate change. In his work on sustainable communities, Blakely advocates for the importance of innovative and forward-thinking planning approaches. He presents planning principles that aimed at guiding practical application, notably by integrating land use, social considerations, and governance into a unified framework, thereby fostering a novel (Blakely and Edward 2004).

Furthermore, Beatley and Manning describe novel settlement patterns that have the potential to effectively address the challenges posed by climate change. These insights contribute to a growing body of knowledge that underscores the critical role of sustainable urban planning and urban design in mitigating and adapting to the impacts of climate change (Beatly et al. 1998). Previous studies on this subject have reached a consensus regarding the influence of urban design on carbon emissions and climate outcomes.

1.2 Research Aims

The research aims to study the effect of urban design on thermal performance and the proportion of carbon dioxide emissions into the air within residential complexes.

1.3 Research Importance

The importance of the research is to shed light on the importance of urban design for residential complexes in influencing the efficiency of environmental performance and carbon dioxide emissions within those communities, which will be directly reflected on human health, energy consumption, and then on the surrounding environment, and their repercussions on global warming phenomena.

1.4 Research Methodology

The research uses the experimental and analytical method by designing three models for residential communities in which the same variables meet (heights, percentage of built areas, percentage of open areas, percentage of green areas, type of plants used in open spaces, finishing materials for the building's outer cover, as well as roads and corridors with open spaces), with one variable differs in it, which is the urban design for the distribution and shape of residential buildings, with the aim of knowing the extent of the impact of the urban design of buildings on environmental performance and the proportion of carbon dioxide, as previously mentioned using Envi-Met program to simulate this relationship.

2 Theory/Calculation

2.1 Thermal and Environmental Simulation

The simulation program Envi-Met 4-4-4 was used, and the following data were adopted for the simulation on the Envi-Met program: Determining the city of Cairo, date: 9-21-2020.

1.1. Envi-Met 4-4-4

ENVI-Met technology allows to create sustainable living conditions in an ever-changing environment, by identifying roof types and building materials, as well as vegetation on walls and ceilings, to scientifically analyze the effects of design measures on the local environment and to help adapt to environmental factors.

ENVI-Met takes the interactions of wind, green space, solar radiation, buildings, and many other factors into account—and helps us design climate-resilient and sustainable cities. ENVI-Met program has been used for over 10 years to study the effect of vegetation cover in built environments. It is one of the few programs that realistically simulate the most important climatic processes—such as interactions between soil, vegetation, and the atmosphere—in urban environments, thus analyzing thermal comfort in cities (Envi-met 2022).

Thermal simulation was done for:

a. Celsius temperature
b. Relative humidity

c. Wind speed
d. Carbon dioxide
e. CO_2 mg/m^3 (milligram per cubic meter): This unit expresses the concentration in one cubic meter of air (equivalent to 1L or 1000 mL) of a substance in terms of its mass (measured in milligrams). It is primarily used for particle-like substances, and only rarely for gaseous concentrations (Gastec. 2022).

The three models were designed and drawn using Autocade, to ensure that the built-up area, open spaces and green spaces of the three models have the same area in each of the three models with different urban design for each model to know the extent of the impact of urban design on carbon emissions and thermal performance efficiency, and then the models are exported to Envi-Met, and the correct scale is applied to reach the actual area designed by Autocade.

2.2 Simulated Urban Design Models

- Total area of land (600 × 360) = (216,000 m^2)
- Built-up area: 43,200 m^2
- North direction: up perpendicular to the main axis of the residential complex (perpendicular to the greatest dimension).
- Height of all residential buildings: 16 m

CO_2 levels in outdoor air typically range:

Carbon dioxide (CO_2) is a colorless, odorless, non-flammable gas that naturally occurs in the atmosphere. CO_2 is produced by body metabolism and is a normal component of exhaled breath. It also results from the burning of fossil fuels and natural sources such as volcanic eruptions. CO_2 levels in outdoor air typically range from 300 to 400 ppm (0.03% to 0.04%) but can be as high as 600–900 ppm in metropolitan areas (osha.gov, 2022). Figure 1 shows the safe levels of (CO_2) for some of the places where we live.

Model No. 1

The first model was designed using of separate residential buildings with a built-up area of land (24 × 24), (24, 36), distributed in a U-shape, and green spaces were placed in the middle, and corridors were cut between the residential buildings and each other (corridor width 12 m), in addition to Putting open spaces with vegetation cover between each grouping of the residential buildings to create a stream of air between the buildings (24 m wide), and distributing in parallel with each other, and vertically on the main axis of the housing as a whole, As shown in Fig. 2.

Model No. 2

The second model was designed using connected residential buildings with striped built-up area of land (24 × 120), distributed in a stripe and perpendicular to the main axis of the residential assembly, striped, green spaces were also placed parallel to the residential buildings, the width of the corridors and the green strip between each building and the other (12 m). In addition to placing open spaces with vegetation cover between each grouping of residential buildings (36 m wide) and (12 m for corridors), and the distribution of those communities parallel to each other, and perpendicular to the main axis of the residential community are shown in Fig. 3.

Model No. 3

The residential complex was designed by dividing it into eight residential buildings interspersed with an open courtyard for a few built-up areas exposed to the sun. Studying the environmental impact of the courtyard on urban design, so that the external dimensions of the residential building were (96 × 84), and the dimensions of the courtyard were (48 × 36). The inner courtyard was covered with an open cover. Corridors and striped, green spaces were placed parallel to each residential building (with a width of 12 m for green space, and 12 m for each of the corridors separately), as shown in Fig. 4. The residential buildings, green spaces,

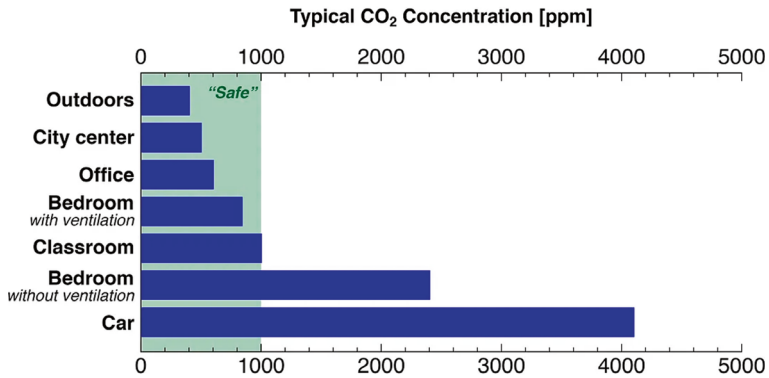


Fig. 1 Typical CO_2 levels in everyday places. Many of the places we live have CO_2 concentrations far above the recommended "safe" levels (Jean 2016).

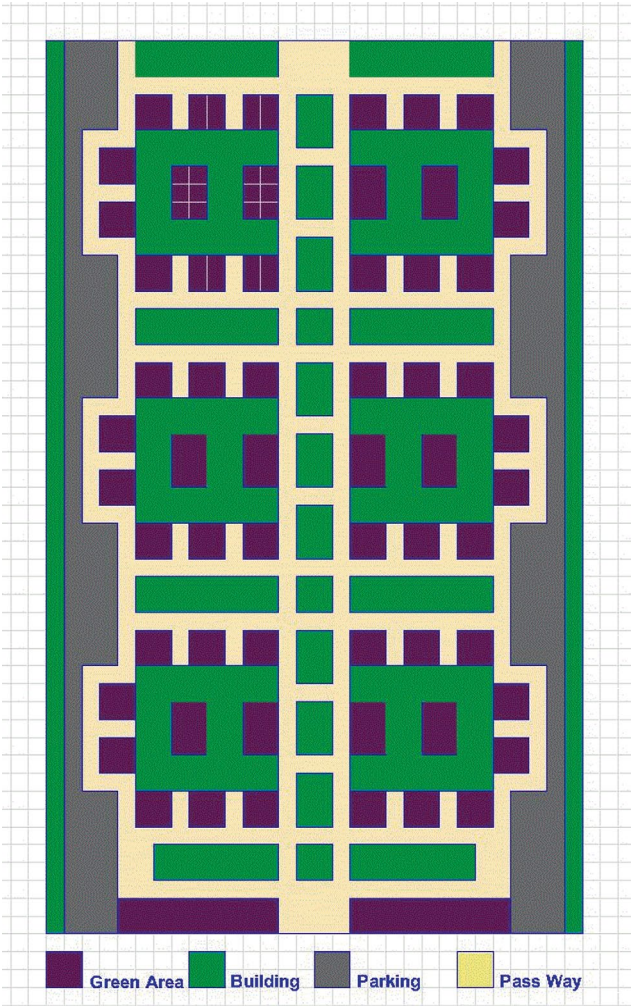


Fig. 2 Model No. 1. *Source* Researcher

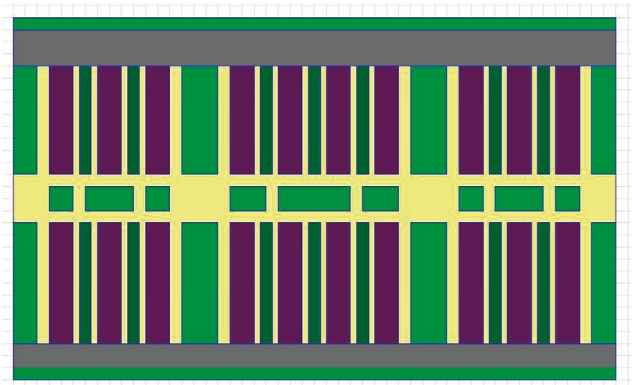


Fig. 3 Model No. 2. *Source* Researcher

and strip lanes are perpendicular to the main axis of the residential complex. The width of the main axis of this model has been increased from the previous models (to be 96 m wide, including 48 m green spaces).

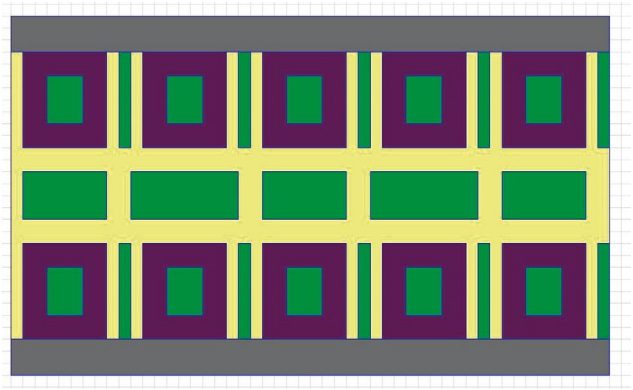


Fig. 4 Model No. 3. *Source* Researcher

3 Discussion

Below is a display to discuss the thermal simulation results.

3.1 Model No. 1

The temperatures of Model No. (1) recorded a noticeable increase (39.5 C), and a clear contrast between day and night temperatures with a difference of (22 C). This may be due to the presence of spaces between buildings and each other, which led to the exposure of large areas to sunlight during the day, and an increase in the flow of winds at night. The humidity also increased significantly to (94.1%), Figs. 5 and 6 shows the thermal simulation results for temperature and relative humidity, while the wind speed around the buildings decreased to (0.01 M/S), but it created some air currents between the buildings as shown in Fig. 8 (the wind ratio between the buildings ranges from (0.4:0.8))). It rose on the main axes parallel to the buildings to reach an average of (2.7 M/S), as shown in Fig. 8. Carbon emissions increased to 380.3 PPM as shown in Fig. 7.

Table 1 shows the maximum and minimum values of the environmental and thermal simulations of Model (1) through temperatures, relative humidity, wind speed, and the percentage of carbon dioxide emissions.

3.2 Model No. 2

In the environmental and thermal simulations of model (2), the temperatures were significantly lower than in model (1), with the maximum temperature reaching (32.5 C), Fig. 9 shows the temperature decrease as we head towards the west, and the humidity percentage also decreased, but not noticeably, to reach the maximum humidity of (93.5%) as shown in Fig. 10. The wind speed on the longitudinal line

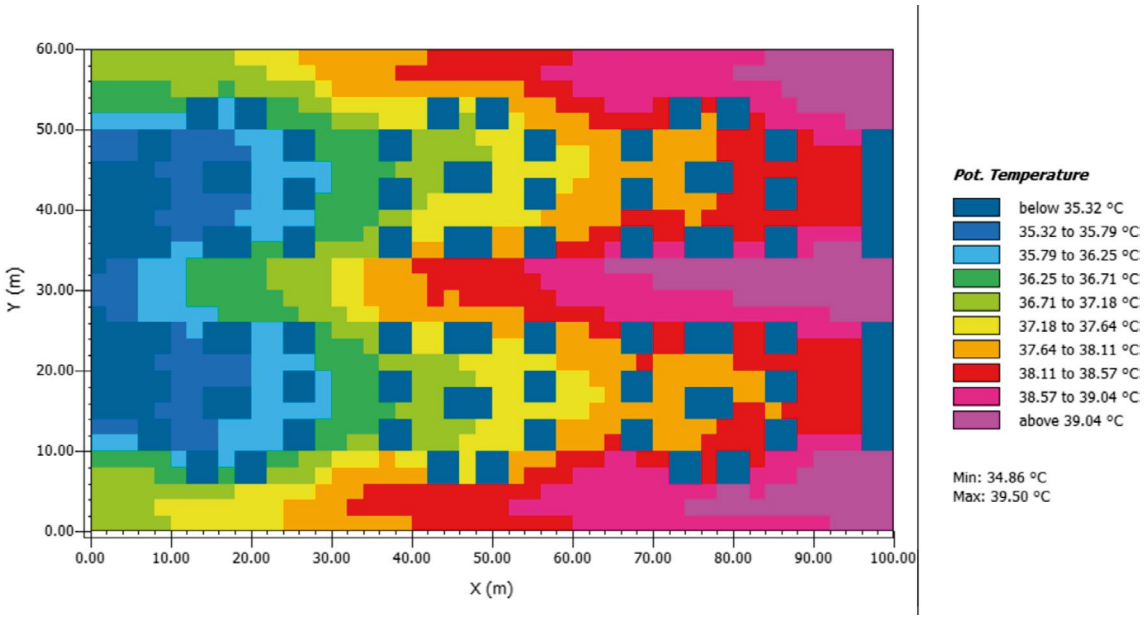


Fig. 5 Simulation of temperatures for a model (No. 1). *Source* Researcher

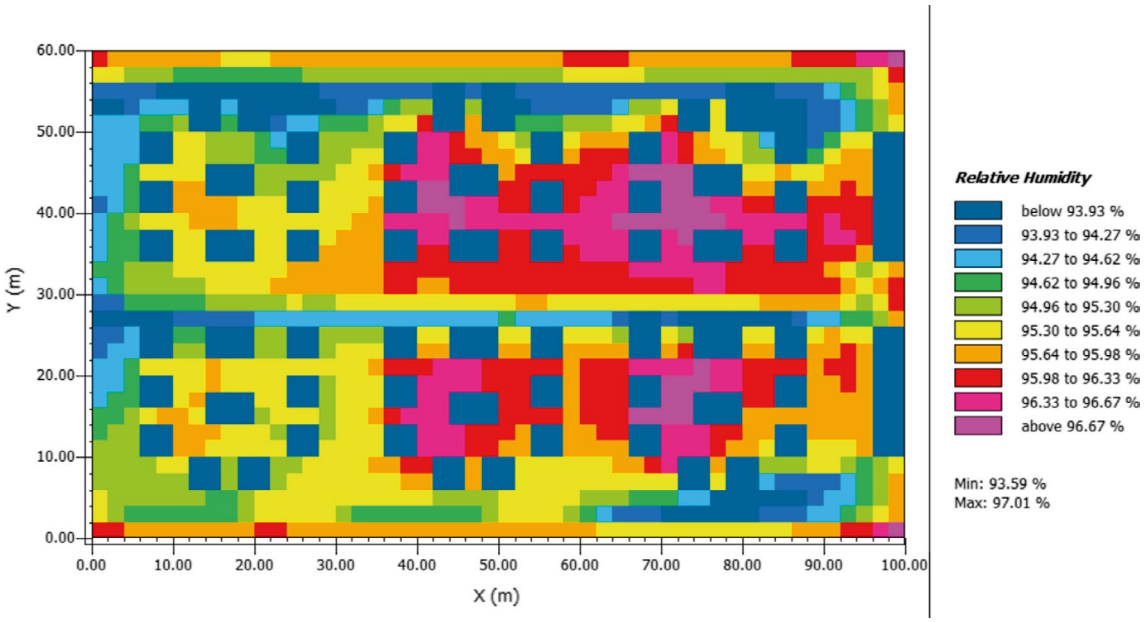


Fig. 6 Relative humidity simulation for a model (No. 1). *Source* Researcher

perpendicular to the buildings also decreased to (0.01 M/S). Where the strip shape of the buildings led to the formation of air deflectors, which significantly reduced the air velocity, as shown in Fig. 12.

The carbon dioxide emissions also decreased from what they are in Model (1), to record its maximum value (367.7 PPM), with a difference of (13 degrees compared to Model 1), Fig. 11 shows part of the simulation results for carbon emissions that were conducted throughout the day. Table 2 shows the maximum and minimum values of the environmental and thermal simulations of Model (1).

3.3 Model No. 3

The environmental and thermal simulations of Model (2) show that its temperatures are significantly lower than in Model No. (1), with a maximum temperature of 32.2 C.

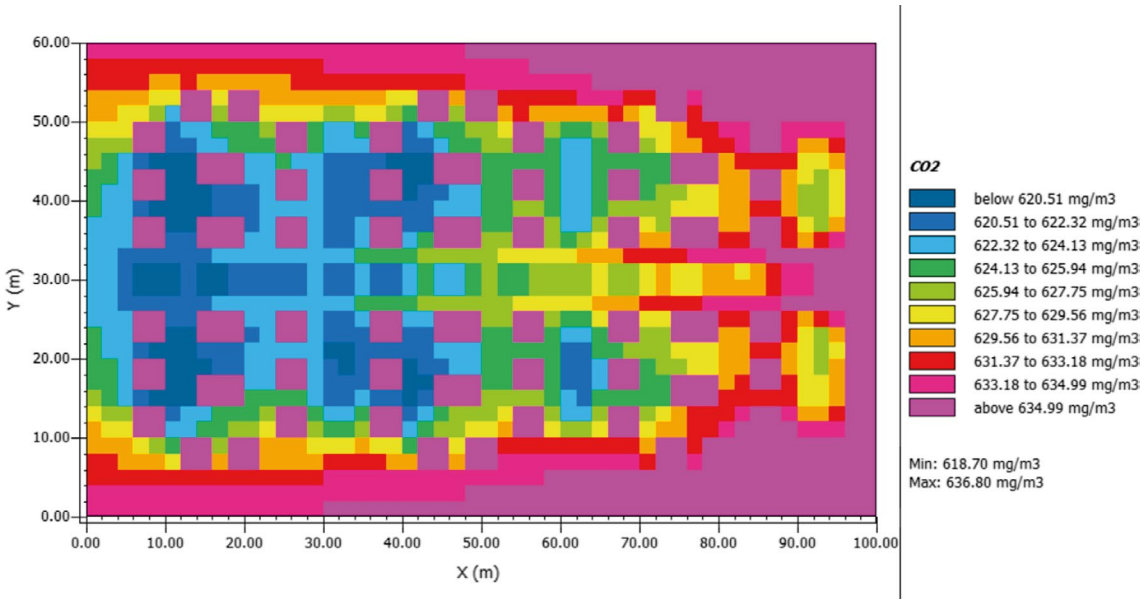


Fig. 7 Carbon dioxide simulation of emissions for a model (No. 1). *Source* Researcher

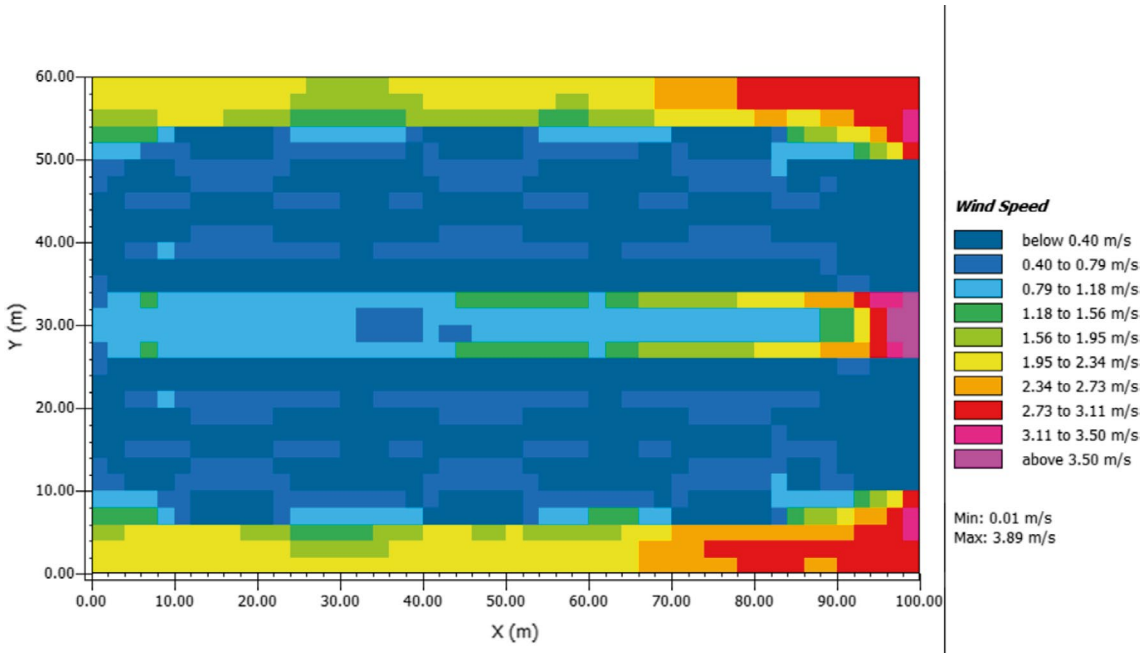


Fig. 8 Wind speed simulation for a model (No. 1). *Source* Researcher

However, the temperatures decreased significantly around the buildings, and rose in the main axis parallel to the buildings, which is the opposite of what happened in both models (1,2), which will greatly help in improving the thermal comfort rates inside the buildings, as shown in Fig. 13.

The humidity also decreased, but not noticeably, to reach the maximum humidity (93%), and the wind speed on the longitudinal line perpendicular to the buildings decreased to (0.01 M/S). Where the strip shape of the buildings led to the formation of air deflectors, which significantly reduced the air velocity.

Carbon dioxide emissions also decreased from what they were in model (1) to record their maximum value (367.7 parts per million), a difference of (13 degrees also from model (1)), as Table 3 shows the maximum and minimum values of for the environmental and thermal simulation of model (3). Figure 14 shows the maximum and minimum percentages of relative humidity, while Fig. 15 shows the carbon emissions values.

Table 1 Thermal simulation results for model 1

model 1	co2 (mg\m3)		co2		Temp		humidity		wind speed m\s	
Time	Min	Max	min	Max	min	Max	Min	max	Min	Max
12:00 AM	636.5	636.8	346	347.1	21.2	22.3	67	75.5	0.09	3.2
2:00 AM	636.5	636.8	345.6	346.8	18.9	19.6	67.8	76.3	0.07	3.19
4:00 AM	636.5	636.8	345.3	346.5	16.7	17.8	70.5	83.2	0.05	3.21
6:00 AM	636.5	636.8	345.2	346.4	17.8	19.1	76.7	87.2	0.04	3.22
8:00 AM	613.7	636.8	346.9	354.1	22.2	22.9	86.6	89.1	0.01	3.19
10:00 AM	615.7	636.8	352.9	369.7	29.7	33.3	87.9	94.1	0.01	3.46
12:00 PM	618.7	636.8	360.3	377.5	33.1	38	85.4	91.3	0.01	3.7
2:00 PM	620.8	636.8	364.4	380.3	34.9	39.5	83.3	90.6	0.01	3.89
4:00 PM	624	636.8	366.2	378.9	33.2	38.6	81.1	88.5	0.01	4.07
6:00 PM	632.3	636.8	368.6	375.8	30.4	36.1	76.9	87.5	0.01	4.21
8:00 PM	636.5	636.8	347.2	348.3	28.2	32.3	70.3	83.3	0.08	3.11
10:00 PM	636.5	636.8	346.4	347.5	24.6	26.6	65.8	74.2	0.11	3.14

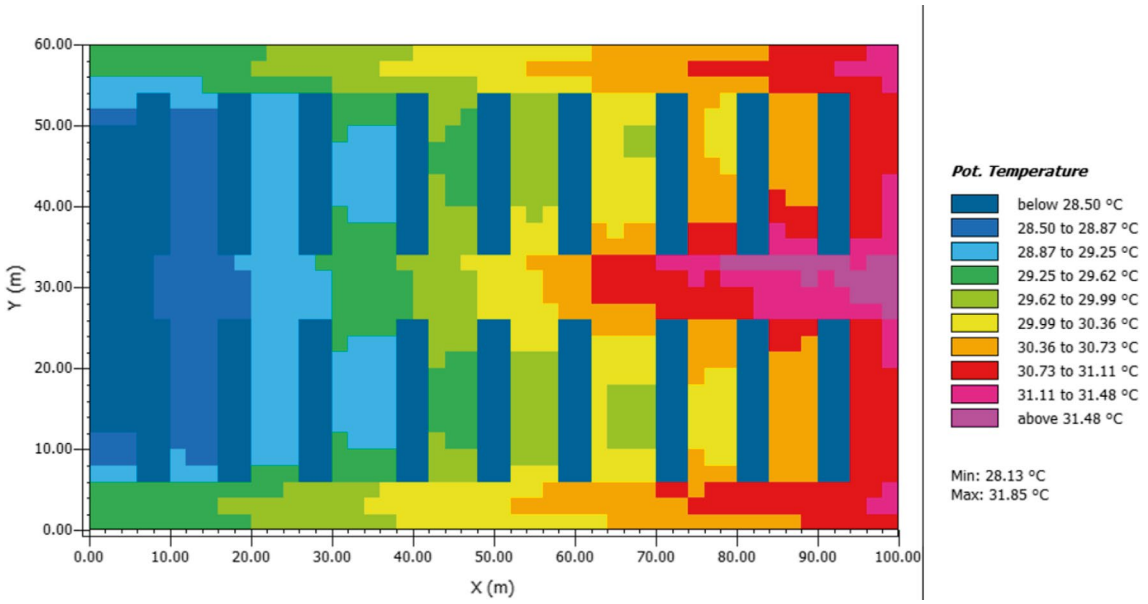


Fig. 9 Temperature simulation for a model (No. 2). *Source* Researcher

4 Results

Below is an explanation of the most important search results.

4.1 Thermal Comparison Results for the Three Models

The difference in the urban design of the three models significantly affected the efficiency of the environmental and thermal performance of the models. The thermal and environmental simulation results for the models differed, as follows.

Temperature

Both models (2, 3) recorded a noticeable decrease in temperature compared to model (1), with a difference of up to (7 degrees Celsius to the highest degrees) at 2 pm, which reduced the great disparity between the minimum and maximum temperatures (day and night) for two models (2, 3), which is noticeably clear in model (1), as shown in Fig. 16, the noticeable decrease in the temperature of the two models (2, 3) may be due to the continuous stacking of buildings than in Model 1, which reduced the walls exposed to sunlight.

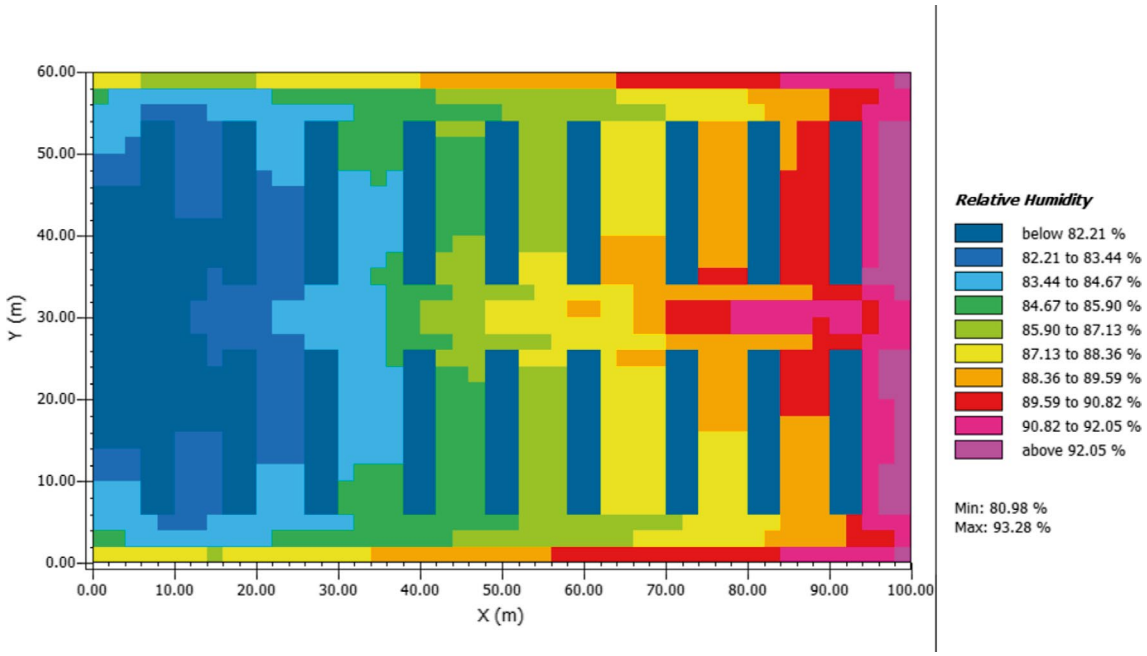


Fig. 10 Relative humidity simulation for a model (No. 2). *Source* Researcher

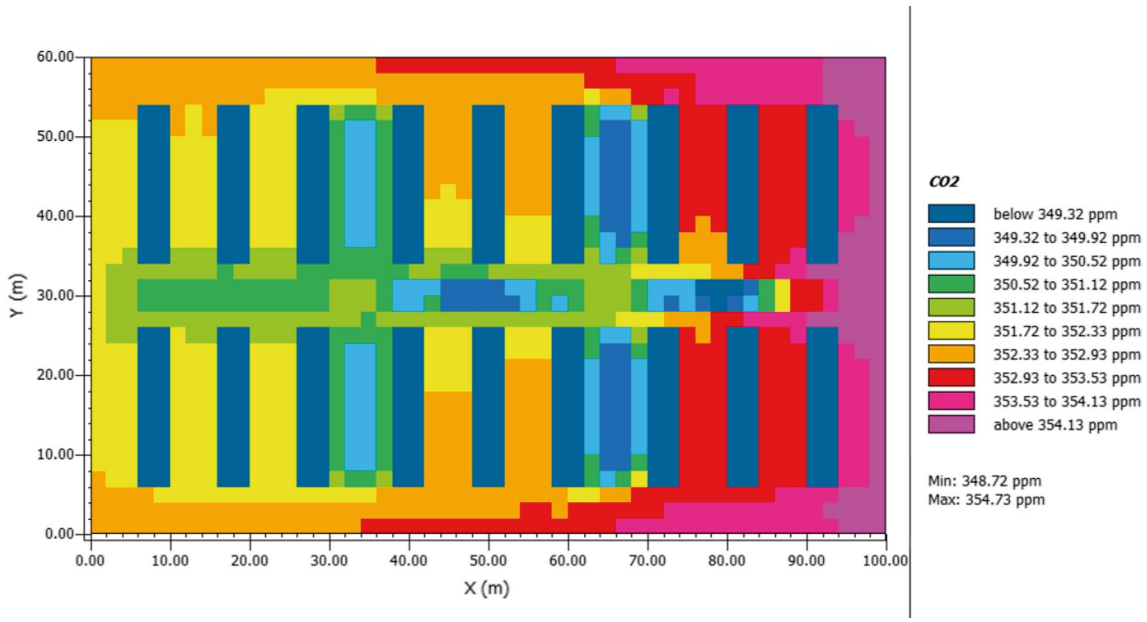


Fig. 11 Carbon dioxide simulation of emissions for a model (No. 2). *Source* Researcher

Relative Humidity

The relative humidity (min) decreased significantly for two models (2, 3) than it was in model (1), but they converged to each other in the three models of the max, as shown in Fig. 17.

Wind Speed

The results of the comparison of the maximum and minimum wind speeds for the three models the comparison have become clear that the wind speed decreased in model (2) than it was in both models (1,3) as shown in Fig. 18, this may be due to the stacking of buildings in Model 2 in a stripe pattern, impeding the flow and movement of the wind, which led to a lower speed than other models.

Dioxide Emissions

The rates of carbon dioxide emissions for each of the (2, 3) models were lower than in the (1) model, by a difference of (12 degrees max).

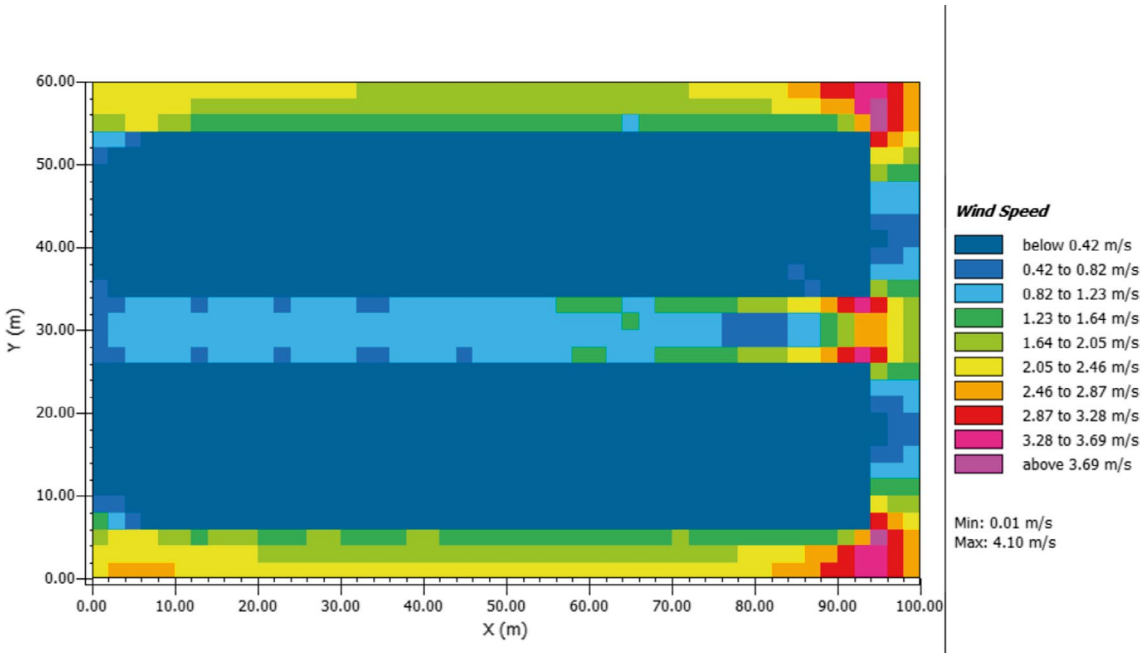


Fig. 12 Wind speed simulation for a model (No. 2). *Source* Researcher

Table 2 Thermal simulation results for Model 2

Model 2	co2 (mg\m3)		co2		Temp		Humidity		wind speed m\s	
Time	Min	max	min	max	min	Max	Min	Max	Min	Max
12:00 AM	636.5	636.8	346	347.1	18.2	18.9	67	75.5	0.09	3.17
2:00 AM	636.5	636.8	345.6	346.8	17.6	18	67.8	78.1	0.07	3.19
4:00 AM	636.5	636.8	345.3	346.5	16.7	17.8	68.5	78.9	0.05	3.21
6:00 AM	636.5	636.8	345.5	346.4	16.8	20.8	76.7	87.2	0.04	3.22
8:00 AM	626.1	636.8	347.9	354.2	22	24.8	88.2	93.5	0.02	3.04
10:00 AM	623.8	636.8	355.5	363.4	26.3	29.4	84.2	92.3	0.01	3.26
12:00 PM	623.2	636.8	358.4	367.6	28.5	23.2	79.5	89.8	0.01	3.46
2:00 PM	624.9	636.8	360.1	367.7	28.7	32.5	80.3	87.3	0.02	3.63
4:00 PM	628	636.8	360.8	366.1	26.4	31.2	81.1	84.2	0.03	3.79
6:00 PM	635.2	636.8	360.2	362.8	24.6	28.7	78.8	82.8	0.03	3,93
8:00 PM	636.5	636.8	347.2	348.3	21.2	23.3	72.1	78.3	0.08	3.11
10:00 PM	636.5	636.5	346.4	347.5	19.6	20.6	74.8	82.2	0.11	3.14

By Analyzing and Comparing the Thermal Simulation Results of the Three Models, the Results Were Concluded as: (Model 3) scored the best thermal and environmental performance among the three models (from achieving thermal comfort rates and carbon dioxide emissions), As it is the only model that has improved the efficiency of thermal performance between buildings, not only on the scale of open spaces but it is also followed by Model No. (2), and at the end comes Model No. (1) with a big difference between the two models (2,3).

5 Conclusion

The urban design and urban planning have a clear effect on the efficiency of thermal and environmental performance within urban communities, bearing far-reaching implications for the following aspects:

- **Thermal Comfort of Spaces**: The meticulous design and planning of urban spaces directly impact the thermal comfort experienced by inhabitants. Thoughtful

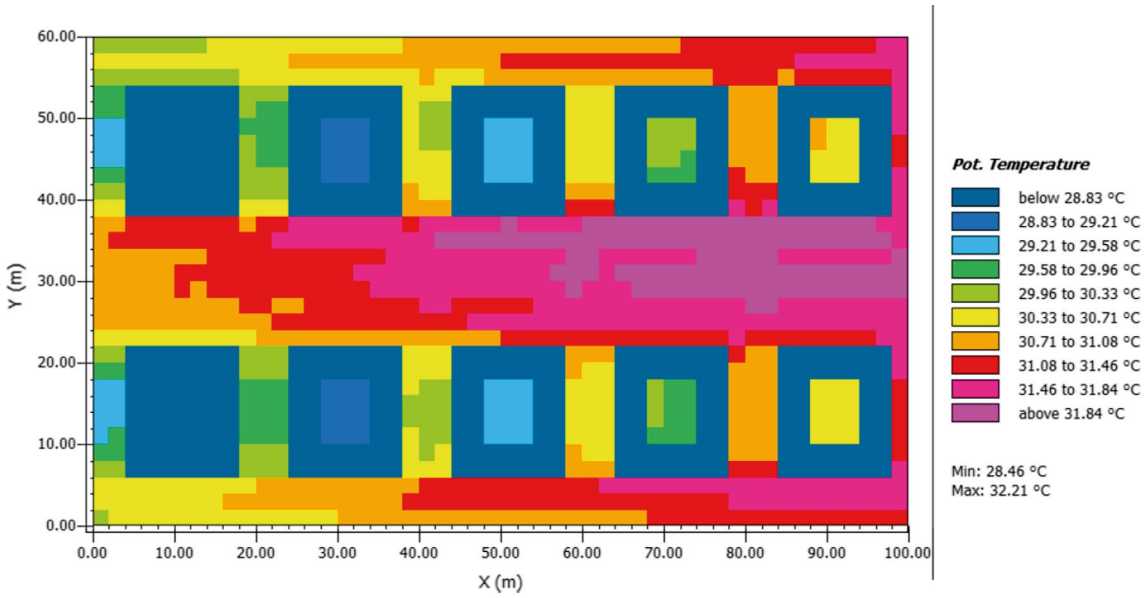


Fig. 13 Simulation of temperatures for a model (No. 3). *Source* Researcher

Table 3 Thermal simulation results for model 3

model 3	co2 (mg\m3)		co2		Temp		Humidity		wind speed m\s	
Time	min	Max	Min	Max	Min	Max	Min	Max	Min	Max
12:00 AM	636.5	636.8	346	347.1	17.2	18.3	66.5	75.5	0.09	3.2
2:00 AM	636.5	636.8	345.6	346.8	16.1	18	67.8	76.3	0.07	3.19
4:00 AM	636.5	636.8	345.3	346.6	15.7	17.8	71.5	77	0.05	3.21
6:00 AM	636.5	636.8	345.5	346.6	19.7	20.8	74.7	82.7	0.04	3.22
8:00 AM	626.5	636.8	348.7	354.7	22.3	23.2	87.2	90.6	0.01	3.32
10:00 AM	626	636.8	354.9	362.2	26	28.5	82	93	0.01	3.54
12:00 PM	626	636.8	357.6	367.3	28.5	32.2	83.3	87.3	0.01	3.75
2:00 PM	626.1	636.8	358.4	367.6	28.2	31.4	82.6	85.6	0.01	3.93
4:00 PM	628.6	636.8	356.3	364.4	27.9	29.9	81.3	83.4	0.01	4.1
6:00 PM	635.1	636.8	358.6	360.4	26.4	27.1	79.2	82.5	0.01	4.25
8:00 PM	636.5	636.8	347.2	348.3	20.2	23.3	76.3	80.3	0.08	3.11
10:00 PM	636.5	636.8	346.4	347.5	18.1	18.6	71.8	78.2	0.11	3.14

arrangement of buildings, green spaces, and public areas can influence factors like shade provision, wind circulation, and temperature moderation. Effective urban design contributes to creating microclimates that enhance outdoor and indoor thermal comfort, ultimately fostering a healthier and more livable environment for residents.

- **Physical Health of the Community**: The design and layout of urban communities significantly influence the physical well-being of their inhabitants. Consideration of factors such as walkability, accessibility to green spaces, and reduction of heat islands can promote physical activity, reduce air pollution exposure, and mitigate health risks associated with extreme temperatures. A well-designed urban environment can contribute to lower rates of respiratory issues and other health problems.
- **Carbon Emissions and Energy Use**: Urban design and planning choices directly impact the energy demands required to maintain thermal comfort within buildings. Poorly designed spaces may necessitate higher energy consumption for heating and cooling, thereby contributing

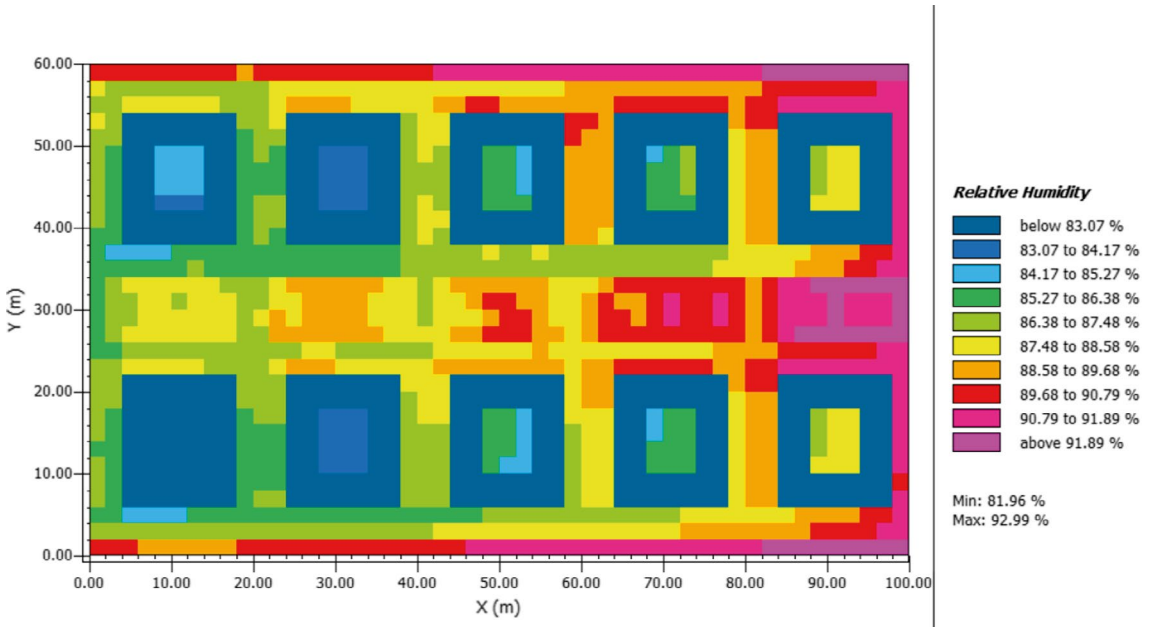


Fig. 14 Relative humidity simulation for a model (No. 3). *Source* Researcher

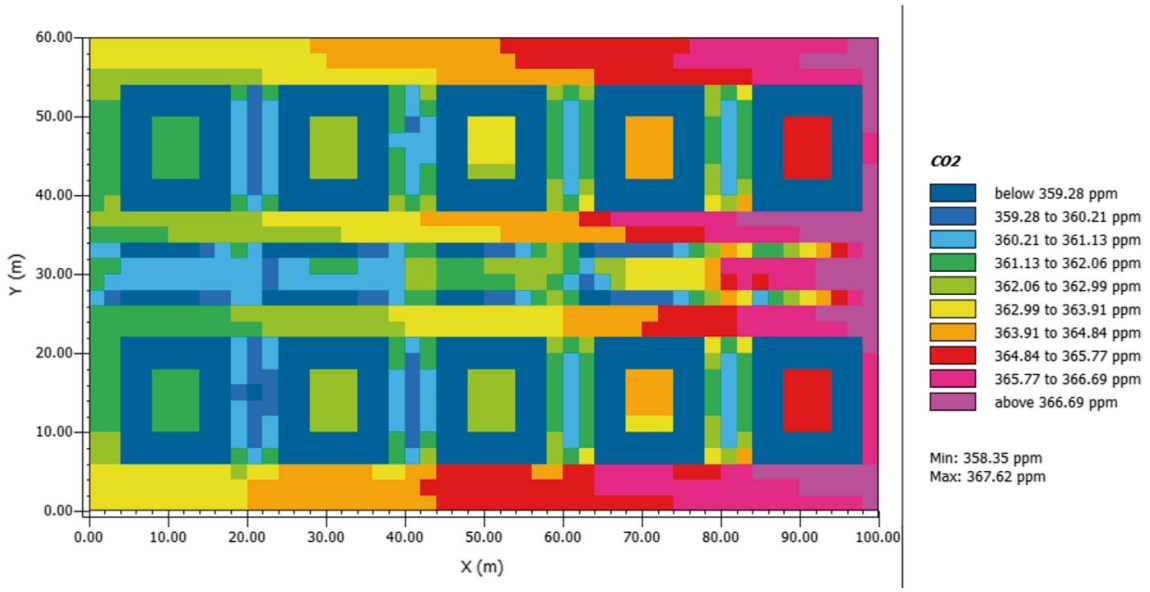


Fig. 15 Carbon dioxide simulation of emissions for a model (No. 3). *Source* Researcher

to increased carbon emissions. Conversely, well-designed urban communities can incorporate energy-efficient buildings, passive cooling strategies, resulting in reduced carbon emissions and a diminished ecological footprint.

- **Environmental Impact and Global Warming**: Inadequate urban design can lead to urban heat islands and heightened energy consumption for temperature regulation, exacerbating local environmental challenges and contributing to global warming. Urban planning decisions influence the amount of green cover, surface reflectivity, and heat-absorbing materials used in construction, all of which contribute to the overall urban heat island effect and its broader environmental repercussions.

Finally, urban design can be perceived as a fundamental constituent within the broader framework of urban planning.

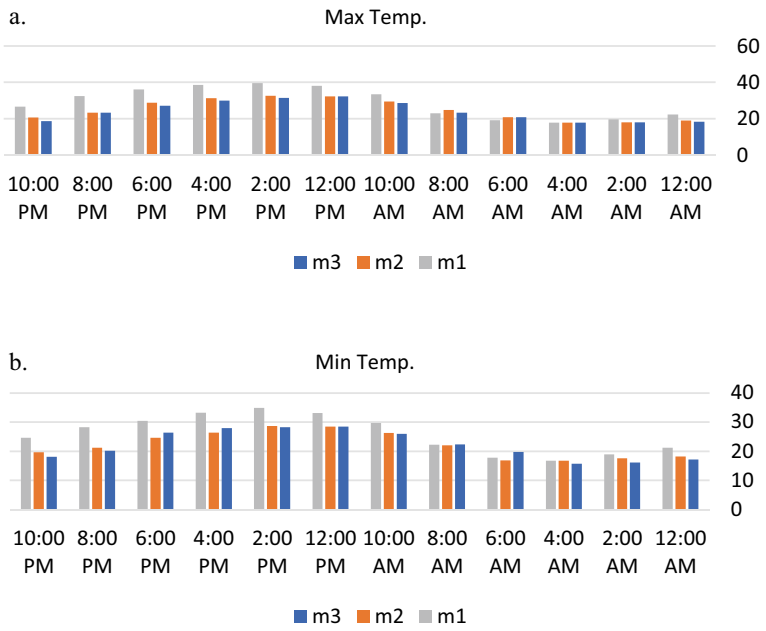


Fig. 16 Comparison of maximum and minimum temperatures for the three models

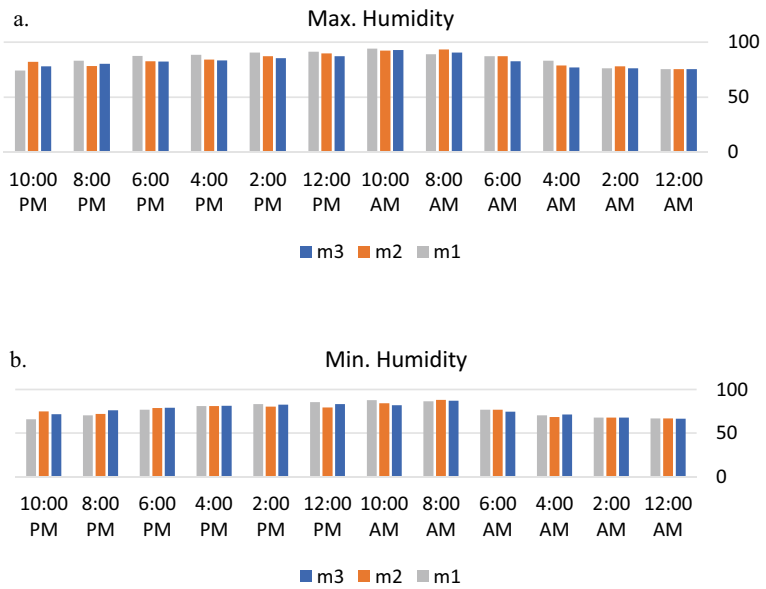


Fig. 17 Comparison of the maximum and minimum relative humidity for the three models

The adept execution of urban design, characterized by optimal thermal performance efficiency and reduced carbon emissions, holds the potential to usher in urban communities marked by enhanced sustainability, environmental preservation, and mitigation of negative global climate impacts.

Consequently, a compelling imperative emerges to cultivate an environment of encouragement and collaboration among urban planners, architects, real estate developers, and specialists engaged in urban legislation development. This collective effort aims to rigorously examine the profound impact of urban design and planning on the intricate interplay of environmental and thermal performance within urban communities. A cornerstone of this endeavor involves the formulation and exploration of multiple viable alternatives, each engineered to yield optimal outcomes in terms of environmental and thermal efficiency. The utilization of sophisticated simulation programs constitutes a pivotal tool in this quest to discern the most effective strategies.

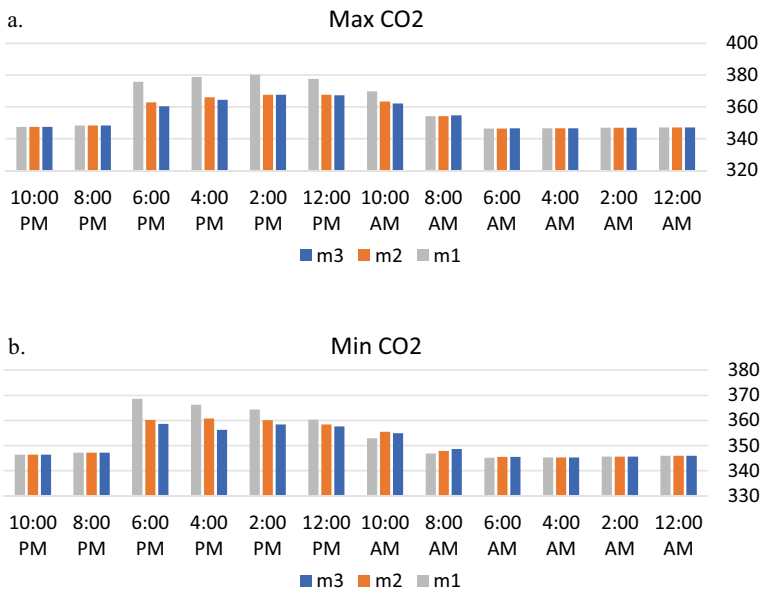


Fig. 18 Comparison of the maximum and minimum wind speeds for the three models

References

Beatly T, Deyle E, French P (1998) Hazard assessment: The factual basis for planning and mitigation. In: Burby R (ed) Cooperating with nature: confronting natural hazards with land use planning for sustainable communities. Joseph Henry/National Academy Press, Washington, DC, pp 119–166

Blakely G, Edward J (2004) Suburbs as sustainable communities: a paradigm for the future. Aust Plan 40(4)

Envi- met software. Microclimate and simulation. Website. https://www.envi-met.com/microclimate-simulation-software/ at 20-3-2022

Gastec. (2022) Gastec corporation - For all types of gas and vapour, Website. https://www.gastec.co.jp/en/technology/knowledge/concentration/ at 18-3-2022

Jean J (2016) I'm living in a carbon bubble. Literally. Medium site. https://medium.com/@joeljean/im-living-in-a-carbon-bubble-literally-b7c391e8ab6 at 17-3-2022

Karen C, Galina C, Angel H, Anu R (2021) From low- to net-zero carbon cities. the next global agenda. Annu Rev Environ Resour 46

Robert D (2011) Landsc Urban Plan 100(4):372–374

Rosenzweig C, Solecki W, Lankao P, Mehrotra S, Dhakal S, Ibrahim S (2018) Climate change and cities: second assessment report of the urban climate change research network. Cambridge University Press, ISBN 978-1-316-94456-1

Seto K, Davis S, Mitchell R, Ronald B, Eleanor C, Gregory U, Diana Ü (2016) Carbon lock-in: types, causes, and policy implications. Annu Rev Environ Resour

United nation climate chang. (2020) Urban climate action is crucial to bend the emissions curve report, https://unfccc.int/news/urban-climate-action-is-crucial-to-bend-the-emissions-curve

United Nations (2004) World urbanization prospects: the 2003 revision. New York

Urban design group (2022) What is Urban design? https://www.udg.org.uk/about/what-is-urban-design at 20-3-2022

Wheeler M (2013) Planning for sustainability: creating livable, equitable and ecological communities. Routledge. ISBN 978-1-136-48201-4

Influence of Embedded Crack on the Mechanical Failure of Ti–6Al–4V Locking Compression Plates Using Finite Element Analysis

Surinder Pal, Waqas Saleem, Xavier Velay, and Kamleshwar Kumar

Abstract

The structural failure of medical implants commonly happens due to the intense and dynamic stress profile generated by daily activities. To ensure long-term service life, an implant should demonstrate high fracture resistance under varying forces. This study investigates the behavior of a stationary crack embedded on the circumference of a Ti–6Al–4V made locking compression plates (LCP), under tension and bending loadings. The criteria for crack simulations and fracture propagations were reviewed, and the fracture modeling was performed using the contour integral approach available in Abaqus CAE software. Parameters, namely crack orientation and loading types, were selected to study their individual effects on stress intensity factors and for comprehensive fracture simulation purposes. The importance of mesh refinement at the tip of the crack was discussed. The simulation results revealed increased crack severity under bending loads, followed by tensile loads. As a noteworthy observation, the LCP implants should be designed and developed for future applications by accounting for the effect of bending stresses.

Keywords

LCP · Crack · Stress · Abaqus · Titanium

S. Pal (✉) · X. Velay
Atlantic Technological University, Sligo, Ireland
e-mail: surinder92.pal@gmail.com

X. Velay
e-mail: xavier.velay@atu.ie

W. Saleem
Technological University, Dublin, Ireland
e-mail: waqas.saleem@tudublin.ie

K. Kumar
National Institute of Technology, Jalandhar, India

1 Introduction

The modern part of the Locking Compression Plate (LCP) is the perfect mixture of two well-known anchorage ideas into one implantation. Earlier to the progress of LCP, limited plates preceded the concept of the LCP by; Frigg (2001, 2003). However, Ahmad et al. (2007) revealed that the widely used LCP can adapt flexibility and mechanical stiffness by configuring the space between plate and bone. Although, Miramini et al. (2015); Miramini et al. (2016a, b) showed that a flexible structure of the Locking Compression Plate could possibly promote secondary bone healing and enhance callus growth. Miramini et al. (2016a, b) stated that significant mechanical loading is a major factor impacting interfragmentary movement, fracture fixation stability, and fracture healing in the end. Numerous researchers such as Geris et al. (2003), González et al. (2010); and Bailón and Van Der Meulen (2003) have reported that the initial weight carrying develops the fracture healing process, and implants similar to the Locking Compression Plate permits primary movement of broken bone next to surgical treatment. Several trials have been presented to investigate the best suitable loading methods for the best crack healing outcomes. However, Wolf et al. (1998); Kenwright and Goodship (1989); Claes and Heigele (1999); and González-Torres et al. (2010) applied the range from 1 to 100 Hz frequencies and revealed that high-frequency cyclic loading stimulates chondrogenesis. Furthermore, Weiss et al. (2009) conducted a clinical survey of 6400 Swedish patients experiencing femoral shaft fractures, and the 1024 patients were healed by indirect reduction with screws and plates. Similarly, Kesemenli et al. (2002) found that around 90% of the crack cases were attained within an average period of 4.20 calendar months. Nevertheless, according to reported articles Kanchanomai et al. (2008), Nassiri et al. (2012), and Henderson et al. (2011), there have been many instances where the Locking Compression Plate was not completely productive.

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*,
Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_11

Reported studies by Chen et al. (2010), and Hoffmeier et al. (2011) included different parameters which further impact the implant stability and working length. Some works also cited the lifespan of LCP and fatigue life. Likewise, investigations studied by Chen et al. (2017); Chen (2018) found that the Locking Compression Plates have been repeatedly centered on compatibility among the geometry of the bone and the given plate. Still, they do not consider the mechanical attributes that influence the selection of LCP and design.

In contrast, very few computational models and experiments have examined the impact of bolt structure and middle transverse cracks (Sheng et al. 2019; Nourisa et al. 2015) but mainly performed the basic models of human load (Xie and Luo 2016; Leonidou et al. 2015). Various numerical studies (Zghal and Dammak 2020; Zghal et al. 2022) have reported the impacts on their mechanical response and geometric factors. Nevertheless, very few studies (Zghal et al. 2021; Trabelsi et al. 2020; Thai and Kim 2015) have combined the genetic algorithm and finite element method (FEM) to examine the best possible amount of screw and their positioning. But many reported studies were found on the extensive use of FEM to examine functionally graded structures. Additionally, Suaimi et al. (2022) investigated the von Mises stresses under different amounts of screws for locked compression plates such as 8, 6, 4 screws. It was found that under treating cracks, screws numbers 8 and 4 proved the optimal conditions for delivering high stability. These types of failures and cracks are mainly influenced by Loads that are dynamic or static; sometimes, it happen due to cyclic stress in the course of running or walking (Kanchanomai et al. 2008; Hosseini et al. 2012). However, the stresses are minimal in a few cases, but the fatigue caused due to stress amplitude is further responsible for the failure of the LCP (Moshfeghi et al. 2015; Sharma et al. 2006)). It is essential to observe that the orthopedic physician's practice is necessary for preventing inappropriate implantation. Moreover, Sharma et al. (2006) demonstrated implant failure in orthopedic surgery. It was found that implantation failure occurred due to screw loss in 40 patients, which led to the cracking initiation at an early stage and fracture, along with stress concentration (Azevedo and Hippert 2002; Tatić et al. 2018). Bone fractures are characterized as mechanical damage, which contains damage to surrounding soft tissue and fractures of the bone. These damages occur due to loads and external forces and with intensities exceeding the bone's elasticity level, thus affecting its injury. Therefore, biomedical implants are used to heal these broken bones. The locking compression plate (LCP) represents an effective bridging device for treating long bone fractures by offering supplementary structural support throughout the healing procedure. It is located on top of the fracture and attached to bone fragments with bolts (Sharma et al. 2006). The biomaterials for orthopedic locking compression plates (LCP) are significantly stronger than bone. Therefore, the titanium alloy Ti–6Al–4V has been widely used for orthopedic plates due to its excellent mechanical properties, fatigue strength, exceptional biocompatibility, and corrosion stability (Nemati et al. 2015; ASTM F1472-14; Hosseini et al. 2012; Azevedo and Hippert (2002); Brunner and Simpson 1980; Sedmak et al. 2019). However, there are several reported cases of the failure of LCP in service due to loading orientation. One of the most comprehensive findings reported by Sharma et al. (2006) covers 41 patients aged 17 to 65 years under various loading conditions. The finding of the experiments comprised the failures of LCP due to improper connection of plate and bone and the loosing of screws. Moreover, it revealed that all types of fractures are possible, fatigue, plastic, and brittle. Mathematical methods are extensively used for simulating these types of different behaviors of various implants under dynamic and static loading, such as LCP and hip joints (Milovanovic et al. 2017; Mijatović et al. 2019; Tatić et al. 2014; Tatić et al. 2018). Also, Sommer et al. (2004) observed that in the past, the locking plate method was extensively used in all metaphysis areas since it delivers good angular stability, which further dispersed the load more consistently and advanced the stability. However, the primary concern is that traditional locking plates repeatedly lead to metal prominence and failure due to different loading orientations, which results in frequent pain and implant removal. This malfunction avoids further healing and necessitates a patient to undergo a new surgery, which is undesirable in modern treatments. Nevertheless, the reported studies (Ahmad et al., (2007); Stoffel et al. (2003); Kubiak et al., (2006); Nassiri et al., (2012); Marciniak et al., (2008); Basiaga et al., (2010)) still lack consideration for the relationship between stress, displacement, loading orientation, and fixation. The biomechanical difficulties may occur following LCP operations, particularly related to the mechanical stresses and strains. However, Duda et al. (2001); Krishna et al. (2008) found that statistical methods, like the finite element method (FEM), are highly acceptable in orthopedics research and are to be applied to analyze the stresses and strains in the LCP plates and screws under different loading conditions. In the current study, the model of Ti–6Al–4V LCP was designed with the help of ABAQUS, and their parameters were optimized to enhance the service life and fracture resistance of Ti–6Al–4V implants under different loading. Additionally, the plate for multiple displacements was analyzed using numerical analysis (Abaqus) to determine the state of deformations and stresses. The obtained results are essential for selecting the mechanical properties and structure of biomaterials for the bone plates.

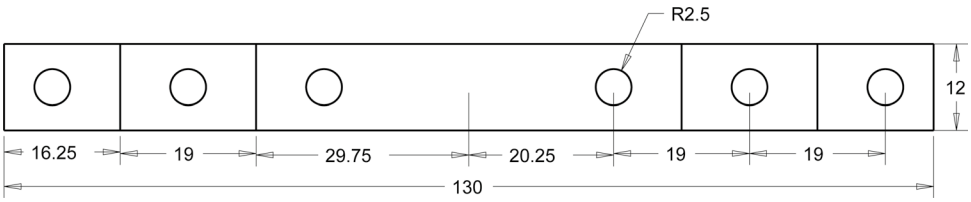


Fig. 1 Sketch of LCP in 2D space. (Dimensions are in mm, *Source* Creo)

Table 1 Properties of Ti–6Al–4V

Elastic modulus, GPa	Poisson's ratio	Density, g/m^2
114	0.32	4.4

2 Materials and Methods

2.1 LCP Modeling and Material

The present study employed a simplified model of 2D LCP, prepared using Creo Parametric software. The graphical model of the LCP shown in Fig. 1 comprises 6 screw holes on the plate. Owning to the biocompatibility and mechanical strength of Ti–6Al–4V alloy, its elastic properties were assigned to the model and summarized in Table 1.

2.2 Boundary and Loading Conditions

A concentrated force of 100 N was applied along the X and Y directions to evaluate the effect of tensile and bending stresses. Instead of individual loadings at different screw holes marked as A to F (Fig. 2), a reference point named RP1 was generated and loaded on the concentrated force. This reference point indicates coupling constraints between screw holes E and F, located in the middle of these screw holes. In contrast, screw holes A and B were chosen for fixed constraint, which eliminated all three degrees of freedom in the X and Y direction. The remaining screw holes named C and D were kept free from boundary and loading conditions, exposing them to induced stresses.

To simply the simulation, the following assumptions were considered:

- The assigned material is homogenous and isotropic in nature.
- The plastic behavior of the material was neglected.

2.3 Mesh Configuration

The LCP model was discretized using the global seeding method. As shown in Fig. 3, the final meshed model comprises two plane stress elements, a 3-node-based triangle (CPS3) and a 4-nodes-based quadrilateral (CPS4R).

2.4 Crack Modeling and Meshing

An edge crack was chosen for the present study and placed on the top of screw hole C. The crack was designed as 1 mm in crack length, while the crack orientation was varied into five different θ angles such as 30°, 60°, 90°, 120° and 150°. The detailed crack modeling was managed by the contour integral approach in ABAQUS CAE software. The crack front and seam were set, as shown in Fig. 4.

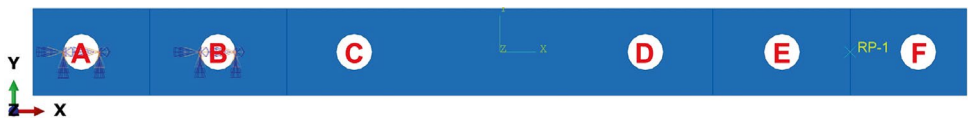


Fig. 2 Boundary conditions employed on LCP model (*Source* Abaqus 2022)

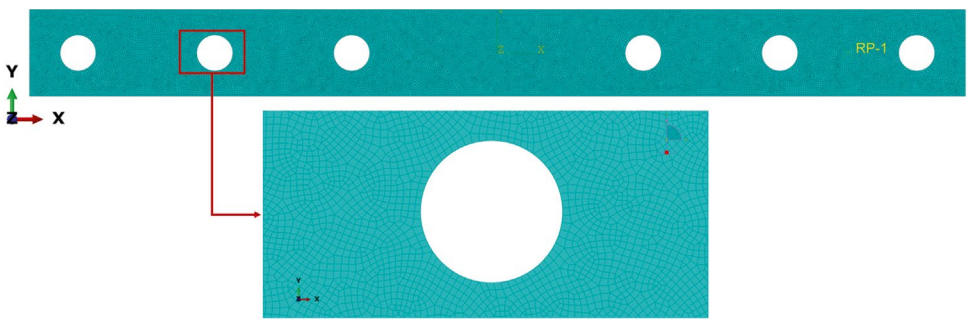


Fig. 3 Mesh configuration of the LCP model (*Source* Abaqus 2022)

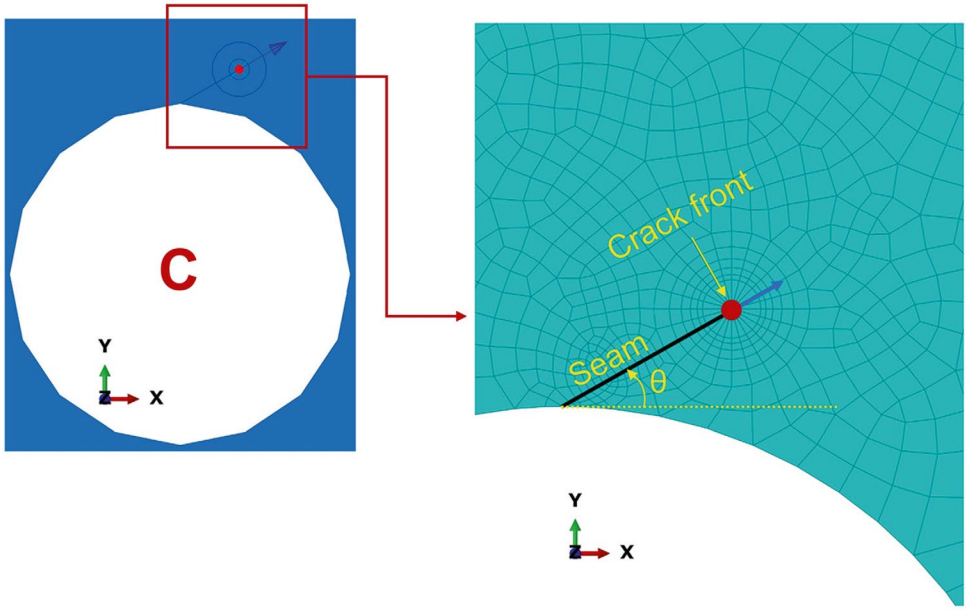


Fig. 4 Crack modeling with enlarged meshing in LCP model (*Source* Abaqus 2022)

The virtual crack extension was controlled by the q vector method. Later, the crack meshing was achieved by the local seeding method, which involved structured meshing of the first circle around the crack front using the CPS3 triangle element and CPS4R quadrilateral elements for the remaining five circles.

3 Results and Discussion

3.1 Pre-Crack Simulation

The pre-crack simulation was performed first to assess the effect of tensile and bending loadings on the LCP model without considering cracks. The von Mises stress distribution shown in Fig. 5 confirmed the formation of two high-stress regions on the LCP model under tension loading. Based on the nomenclature assigned in Fig. 2 earlier, it becomes evident that the potentially critical stressed regions on the respective circumferences of screw holes C and D, perpendicular to the direction of tension loading, are of particular significance. The distribution of similar and symmetrical stress across the LCP model directly resulted from the balanced and opposing nature of the applied loads.

This uniform stress distribution underscores the LCP's ability to effectively transmit tensile forces throughout its structure. Zhang et al. (2023) Identified that the arrangement of screws near the bone's fracture zone also plays

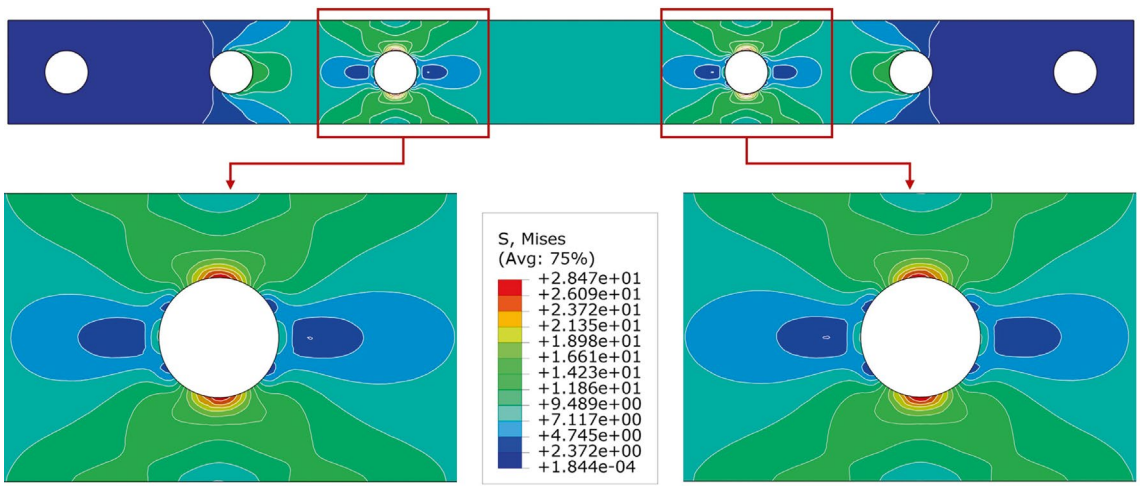


Fig. 5 Von-mises stress distribution on LCP model under tension (*Source* Abaqus 2022)

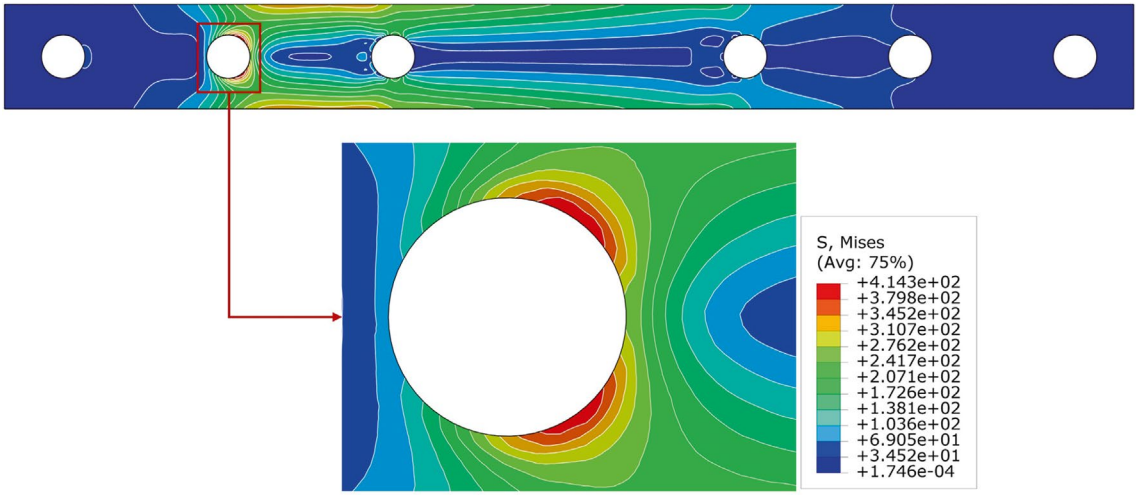


Fig. 6 Von-mises stress distribution on LCP model under bending (*Source* Abaqus 2022)

a role in influencing the vulnerability of LCP failure. Notably, the screw hole B, which benefits from fixed support conditions, exhibited von Mises stress in magnitudes lower than those observed in the corresponding screw hole E, which experienced a 100 N load in the X direction. In comparison, applying a 100 N bending force led to the forming of a single high-stressed region within the LCP model, as illustrated in Fig. 6. Remarkably, this critical region emerged specifically at screw hole B, which was subjected to fixed support boundary conditions. Unlike the regions of high stress observed under tension loading, the free screw holes C and D displayed more moderate von Mises stresses. This divergence in stress distribution patterns further highlights the LCP's responsiveness to different loading conditions. Moreover, Siva Kumar et al. (2023) uncovered that these stress ramifications prove instrumental in the iterative enhancement of device design, circumventing the necessity for measurement and thereby economizing on time and cost entailed by recurrent experimentation.

3.2 Crack Simulation

Earlier research has indicated by Kiyono et al. (2019); Chao et al. (2013), that the stiffness is minimally affected by the working length, as each plate is meticulously designed within an optimal range, and the choice of plates may vary depending on the specific fracture type. Based on the pre-crack simulation methodology, a carefully engineered crack was intentionally introduced into critical regions of the Load-bearing Cranial Plate (LCP) model to systematically examine its structural integrity in the presence of severe crack formations. This approach allowed for a comprehensive investigation into the behavior of the LCP model under different loading conditions, specifically focusing on crack propagation and stress distribution. As per the evaluation carried out by Xue et al., (2022), it was inferred that the highest equivalent stresses were concentrated at the fixation site of the model and the point of application for the load. The simulation specifically targeted a crack placement at screw hole C, strategically chosen to analyze the effects of tension and bending loads on the LCP's performance. In the context of tension loading, a meticulous analysis of stress intensity factors (SIF) was conducted, revealing significant insights into crack behavior. The pivotal observation emerged when the crack was positioned at screw hole C under tension loading, manifesting the highest mode-I SIF of 37.58 MPa$\sqrt{\text{mm}}$. This SIF magnitude was notably observed at a 90° orientation relative to the loading direction, signifying a perpendicular alignment. Intriguingly, a discernible pattern emerged where SIF values were reduced at 120° and 150° orientations in contrast to their counterparts at 30° and 60° orientations, as visually represented in Fig. 7 with Origin 2022. It's important to underscore that these SIF values were meticulously assessed using five distinct contours and then averaged to ensure the robustness and reliability of the findings. A distinct shift in crack behavior was observed upon transitioning to bending load scenarios. Strikingly consistent with the earlier tension-based outcomes, the 90° orientation again exhibited paramount significance. In this context, the simulation under 100 N bending unveiled a substantial increase in SIF magnitude, with a prominent value of 470.7 MPa$\sqrt{\text{mm}}$ observed at the same 90° orientation. These findings were consistent with the prevailing trend, reinforcing the critical impact of the 90° orientation on crack propagation and structural response under both

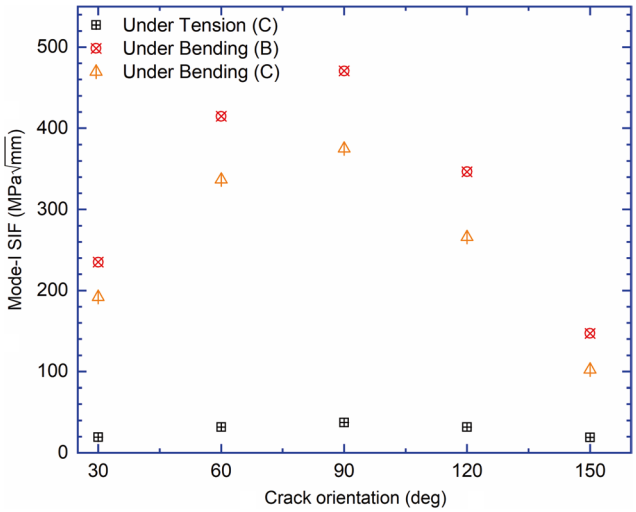


Fig. 7 Mode-I SIF of the crack in the LCP model strained under tension and bending loadings (*Source* Origin 2022). (B and C represent screw holes where the crack was placed)

tension and bending loads. Notably, to comprehensively evaluate the LCP's resilience and behavior, additional simulations were meticulously carried out, this time centering on screw hole C. This deliberate comparison aimed to discern the direct effects of tension and bending loads on the SIF magnitude. The resultant SIF of 375.27 MPa $\sqrt{\text{mm}}$, although smaller in comparison to screw hole B results, demonstrated a substantial magnitude when subjected to bending loads. This observation underscores the dominance of bending-induced stresses on crack propagation and the overall structural integrity of the LCP model when contrasted with tension-loading conditions. Furthermore, a comparable investigation was documented by Filip et al., focusing on the progression of cracks in orthopedic locking compression plates (LCP) made from the Ti–6Al–4V alloy. An analogous investigation conducted by Vučetić et al. (2020) examined plates under a four-point bending load, highlighting the pivotal influence of the structural design of LCPs on their enduring longevity. Similarly, Gautier and Sommer (2003) emphasized the indispensability of conscientiously accounting for bending loads to enhance the reinforcement and protection of LCPs, especially in contexts characterized by complex and routine physical activities.

Overall, the meticulous simulation-based investigation presented in this discussion sheds light on the intricate interplay between crack placement, loading conditions, and structural integrity of the Load-bearing Cranial Plate (LCP) model. The findings underscore the paramount significance of the 90° orientation in dictating crack behavior and stress distribution, with bending loads emerging as a dominant factor influencing crack propagation and structural response. These insights deepen our understanding of the LCP's behavior and emphasize the necessity of accounting for bending loads in the design and reinforcement of the LCP, especially in real-world scenarios involving complex and routine physical activities.

4 Conclusion

The present study investigated the behavior of a stationary crack embedded on the circumference of a Ti–6Al–4V made locking compression plates (LCP) under tension and bending loadings. The fracture modeling was performed using the contour integral approach with detailed simulations of pre-crack and post-crack of the LCP model under varying loading types and crack orientations, and the following conclusions are drawn:

- The free screw holes under tension and fixed screw holes close to bending loadings are critical to the structural integrity of the LCP model.
- The crack oriented at 90° under tension and bending leads to increased severity and stress concentration.
- The bending load poses a more significant threat to the structural integrity of the LCP model over the tension load of the same magnitude for the same crack profile.

References

Ahmad M, Nanda R, Bajwa AS, Candal-Couto J, Green S, Hui AC (2007) Biomechanical testing of the locking compression plate: when does the distance between bone and implant significantly reduce construct stability? Injury 38(3):358–364

ASTM F1472-14, Standard Specification for Wrought Titanium-6Aluminum-4Vanadium Alloy for Surgical Implant Applications (UNS R56400)

Azevedo CRF, Hippert EDUARDO Jr (2002) Failure analysis of surgical implants in Brazil. Eng Fail Anal 9(6):621–633

Bailón-Plaza A, Van Der Meulen MC (2003) Beneficial effects of moderate, early loading and adverse effects of delayed or excessive loading on bone healing. J Biomech 36(8):1069–1077

Basiaga M, Paszenda Z, Szewczenko J (2010) Biomechanical behaviour of surgical drills in simulated conditions of drilling in a bone. In: Information technologies in biomedicine: volume 2. Springer Berlin Heidelberg, pp 473–481

Brunner H, Simpson JP (1980) Fatigue fracture of bone plates. Injury 11(3):203–207

Chao P, Conrad BP, Lewis DD, Horodyski M, Pozzi A (2013) Effect of plate working length on plate stiffness and cyclic fatigue life in a cadaveric femoral fracture gap model stabilized with a 12-hole 2.4 mm locking compression plate. BMC Vet Res 9:1–7

Chen G, Schmutz B, Wullschleger M, Pearcy MJ, Schuetz MA (2010) Computational investigations of mechanical failures of internal plate fixation. Proc Inst Mech Eng, Part H: J Eng Med 224(1):119–126

Chen X (2018) Parametric design of patient-specific fixation plates for distal femur fractures. Proc Inst Mech Eng, Part H: J Eng Med 232(9):901–911

Chen X, He K, Chen Z (2017) A novel computer-aided approach for parametric investigation of custom design of fracture fixation plates. Comput Math Methods Med

Claes LE, Heigele CA (1999) Magnitudes of local stress and strain along bony surfaces predict the course and type of fracture healing. J Biomech 32(3):255–266

Duda GN, Mandruzzato F, Heller M, Goldhahn J, Moser R, Hehli M, …, Haas NP (2001) Mechanical boundary conditions of fracture healing: borderline indications in the treatment of unreamed tibial nailing. J Biomech 34(5):639–650

Frigg R (2001) Locking compression plate (LCP). An osteosynthesis plate based on the dynamic compression plate and the point contact fixator (PC-Fix). Injury 32:63–66

Frigg R (2003) Development of the locking compression plate. Injury 34:6–10

Gautier E, Sommer C (2003) Guidelines for the clinical application of the LCP. Injury 34:63–76

Geris L, Van Oosterwyck H, Vander Sloten J, Duyck J, Naert I (2003) Assessment of mechanobiological models for the numerical simulation of tissue differentiation around immediately loaded implants. Comput Methods Biomech Biomed Engin 6(5–6):277–288

González-Torres LA, Gómez-Benito MJ, Doblaré M, García-Aznar JM (2010) Influence of the frequency of the external mechanical stimulus on bone healing: a computational study. Med Eng Phys 32(4):363–371

Henderson CE, Lujan TJ, Kuhl LL, Bottlang M, Fitzpatrick DC, Marsh JL (2011) Healing complications are common after locked plating for distal femur fractures. Clin Orthop Relat Res 469(6):1757–1765

Hoffmeier KL, Hofmann GO, Mückley T (2011) Choosing a proper working length can improve the lifespan of locked plates: a biomechanical study. Clin Biomech 26(4):405–409

Hosseini S, Hudak R, Penhaker M, Majernik J (2012) Fatigue of ti-6al-4v. Biomedical engineering-technical applications in medicine, 75–92.

Kanchanomai C, Phiphobmongkol V, Muanjan P (2008) Fatigue failure of an orthopedic implant–a locking compression plate. Eng Fail Anal 15(5):521–530

Kenwright JOHN, Goodship AE (1989) Controlled mechanical stimulation in the treatment of tibial fractures. Clin Orthop Relat Res 241:36–47

Kesemenli C, Subasi M, Necmioglu S, Kapukaya A (2002) Treatment of multifragmentary fractures of the femur by indirect reduction (biological) and plate fixation. Injury-Int J Care Injured. 33(8):691–699

Kiyono M, Noda, T, Nagano H, Maehara T Yamakawa Y, Mochizuki Y …, Ozaki T (2019) Clinical outcomes of treatment with locking compression plates for distal femoral fractures in a retrospective cohort. J Orthop Surg Res 14:1–9

Krishna KR, Sridhar I, Ghista DN (2008) Analysis of the helical plate for bone fracture fixation. Injury 39(12):1421–1436

Kubiak EN, Fulkerson E, Strauss E, Egol KA (2006) The evolution of locked plates. JBJS 88(suppl_4):189–200

Leonidou A, Moazen M, Lepetsos P, Graham SM, Macheras GA, Tsiridis E (2015) The biomechanical effect of bone quality and fracture topography on locking plate fixation in periprosthetic femoral fractures. Injury 46(2):213–217

Marciniak J, Szewczenko J, Walke W, Basiaga M, Kiel M, Mańka I (2008) Biomechanical analysis of lumbar spine stabilization by means of transpedicular stabilizer. In: Information technologies in biomedicine. Springer Berlin Heidelberg, pp 529–536

Mijatović T, Milovanović A, Sedmak A, Milović L, Čolić K (2019) Integrity assessment of reverse engineered Ti-6Al-4V ELI total hip replacement implant. Struct Integr Life 19(3):237–242

Milovanovic A, Sedmak A, Colic K, Tatic U, Dordevic B (2017) Numerical analysis of stress distribution in total hip replacement implant. Struct Integr Life-Integritet I Vek Konstrukcija 17(2):139–144

Miramini S, Zhang L, Richardson M, Mendis P, Ebeling PR (2016a) Influence of fracture geometry on bone healing under locking plate fixations: a comparison between oblique and transverse tibial fractures. Med Eng Phys 38(10):1100–1108

Miramini S, Zhang L, Richardson M, Mendis P, Oloyede A, Ebeling P (2016b) The relationship between interfragmentary movement and cell differentiation in early fracture healing under locking plate fixation. Australas Phys Eng Sci Med 39(1):123–133

Miramini S, Zhang L, Richardson M, Pirpiris M, Mendis P, Oloyede K, Edwards G (2015) Computational simulation of the early stage of bone healing under different configurations of locking compression plates. Comput Methods Biomech Biomed Engin 18(8):900–913

Moshfeghi T, Nemati NH, Dehlaghi V, Sadrnezhaad S (2015) Fracture analysis of the orthopedic plates used in joining of the lower-limb bones

Nassiri M, MacDonald B, O'Byrne JM (2012) Locking compression plate breakage and fracture non-union: a finite element study of three patient-specific cases. Eur J Orthop Surg Traumatol 22(4):275–281

Nemati NH, Moshfeghi T, Dehlaghi V, Sadrnezhaad SK (2015) Fracture Analysis of the Orthopedic Plates used in Joining of the Lower-Limb Bones. Cumhuriyet Üniversitesi Fen Edebiyat Fakültesi Fen Bilimleri Dergisi 36(3):2718–2724

Nourisa J, Baseri A, Sudak L, Rouhi G (2015) The effects of bone screw configurations on the interfragmentary movement in a long bone fixed by a limited contact locking compression plate. J Biomed Sci Eng 8(09):590

Sedmak A, Čolić K, Grbović A, Balać I, Burzić M (2019) Numerical analysis of fatigue crack growth of hip implant. Eng Fract Mech 216:106492

Sharma AK, Kumar A, Joshi GR, John JT (2006) Retrospective study of implant failure in orthopaedic surgery. Med J, Armed Forces India 62(1):70

Sheng W, Ji A, Fang R, He G, Chen C (2019) Finite element-and design of experiment-derived optimization of screw configurations and a locking plate for internal fixation system. Comput Math Methods Med

Siva Kumar KGV, Mohammed AT, Rajeev A, Muraleedharan CV (2023) Finite element study of bending of a custom 2-level anterior cervical plate in 4-point configuration. Trans Indian Natl Acad Eng 8(2):263–272

Sommer C, Babst R, Müller M, Hanson B (2004) Locking compression plate loosening and plate breakage: a report of four cases. J Orthop Trauma 18(8):571–577

Stoffel K, Dieter U, Stachowiak G, Gächter A, Kuster MS (2003) Biomechanical testing of the LCP–how can stability in locked internal fixators be controlled? Injury 34:B11–B19

Suaimi MKA, Rashid AMA, Nasution AK, Seng GH, Wui NB, Ramlee MH (2022) Stress distribution and stability evaluation of difference number of screws for treating tibia transverse fracture: analysis on patient-specific data. Malays J Med & Health Sci 18(3)

Tatić U, Čolić K, Sedmak A, Mišković Ž, Petrović A (2018) Evaluation of the locking compression plates stress-strain fields. Tehnički Vjesnik 25(1):112–117

Tatić U, Miletić V, Sedmak S, Mitrović N, Ezdenci A, Gubeljak L, Milošević M (2014) Influence of the cavity shape in restorative dentistry on the stress-strain distribution in dentine and enamel caused by polymerization. Integritet i Vek Konstrukcija 14(3):199–204

Thai HT, Kim SE (2015) A review of theories for the modeling and analysis of functionally graded plates and shells. Compos Struct 128:70–86

Trabelsi S, Zghal S, Dammak F (2020) Thermo-elastic buckling and post-buckling analysis of functionally graded thin plate and shell structures. J Braz Soc Mech Sci Eng 42(5):1–22

Vučetić F, Čolić K, Grbović A, Radaković Z, Sedmak S (2020) Extended FEM analysis of fatigue crack growth in Ti-6Al-4V orthopaedic plates. Procedia Struct Integr 28:555–560

Weiss RJ, Montgomery SM, Al Dabbagh Z, Jansson KÅ (2009) National data of 6409 Swedish inpatients with femoral shaft fractures: stable incidence between 1998 and 2004. Injury 40(3):304–308

Wolf S, Janousek A, Pfeil J, Veith W, Haas F, Duda G, Claes L (1998) The effects of external mechanical stimulation on the healing of diaphyseal osteotomies fixed by flexible external fixation. Clin Biomech 13(4–5):359–364

Xie T, Luo Y (2016) Screw layout optimization to solve fatigue fracture of femoral Ti alloy metal plate. In: MATEC web of conferences, vol 67. EDP Sciences, p 03019

Xue Y, Wang XF, Zhao FL, Han DC, Long AH, Wang J, ..., Wang XF (2022). Finite element analysis of new headless compression supporting screw for the treatment of unstable femoral neck fracture. Bio-Med Mater Eng 33(4):279–292

Zghal S, Dammak F (2020) Vibrational behavior of beams made of functionally graded materials by using a mixed formulation. Proc Inst Mech Eng C J Mech Eng Sci 234(18):3650–3666

Zghal S, Ataoui D, Dammak F (2022) Static bending analysis of beams made of functionally graded porous materials. Mech Based Des Struct Mach 50(3):1012–1029

Zghal S, Trabelsi S, Frikha A, Dammak F (2021) Thermal free vibration analysis of functionally graded plates and panels with an improved finite shell element. J Therm Stresses 44(3):315–341

Zhang NZ, Liu BL, Luan YC, Zhang M, Cheng CK (2023) Failure analysis of a locking compression plate with asymmetric holes and polyaxial screws. J Mech Behav Biomed Mater 138:105645

Analysis of the Impact of New Singular Ventilation Technologies on Enhancing Indoor Air Quality in Schools

Laura Quant, Olga Macias-Juez, Ander Romero-Amorrortu, Asier Urrutia-Sustatxa, Antxon Urrutia-Sustatxa, and Javier de-Iribas

Abstract

Concern about indoor air quality (IAQ) in schools has grown in recent years, particularly in the aftermath of the COVID-19 pandemic, which underscored its impact on children's health. Existing educational buildings employ simple strategies such as opening windows to ventilate classrooms. While this approach achieves the goal of air renewal, it compromises energy efficiency and user comfort. In this context, there is a need to address ventilation in schools from a new perspective, providing innovative technologies that allow quick and simple installation while guaranteeing high standards of air quality, energy efficiency, and user comfort. In response to this challenge, a new solution has been developed consisting of autonomous equipment installed inside each classroom and featuring independent intelligent control. The objective of the present study is to evaluate the applicability and the social impact that the widespread implementation of this alternative technology could have compared to conventional methods. The general characteristics and the specific peculiarities and needs of schools in the Basque Country (CAPV) in Spain were determined. The study first approached the topic theoretically through bibliographic references and statistical analysis, and subsequently, fieldwork to assess the reality of existing buildings. Additionally, an air quality monitoring campaign was carried out in pilot schools, conducted in two stages: first without ventilation and later with the new solution. The study evaluated the benefits in terms of improved air quality achieved, as well as the improvements in the implementation and operational processes. These results were extrapolated to Basque educational buildings, providing an estimation of the potential impact of this new ventilation approach. Highly positive results were yielded in terms of acceptance, feasibility, and ultimately, addressing the identified challenges.

Keywords

Indoor air quality · Schools · Ventilation · Energy efficiency · CO_2 · Demand-controlled ventilation

L. Quant (✉) · O. Macias-Juez · A. Romero-Amorrortu
TECNALIA, Basque Research and Technology Alliance (BRTA), Derio, Spain
e-mail: laura.quant@tecnalia.com

O. Macias-Juez
e-mail: olga.macias@tecnalia.com

A. Romero-Amorrortu
e-mail: ander.romero@tecnalia.com

A. Urrutia-Sustatxa · A. Urrutia-Sustatxa
AIDEKO, Derio, Spain
e-mail: asier@aideko.com

A. Urrutia-Sustatxa
e-mail: antxon@aideko.com

J. de-Iribas
Bikat Manufactures–Coproven Group, Munguia, Spain
e-mail: jdeiribas@bikat.es

1 Introduction

Indoor air quality (IAQ) in buildings has a major impact on people's health, with poor indoor air quality causing discomfort in the form of a range of symptoms and adverse effects that can vary from mild discomfort to the development of major illnesses. According to the European Environment Agency, acute lower respiratory tract infections attributable to indoor air pollution account for 4.6% of all deaths worldwide (European Environment Agency 2019).

In a school environment, poor air quality causes lack of concentration, loss of performance and can lead to stress, depression, and absenteeism (Toftum et al. 2015). A recent

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_12

review regarding air quality and health in schools worldwide (Sadrizadeh et al. 2022) claims that pupils' exposure to insufficient air quality in school buildings may induce serious consequences for their health, as their bodies require more air volume per unit of body weight than adults do. Children's respiratory, digestive, reproductive, immune, central nervous, and neural systems are still developing. In addition to the odours and vapours from art classes and laboratories, other typical sources of pollutants in schools include the chalk dust, fungus, germs, and viruses that children and adults bring into the school environment. Exposure to air pollution has been proven to elevate the risk of various health issues in children, including acute respiratory illnesses and asthma, ultimately leading to a rise in child mortality rates. The variety of symptoms and chronic diseases that have been linked to varied immune system reactions in children due to indoor air exposures have been identified as "sick building syndrome". Taking into account the previous statements, it is considered necessary to control indoor air quality in educational establishments. The present study focuses on the impact of a new solution for improving the Indoor Air Quality (IAQ) of educational centres of the autonomous community of the Basque Country (CAPV).

Several researchers have studied the concentration of indoor air pollutants in Spanish schools (Becerra et al. 2020; Mazaheri et al. 2016; Rivas et al. 2015), identifying indoor and outdoor air pollutants in Mediterranean schools that may affect children's health when there is insufficient ventilation. Studies regarding ventilation in the territory focus on natural ventilation strategies to achieve appropriate ventilation (Aguilar et al. 2022; Monge-Barrio et al. 2022; Rodríguez-Vidal et al. 2022). However, depending on the season and the climatic typology of some locations such as the Basque Country (northern Spain) opening windows to achieve a good IAQ may result in thermal discomfort and higher energy consumption. Then, mechanical ventilation is considered as a good option, because it can include both ventilation and heat exchange functionalities. Besides, Canha et al. (2016) work regarding the assessment of ventilation in nursery and elementary schools in France concluded that the ventilation rate was significantly higher in mechanically ventilated classrooms compared with naturally ventilated rooms. Toftum et al. (2015) claimed that pupils in schools with balanced mechanical ventilation had significantly higher achievement indicators than pupils in schools with natural ventilation, where airing took place mostly by manual window opening. And finally, regarding efficiency, Merema et al. (2018) pointed out that demand-controlled ventilation produces significant energy consumption and heat loss reductions in lecture rooms with a varying occupancy profile.

Mechanical ventilation is a relatively new concept, which in Spain was not mandated in the regulations for school buildings until 2007. As can be seen in Fig. 1 Number of cultural buildings (mainly educational facilities) built from 1900 to 2020 in Spain, more than 95% of the schools in the territory were built prior to the obligation of mechanical ventilation. Consequently, requirements concerning air renewal or CO_2 levels are not met. In these schools, ventilation is dealt with by opening windows, which have little control of pollutant levels and can lead to thermal or acoustic discomfort. This highlights a major underlying health problem that affects the most vulnerable population: children.

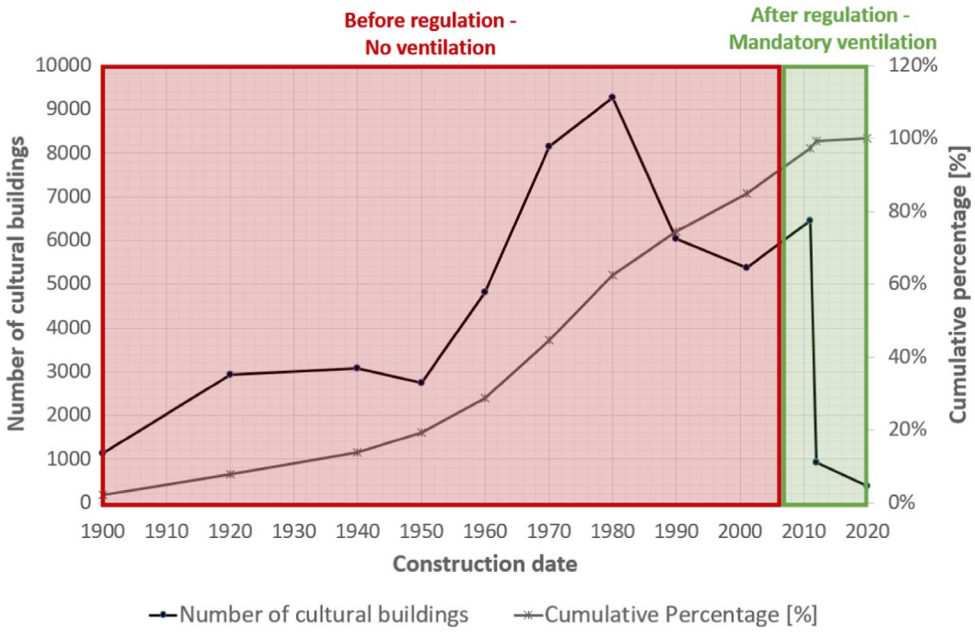


Fig. 1 Number of cultural buildings (mainly educational facilities) built from 1900 to 2020 in Spain based on data from ERESEE (Ministerio de Transportes Movilidad y Agenda Urbana. Gobierno de España (2020)

The case study focuses on the autonomous community of the Basque Country. In this region, following the same trend as in the rest of the country, most of the schools were built prior to the incorporation of mandatory mechanical ventilation into building regulations.

Concerning the availability of affordable technical solutions and the quality-to-simplicity ratio of the process, air quality levels in a space can be evaluated directly by measuring the existing pollutants (mainly CO_2 concentration), or indirectly by ensuring an air renewal rate. In recent decades, IAQ has garnered interest and has given rise to standards, guidelines, and recommendations. Some of these references are gathered in Table 1. In most of them, optimal air quality levels are defined by one of the following parameters: required/recommended fresh air flow rate, air change per hour (ACH), or CO_2 concentration.

From the analysis of these references, it can be concluded that, although the normative values required in Spain are around 1000 ppm of CO_2, it is advisable to reduce these values to around 700–800 ppm. However, to reach these levels, energy cost is high summarizes these recommendations considering the energy costs, health, and air quality parameters.

To solve the problem of poor ventilation in schools, it would be necessary to install mechanical ventilation systems. However, conventional systems present certain problems: Existing commercial solutions applied to constructed buildings may not be suitable as they are generalist solutions, mainly designed for buildings in the tertiary sector. They usually lack case-specific adaptation, require a lot of space, consist of decoupled heating and ventilation systems, have limited control options (continuous operation, uncertainty about air quality), and have low energy efficiency.

In response to these barriers and limitations, the ESKOAIRE project has developed an autonomous demand-controlled ventilation system that can significantly improve the current situation in Basque schools, thereby enhancing children's health and performance. ESKOAIRE devices are intended to be installed in existing educational buildings that are not equipped with mechanical ventilation systems. They are specially adapted to the characteristics and needs of school classrooms. The system approaches ventilation in schools from a new perspective, different from traditional solutions. The system is composed of compact ventilation units, without ductwork, and CO_2 probe-based variable flow control that works independently for each classroom. These devices produce minimum noise, and their dimensions are suitable for typical classrooms. They have a logical connection between the ventilation control systems and the heating terminal units that, when necessary, enable manual operation or control.

The ESKOAIRE solution has been set up in two pilot schools. An air quality monitoring campaign was carried out, conducted in two stages: first without ventilation and later with the new solution. The results show an important improvement in classroom air quality, global energy consumption, and user comfort. The present study aims to highlight the impact of this solution if applied to all schools in the CAPV. This study and its results are considered highly innovative, as no literature has been found that addresses the improvements introduced by the use of mechanical ventilation in Spanish schools or those belonging to the CAPV.

Table 1 Recommendations and standards for indoor ventilation in residential sector and classrooms

	Supply air	CO_2 Concentration [ppm]	References
RITE. Spanish regulation of thermal installations in buildings	12.5 l/s.pers	$C_{int} \leq 500 + C_{ext}$ $C_{int\_max} = 1000$	Gobierno de España (2021)
Recommendations for O&M of monitorization and ventilation of buildings for SARS-COV-2 prevention	12.5 l/s.pers	–	Ministerio de Sanidad & IDAE (2020)
5-step guide to checking ventilation rates in classrooms	–	$C_{int\_max} = 700$	Harvard Healthy Buildings program (2020)
Guide for ventilation in classrooms. IDAE-CSIC	5 ACH	–	CSIC-IDAEA & Ministerio de Ciencia e Innovación y Mesura (2020)
ANSI/ASHRAE Standard 62.1–2019. Ventilation for Acceptable Indoor Air Quality	5 ACH	–	ANSI/ASHRAE (2019)
DIN 1946. Ventilation and air conditioning. 2019 (Classrooms)	5–7 ACH	–	Heating, Ventilation, and Air Conditioning. Requirements Relating to Health. DIN (1946, n.d.)
Efficient air renewal guide in the residential sector	–	$C_{int\_max} = 1000–1200$	Fenercom (2014)
Guide to good ventilation practices. Criteria to prevent the transmission of COVID-19	–	$C_{int\_max} = 800$	Asepeyo (2021)

2 Methodology

The methodology has been developed according to the needs of the project. This section describes the methodology followed for developing the technical solution and the research conducted to evaluate its impact. The methodology is shown in Fig. 3.

Initially, the identification of the applicability archetypes where the solution could potentially be installed was performed. To achieve this, bibliographical and statistical studies regarding the number of schools and their characteristics were carried out mainly using data from the Basque Statistics Institute (Eustat), standards, regulations, and reference literature on school building design and best practices. In addition to the existing general characteristics of the school buildings and classrooms, the number of schools and their year of construction were determined.

Then, a field study consisting of visits to educational facilities was conducted to validate the data set obtained from literature. The previous statistical study allowed the selection of representative sample schools, from which 26 were selected to be surveyed. Considering that some of them offer different educational stages, the result was the inspection of 51 classrooms of various typologies: 17 preschool classrooms, 18 primary classrooms, and 16 secondary classrooms. These actions allowed for obtaining the characterization of applicability archetypes of classrooms. As a result, the most common classroom typologies were identified.

Additionally, the identification of the shortcomings of current ventilation systems, the definition of the main technical challenges faced by the project, and the study of the regulatory requirements affecting ventilation systems in classrooms were performed. This work was completed by an analysis of recommendations and good practices in the field of ventilation in classrooms in order to identify the references and best practices to be applied.

Following this, the case study was defined considering the results of the archetype characterization, the problems to be addressed, and the regulations to be complied with.

Next, the technical development was faced. First, to overcome the problem posed by the case study, the characteristics that the ESKOAIRE solution must fulfil were defined.

Then, in the context of the project ESKOAIRE, three prototypes were developed and built. The final prototype was tested experimentally at KUBIK research building at the Tecnalia facilities (Chica et al. 2011). The KUBIK installation allows CO_2 and temperature levels to be controlled to simulate classroom occupation. A control system was developed and implemented as well, and different ventilation strategies were studied.

Finally, the ESKOAIRE solution was installed in two pilot schools where air quality measurements were taken before and after the installation of the equipment. In

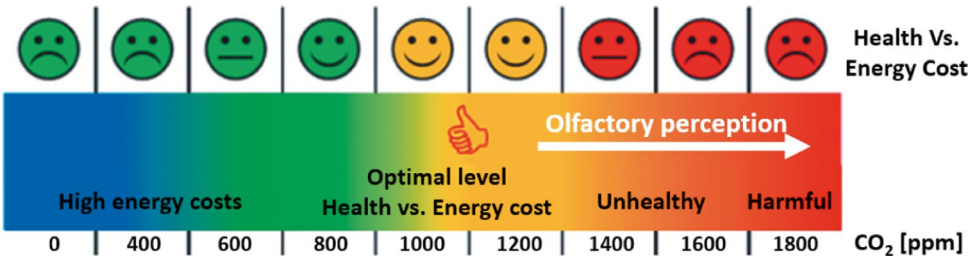


Fig. 2 Energy cost, health, and air quality balance (Fenercom 2014)

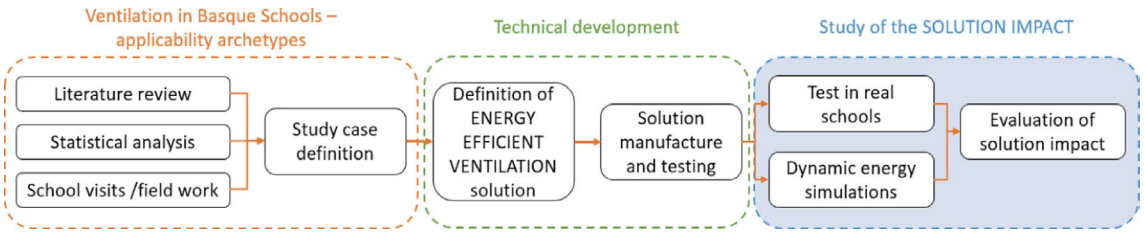


Fig. 3 Methodology scheme developed for ESKOAIRE project. *Source* The authors

addition, complementing field measures, dynamic thermal simulations were carried out to predict the impact of alternative ventilation scenarios. The study case solution impact was conducted by extrapolating the analysed data from the pilots and simulations to the whole CAPV.

3 Results

This section includes the main results of the ESKOAIRE project research described in the previous section: characterization of archetypes, technical development, and impact of the solution. The results allow the evaluation of the applicability of this new alternative technology in educational centres in the CAPV compared to conventional methods, the improvement of IAQ, energy consumption, and the social impact linked to its widespread implementation.

3.1 Results from the Study of the Applicability Archetypes

Results were obtained in terms of the number of pupils per classroom, area per classroom, area per pupil in the classroom, classroom dimensions, layout of furniture, distance between windows, window-ceiling distance, distance between lights and wall-lights, types of façades, compartmentalization, types of windows, existence of false ceilings, heating terminal units and type of thermal source, existence of renewable energies, etc. Two main results were obtained: on the one hand, the characterization of the existing stock of educational centres in the CAPV and, on the other hand, the characteristics of the classrooms.

Table 2 Construction year and number of educative buildings built in the CAPV

Construction period	Built schools per period	%	Built schools (cumulative)	%
Not available	84	12%	84	12%
<1941	48	7%	132	20%
1941–1960	48	7%	180	27%
1961–1980	255	38%	435	64%
1981–2007	186	28%	621	92%
2008–2011	40	6%	661	98%
>2011	14	2%	675	100%
Total	675	–	675	–

Regarding the existing stock, the results indicate that 92% of schools were built before mechanical ventilation was mandatory; thus, they probably do not have mechanical ventilation systems. Therefore, there are 661 school buildings on which this new solution could have a significant impact (see Table 2 and Fig. 4).

Regarding the characterization of applicability archetypes, the results obtained from the theoretical and field studies show a relative uniformity for certain characteristic parameters of the centres in the CAPV, particularly those related to the dimensional properties of the classrooms. The range of characteristic values, as well as the values for the typical classroom are indicated in Table 3 .

The theoretical studies and the fieldwork, point out that the target buildings for the ESKOAIRE project should preferably be those housing the primary and secondary education stages. The results from the gathered data show that generally classrooms do not have enough space to install

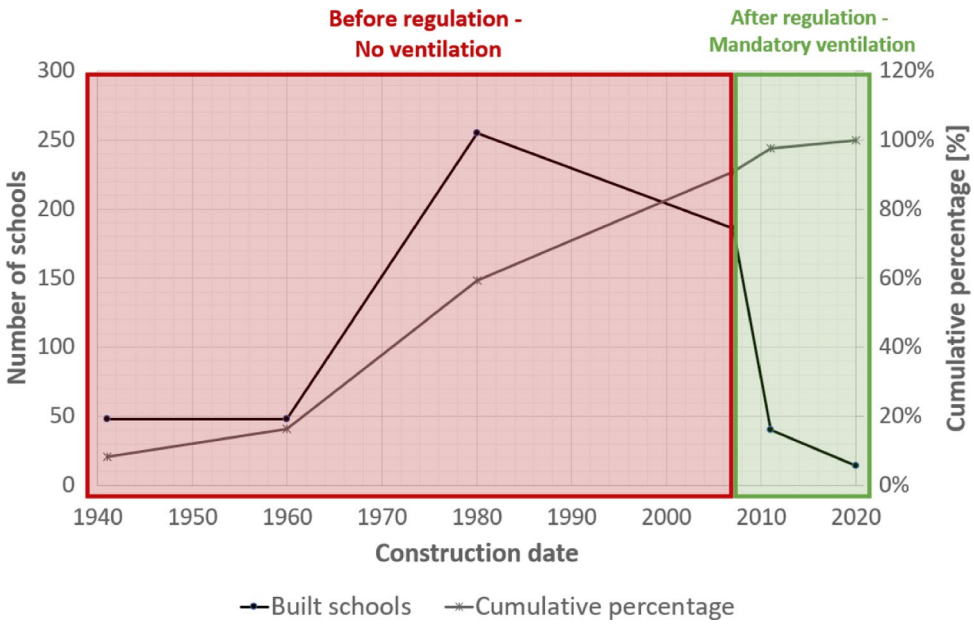


Fig. 4 Number of schools built from 1941 to 2021 in the CAPV based on data from Eustat (Eustat 2020)

Table 3 Characterization of the ESKOAIRE project-type classroom

	Predominant value	Typical classroom
No. of students	17–25	20
Classroom surface (m^2)	45–53	50
Facade wall (m)	6–7.5	7
Perpendicular to facade wall (m)	6–8	7.2
Height (m)	2.8–3.5	3
Desk layout	Single–Double	Single
Heating element	Hot water radiators under the windows	Hot water radiators under the windows

ventilation equipment with duct distribution, and therefore require a special and innovative solution to overcome this limitation.

Thus, for the design and development of the ESKOAIRE equipment, a typical classroom was considered with dimensions of $7 \times 7 \times 3$ m^3, an occupancy of 20 students, and a required ventilation flow rate of 250 l/s (following the indirect method).

3.2 Technical Developments

The resulting autonomous ventilation device includes supply and extract air sections, fans, filters, heat exchangers, and monitoring devices. Basic scheme of the developed device is presented in Fig. 5. Three prototypes were built to achieve the desired ventilation performance with adequate size, noise, and energy requirements. The final prototype was tested in the Kubik experimental installation in Tecnalia (Derio, Spain), and the ventilation strategies and use modes were preliminarily defined validating a previously performed Computational Fluid Dynamics (CFD) analysis. Besides, a configuration software, ventilation control system, and management and maintenance platform in the cloud were developed. The main characteristics of the ESKOAIRE solution classroom will be published in forthcoming articles.

3.3 Results from the Study of Solution Impact

The impact of the solution was evaluated from three perspectives: impact on improving air quality, energy impact, and, finally, social impact.

Impact on the Improvement of Air Quality

In order to evaluate the impact of the solution on improving air quality, two schools were selected as pilots. There, two typical classrooms were monitored in advance to the installation of ESKOAIRE devices. Then, ventilation units were installed in April 2023 (Fig. 6).

The indoor air quality and thermal comfort conditions were monitored until the end of school year to evaluate the spring–summer improvements in the indoor environmental quality (IEQ) while using ESKOAIRE ventilation devices. These classrooms will continue to be monitored until the end of 2023 to evaluate the fall-winter performance, which is not included in the present study.

The measurements taken before installing the equipment show that the CO_2 average concentration value during

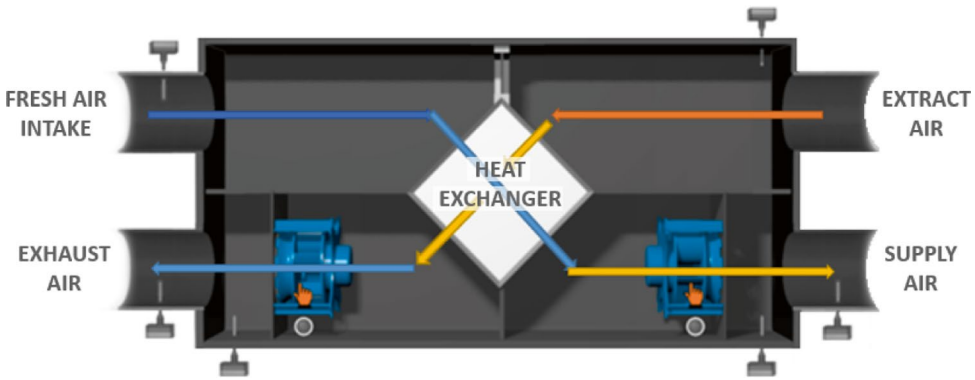


Fig. 5 ESKOAIRE Energy efficient ventilation device scheme. *Source* The authors

Fig. 6 ESKOAIRE devices installed at two pilot classrooms. *Source* The authors

school hours was 1973 ppm and the maximum value was 7380 ppm. During 70% of the school time in the whole measured period, the CO_2 values reached were higher than 1000 ppm, and during 52% of the time they were higher than 1500 ppm. It is worth noting that worse IAQ levels were reached in January, where the CO_2 concentration was above 1500 ppm during 91% of the school time. This could be justified by the fact that during this period, windows were opened less often. On the contrary, in October the values were much better as the windows were open most of the time. These values agree with the results obtained from a school located in Pamplona (northern Spain) which poses a similar climate type. 9 classrooms were monitored under the usual natural ventilation strategy during Spring (March). The mean values ranged from 1800 to 2900 ppm and the maximum values ranged from 4200 to 7100 (Monge-Barrio et al. 2022). Besides, the results of the monitorization of 9 schools in southern Spain during spring (March) show lower values, but in the same order of magnitude, considering that milder climates are more favourable to the opening of windows. The average values ranged from 1200 to 2000 ppm and the maximum values ranged from 1900 to 4100 (Becerra et al. 2020).

From the monitoring results of the pilot schools in the CAPV, it is observed that the air quality levels achieved with a natural ventilation strategy in monitored classrooms were insufficient, mainly because the air flow rate introduced cannot be controlled. Consequently, it may be concluded that natural ventilation strategies applied in these schools do not always guarantee an adequate air quality, so in order to have an efficient control it is necessary to combine them with mechanical ventilation, as previous studies have suggested (Canha et al. 2016; Merema et al. 2018).

The devices were installed, and first experiments were run. The maximum flow rate allowed for the equipment was set below the pre-set values to limit the noise inside the classroom (first tests and feedback from students and teachers indicated that the optimal ventilation level produced too much noise). The resulting flow rate setpoint was lower than what would be desired or required to achieve optimum CO_2 levels. The average CO_2 value after ESKOAIRE system installation was 919 ppm and the maximum value was 1415 ppm. CO_2 levels remained below the regulatory requirement of 1000 ppm during 62% of the school time. In addition, a correct energy cost vs. health balance was achieved in 94% of the teaching time by not exceeding 1200 ppm. In no case did they exceed 1500 ppm, which is the threshold value considered unhealthy. Table 4 shows the pilot CO_2 concentration during lecture time regarding key threshold values.

Figure 2 indicates that the ideal CO_2 levels for health and energy balance are in the range of 1000–1200 ppm, with more than 1500 ppm being regarded as unhealthy. This value is dramatically exceeded in the classrooms tested (variable occupancy of 15 to 30 students) before the installation of the ESKOAIRE devices, reaching values 6 times higher than recommended in a classroom. On some days peak values of up to 7000 ppm CO_2 were reached. Figure 7 compares the CO_2 concentration evolution with occupancy on a typical day before and after ESKOAIRE solution installation. These findings support the requirement

Table 4 CO_2 concentration in pilot classrooms before and after the ESKOAIRE installation

Lecture time before ESKOAIRE (%)	CO_2 concentration before ESKOAIRE (ppm)	Lecture time with ESKOAIRE (%)	CO_2 concentration with ESKOAIRE (ppm)
70	>1000	38	>1000
61	>1200	6	>1200
52	>1500	0	>1500

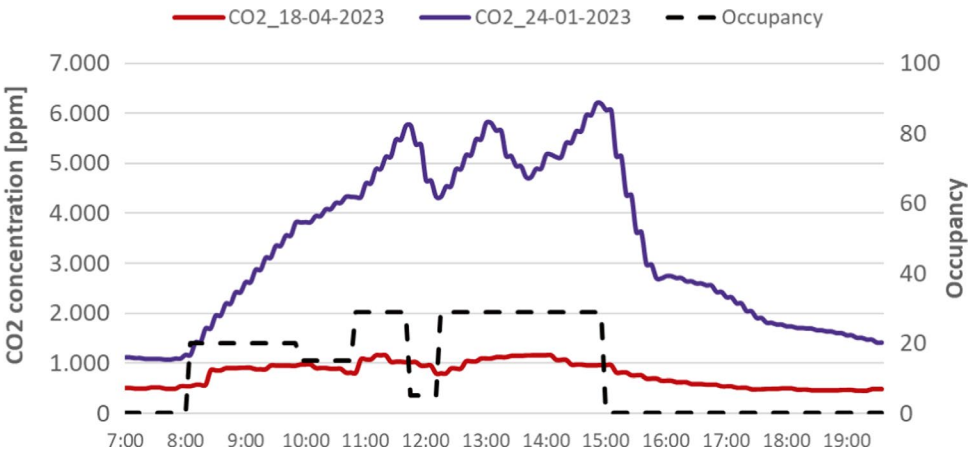


Fig. 7 Comparison between CO_2 concentration and temperature profiles of two monitored days at pilot classrooms: before and after the ESKOAIRE ventilation system. *Source* Prepared by the authors based on results from monitoring campaigns

for installing ventilation systems in the CAPV schools. It is concluded that the ESKOAIRE system improved the IAQ of the case study. The average CO_2 concentration during school hours was reduced by 53% and the maximum CO_2 concentration was reduced by 81%.

Impact on the Energy Consumption

Although natural ventilation strategies can be considered a very energy efficient way to improve air quality in spaces as they do not require the use of mechanical equipment, the lack of control over the temperature of the fresh air introduced through the windows results in a higher heating demand during winter.

The energy consumption evolution of actual pilots before and after the ESKOAIRE system installation could not be assessed since the buildings' heating systems were centralized and the ESKOAIRE solution was installed in one single classroom. Only data regarding the ESKOAIRE energy consumption are available.

Moreover, the energy performance of the present case study can be compared to existing systems with different functionalities. A campaign of dynamic energy simulations was carried out to compare the energy consumption of four consumption scenarios to evaluate the energy impact in the studied case study. Case 4 corresponds to the functionalities of the ESKOAIRE system. The results of these simulations, presented in Table 5, show that, for a typical classroom as defined above, the proposed solution with advanced demand control, heat recovery, and free-cooling provides heating demand savings of up to 97% if compared to a traditional constant flow ventilation strategy without heat recovery (Case 1). The implementation of variable flow and demand control ventilation (Case 2) can result in heating demand savings of 29%; if heat recovery is also included heating demand savings can increase to 97%. In addition, the system allows reducing the cooling demand by 81% by taking advantage of the free-cooling capacity of the system.

Table 5 Summary of energy consumption scenarios (% savings compared to Case 1)

	Case 1	Case 2	Case 3	Case 4
Mechanical ventilation	Yes	Yes	Yes	Yes
Variable flow—demand control (CO_2)	No	Yes	Yes	Yes
Heat recovery	No	No	Yes	Yes
Freecooling	No	No	No	Yes
Heating demand (kWh/m^2)	62,9	44,7	1,9	1,9
Cooling demand (kWh/m^2)	1,2	1,5	1,5	0,2
Heating demand savings	–	29%	97%	97%
Cooling demand savings	–	−32%	−32%	81%

Social Impact

According to the information received from school principals involved in the first phase of the project (26 schools), the implementation of solutions that can improve air quality in terms of pollutant concentration, but also improve indoor environmental conditions in general, is essential.

Based on the feedback obtained from both students and teachers who have participated in the pilot schools, the evaluation of the implementation of the ESKOAIRE solution has been very positive, with a high level of acceptance. Furthermore, some of the problems mainly highlighted were the situation of overheating in spring and the excessive noise caused in some centres by the opening of windows for ventilation. These problems could be addressed with the ESKOAIRE solution.

Therefore, the mass implementation of this type of solution could have a high social impact as it could improve the

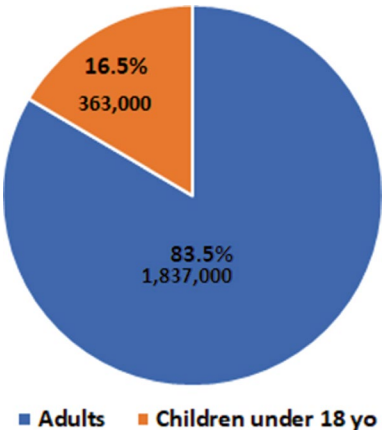


Fig. 8 Percentage of adults and children in the Basque Country based on data from municipal population statistics reports (Eustat 2023)

situation regarding limited IAQ in schools in the CAPV. Simoni et al. (2010) studied the correlation of health problems in children when exposed to different air pollutants, including mean CO_2 in classrooms in different countries. For example, the results point out that the prevalence of dry cough at night could be as high as 45% when classroom CO_2 mean concentration values were between 1000 and 2000 ppm. As stated before, several studies conclude that the consequences of poor IAQ in schools include absenteeism, health problems, and loss of school performance, among others, which ultimately translates into an economic impact. These problems are not easily quantified as it depend on many correlating factors (Chithra and Shiva Nagendra 2018).

The solution has an impact on an important sector of society, children. The CAPV population is 2.2 million people, 16.5% of whom are under 18 years of age (Fig. 8). The improvement of the IAQ potentially conducted by the installation of ventilation units in schools would improve the health and academic performance of around 360 thousand students in the non-university educational stages, from which around 300 thousand are under compulsory school age.

4 Conclusion

The studies carried out have revealed the problem of the lack of adequate ventilation in educational centres in the CAPV. It affects more than 300,000 pupils and can lead to widespread health problems, absenteeism, and other serious problems.

To determine the best way to solve the ventilation problem, a study regarding the characterization of CAPV school buildings was carried out. Results pointed out that out of 675 existing school buildings, 92% may not have controlled ventilation. Information regarding classroom construction parameters, occupation, layout, etc. was collected and validated by a fieldwork campaign consisting of visiting and collecting data from 26 school buildings. The case study was defined, and a technical solution was developed.

The ESKOAIRE project has developed a solution adapted to CAPV classroom characteristics, approaching air renovation in schools from a different perspective than traditional solutions based on duct ventilation systems. The solution consists of an autonomous demand-controlled ventilation system, comprised of compact ventilation units without ductwork, and CO_2 probe-based variable flow control that works independently for each classroom. The ESKOAIRE system was installed in two pilot classrooms.

The results allow the evaluation of the applicability, improvement of the IAQ, and the social impact of the widespread implementation of this new alternative technology in educational centres compared to conventional methods. The objectives of the study were successfully achieved and showed positive outcomes: a drastic reduction in CO_2 levels was achieved: the percentage of teaching time in which CO_2 levels remained below regulatory requirements increased from 30 to 62%. In no case did they exceed the threshold value considered unhealthy; a correct energy cost vs. health balance was achieved in 94% of the teaching time. Although the ventilation rates could be higher, the applied rates were adjusted to meet acoustic and thermal comfort. Thus, the ESKOAIRE application posed a great acceptance among users; it has been determined that the implementation of this solution in educational centres in the CAPV would have a significant energy impact, resulting in a 97% reduction in heating demand consumption compared to traditional systems with neither heat recovery nor variable flow/demand control. Above all, it would have a significant social impact by improving the health and well-being of more than 300,000 pupils under compulsory school age, potentially reducing absenteeism, health problems, and improving school performance linked to poor IAQ.

Acknowledgements This material is based on work developed within the scope of the ESKOAIRE project, which has received funding from the Basque Government's Hazitek Business R&D Support Programme and the European Regional Development Fund (FEDER) under the grant ZL-2021/00841.

References

Aguilar AJ, de la Hoz-Torres ML, Costa N, Arezes P, Martínez-Aires MD, Ruiz DP (2022) Assessment of ventilation rates inside educational buildings in Southwestern Europe: Analysis of implemented strategic measures. J Build Eng 51(January). https://doi.org/10.1016/j.jobe.2022.104204

ANSI/ASHRAE (2019) Ventilation for acceptable indoor air quality. In ANSI/ASHRAE Standard 62.1-2019

Asepeyo (2021) Buenas prácticas. Ventilación. Criterios para prevenir la transmisión de la COVID-19. https://www.hrome-extension://efaidnbmnnnibpcajpcglclefindmkaj, https://prevencion.asepeyo.es/wp-content/uploads/R1E20217V02-GUIA-BUENAS-PRACTICAS-Ventilación-COVID-19.pdf

Becerra JA, Lizana J, Gil M, Barrios-Padura A, Blondeau P, Chacartegui R (2020) Identification of potential indoor air pollutants in schools. J Clean Prod 242.https://doi.org/10.1016/j.jclepro.2019.118420

Canha N, Mandin C, Ramalho O, Wyart G, Ribéron J, Dassonville C, Hänninen O, Almeida SM, Derbez M (2016) Assessment of ventilation and indoor air pollutants in nursery and elementary schools in France. Indoor Air 26(3):350–365. https://doi.org/10.1111/ina.12222

Chica JA, Apraiz I, Elguezabal P, Rrips MO, Sánchez V, Tellado B (2011) Kubik: open building approach for the construction of an unique experimental facility aimed to improve energy efficiency in buildings. Open House Int 36(1):63–72. https://doi.org/10.1108/ohi-01-2011-b0008

Chithra VS, Shiva Nagendra SM (2018) A review of scientific evidence on indoor air of school building: Pollutants, sources, health effects and management. Asian J Atmos Environ 12(2):87–108. https://doi.org/10.5572/ajae.2018.12.2.87

CSIC-IDAEA, & Ministerio de Ciencia e Innovación y Mesura (2020) Guía para ventilación en aulas (p. 43). https://www.chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj, https://www.csic.es/sites/default/files/guia_para_ventilacion_en_aulas_csic-mesura.pdf

European Environment Agency (2019) Healthy environment, healthy lives: how the environment influences health and well-being in Europe (Issue 21)

Eustat (2020) Censo de edificios y locales. https://www.eustat.eus/estadisticas/tema_24/opt_1/tipo_1/ti_censo-de-edificios-y-locales/temas.html

Eustat (2023) Población de la C.A. de Euskadi por año de nacimiento, según el territorio histórico y el sexo. 01/01/2023. https://www.eustat.eus/elementos/tbl0011424_c.html

España (2020) ERESEE 2020 - Actualización de la Estrategia a largo plazo para la Rehabilitación Energética en el Sector de la Edificación en España. https://www.mitma.es/el-ministerio/planes-estrategicos/estrategia-a-largo-plazo-para-la-rehabilitacion-energetica-en-el-sector-de-la-edificacion-en-espana

Fenercom (2014) Guía de renovación de aire eficiente en el sector residencial. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.fenercom.com/wp-content/uploads/2014/03/Guia-de-Renovacion-de-Aire-Eficiente-en-el-Sector-Residencial-fenercom-2014.pdf

Gobierno de España (2021) Reglamento de instalaciones térmicas en los edificios (RITE). Boletín Oficial Del Estado, 1–97. https://www.boe.es/buscar/act.php?id=BOE-A-2007-15820

Harvard Healthy Buildings program (2020) 5-Step Guide To Checking Ventilation Rates in Classrooms. In Harvard Healthy Buildings program (p. 46). https://schools.forhealth.org/ventilation-guide/Heating, ventilation and air conditioning. Requirements relating to health. DIN 1946. (n.d.)

Mazaheri M, Reche C, Rivas I, Crilley LR, Álvarez-Pedrerol M, Viana M, Tobias A, Alastuey A, Sunyer J, Querol X, Morawska L (2016) Variability in exposure to ambient ultrafine particles in urban schools: comparative assessment between Australia and Spain. Environ Int 88:142–149. https://doi.org/10.1016/j.envint.2015.12.029

Merema B, Delwati M, Sourbron M, Breesch H (2018) Demand controlled ventilation (DCV) in school and office buildings: lessons learnt from case studies. Energy Build 172:349–360. https://doi.org/10.1016/j.enbuild.2018.04.065

Ministerio de Sanidad, & IDAE (2020) Recomendaciones de operación y mantenimiento de los sistemas de climatización y ventilación de edificios y locales para la prevención de la propagación del SARS-CoV-2. 15. https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Recomendaciones_de_operacion_y_mantenimiento.pdf

Monge-Barrio A, Bes-Rastrollo M, Dorregaray-Oyaregui S, González-Martínez P, Martin-Calvo N, López-Hernández D, Arriazu-Ramos A, Sánchez-Ostiz A (2022) Encouraging natural ventilation to improve indoor environmental conditions at schools. Case studies in the north of Spain before and during COVID. Energy Build 254:111567. https://doi.org/10.1016/j.enbuild.2021.111567

Rivas I, Viana M, Moreno T, Bouso L, Pandolfi M, Alvarez-Pedrerol M, Forns J, Alastuey A, Sunyer J, Querol X (2015) Outdoor infiltration and indoor contribution of UFP and BC, OC, secondary inorganic ions and metals in PM2.5 in schools. Atmos Environ 106:129–138. https://doi.org/10.1016/j.atmosenv.2015.01.055

Rodríguez-Vidal I, Martín-Garín A, González-Quintial F, Rico-Martínez JM, Hernández-Minguillón RJ, Otaegi J (2022) Response to the COVID-19 pandemic in classrooms at the university of the Basque country through a user-informed natural ventilation demonstrator. Int J Environ Res Public Health 19(21). https://doi.org/10.3390/ijerph192114560

Sadrizadeh S, Yao R, Yuan F, Awbi H, Bahnfleth W, Bi Y, Cao G, Croitoru C, de Dear R, Haghighat F, Kumar P, Malayeri M, Nasiri F, Ruud M, Sadeghian P, Wargocki P, Xiong J, Yu W, Li B (2022) Indoor air quality and health in schools: a critical review for developing the roadmap for the future school environment. J Build Eng 57(March):104908. https://doi.org/10.1016/j.jobe.2022.104908

Simoni M, Annesi-Maesano I, Sigsgaard T, Norback D, Wieslander G, Nystad W, Cancianie M, Sestini P, Viegi G (2010) School air quality related to dry cough, rhinitis and nasal patency in children. Eur Respir J 35(4):742–749. https://doi.org/10.1183/09031936.00016309

Toftum J, Kjeldsen BU, Wargocki P, Menå HR, Hansen EMN, Clausen G (2015) Association between classroom ventilation mode and learning outcome in Danish schools. Build Environ 92:494–503. https://doi.org/10.1016/j.buildenv.2015.05.017

Disruptive Business Models and Innovative Market Strategies: Digital Marketing and Social Innovations

Shaping Disruptive Solutions for Sustainable Futures: Zooming in on the Social in Socio-Technical Transformation

Taina Eriksson and Titiana Ertiö

Abstract

Creating sustainable futures is one of the grand challenges of our time and one that requires a suit of disruptive solutions to act in concert towards the shared goal. For too long now, businesses have focused disproportionately on maintaining the status quo through sustaining innovations. Our current technologies, with their interest in existing users' needs and product-market fit, miss opportunities to disrupt at scale. What is needed are disruptive solutions that tackle the sustainability deficiency. In the Science and Technology Studies field, it is well established that technology and society mutually shape each other. Thus, focusing on collective social needs, businesses can shape technologies such that they become fit for tackling sustainability issues. Today, businesses have opportunities to develop economically viable sustainable solutions. Based on signals from both industry and academia, we believe the time is ripe for disruptive solutions that incorporate social actors as active agents in the sustainability transformation. We propose a conceptual study that addresses the following research question: How might social interactions shape and drive disruptive solution development in businesses? To operationalize our research question, Christensen's theory of disruptive innovation sheds some light into the social aspects of technologies, for instance by focusing on the process rather than the product or service. Additional perspectives are needed to grasp the complex and systemic phenomenon of purposefully crafting disruptive solutions in the digital age, in particular around how technologies can be co-created among social actors with competing interests but united by the drive of solving grand challenges. By combining disruptive innovation theory with social shaping of technology and social construction of technology, we seek to understand how businesses might initiate, craft, and shape disruptive technologies together with social actors rather than just adopt otherwise sustaining innovations.

Keywords

Disruptive solution · Innovation · Interpretive flexibility · Social interaction · Social shaping · Sustainability disruption · Sustainability transition

1 Introduction

Plant-based proteins represent one of the innovations of our times, which respond to the sustainability gap between eating needs, global food ecosystem pressures, and planetary boundaries. Granted, producing protein out of plants involves sophisticated food and technological innovations. But what makes plant-based proteins disruptive is the fact that it has the potential to disrupt the food industry and farming at scale; long-held traditions and cultural aspects of food may gain new meanings. Commercially, the business model of plant-based proteins targets (still) niche customers, those environmentally and ethically aware. From a marketing perspective, products have emphasized similarities with meat-based products rather than the unique and distinctive qualities of plant-based proteins. These observations largely correspond to the initial stages of disruptive innovations outlined by Christensen and Raynor (2003), which we shall discuss shortly. However, responding to calls to reduce the environmental load globally, plant-based

T. Eriksson (✉) · T. Ertiö
Centre for Collaborative Research CCR, University of Turku, Turku, Finland
e-mail: taina.eriksson@utu.fi

T. Ertiö
e-mail: titiana.ertio@utu.fi

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_13

proteins pose all the "right ingredients" for sustainable disruptive innovation. Another example of a domain with considerable potential to contribute to sustainability transitions through disruptive innovation is bioeconomy (Losacker et al. 2023). Creating sustainable futures is a complex set of challenges and can be conceptualized as a grand challenge of our time; one that has profound implications for future generations and requires the collaboration of multiple actors globally (Scott 2021).

Most innovations are so-called sustaining innovations, which improve product performance against established evaluation criteria, such as those important to the mainstream customers (Christensen et al. 2018). Therefore, sustaining innovations target the needs of current customers or users. Even though sustaining innovations are an important and necessary part of sustainable development, we argue that disruptive innovations are the key to achieving sustainability transformation (Kivimaa et al. 2021). To borrow a common wisdom: "what got us here, won't get us there". In other words, solving the sustainability deficit and its challenges necessitates radical, disruptive innovation. Significant systemic changes that challenge established structures are called for in the pursuit of sustainable futures (Rutting et al. 2023).

In contrast to sustaining innovation, disruptive innovation targets the needs of potential customers and non-users or those customers that are not currently optimally served by existing services and offerings (Christensen et al. 2018). Thus, disruptive innovation looks beyond efficiency and optimization; it invites divergent thinking into expanding the customer base and anticipating their latent needs and wants, thus expanding existing markets or creating completely new markets. Disruptive innovations carry with themselves the potential to enforce sustainable development trajectories (Rutting et al. 2023).

The initial innovation that helps users in solving some challenges is only one piece in the puzzle. By an innovation, we mean an invention that typically builds on an enabling technology and it has been commercialized or adopted into use within an organization. To understand the sustainability implications and the development trajectory, it is important to examine how the use of the technology shapes people's interactions and behaviors as well as the implications the interactions and behaviors have on the technology. Therefore, to fully understand what kind of role disruptive innovations play in sustainability transformations and how the transformation unfolds, we need to take even broader approach and to consider the dynamics of social interactions.

Including the users into the analysis means extending the conceptualization of disruptive innovation from systems-of-innovation level to socio-technical system, which includes both the production domain of the innovations as well as the application domain (Geels 2004). Sustainability transitions call for significant change in both the production side as well as the consumption side. In addition to disruptive change on both of these domains, systemic change is needed (Kuokkanen et al. 2019).

In this conceptual study, we address the following research question: **How might social interactions shape and drive disruptive solution development in businesses?**

This study presents a theory-based model of the social shaping of disruptive solutions as part of sustainability transformations. Disruptive innovations are rarely discussed beyond an industry (see for example Christensen 1997), with some few notable exceptions (Schuelke-Leech 2018; Hopster 2021). Here, we present the sustainability potential of disruptive technologies and discuss how these might be shaped collectively bearing in mind the complex systemic nature of sustainability challenges. The paper is structured as follows. After the introductory section, we review the theory of disruptive innovation and thereafter look into sustainability transitions as a form of socio-technical transformation that is highly relevant for this study. We also draw insights from the social construction of technologies and social shaping of technologies to build a theory-based framework of disruptive solutions.

2 Disruptive Innovations

The theory of disruptive innovation dates back to Christensen's writings of disruptive technologies in the 1990s. In 2003, Christensen and Raynor coined the concept of disruptive innovation, which is a combination of technology and business model that is an innovation targeted initially at niche markets where there are potential customers unhappy with or even neglected by existing offerings (Christensen et al. 2018).

Since disruptive innovation introduces novel value or performance attributes, it tends to initially perform worse than established products or services along the evaluation criteria of the mainstream customers. However, it meets the needs of a particular niche and hence can start developing into a viable business. Over time the technology and its applications develop, and the performance soon meets the requirements of the mainstream customers as well (Christensen and Raynor 2003). Following the technology development s-curve, after the initial phases, the development accelerates and changes in performance can happen fast (Foster 1986). This is one of the reasons it is difficult to grasp disruptive innovation at early phases. Yet, there is some initial evidence that the theory of disruptive innovation can be used ex ante to identify disruptions (Christensen et al. 2018). This is an important aspect for this study that strives for understanding the proactive approach to

purposefully shaping the solutions that disrupt the market towards sustainability.

An important mechanism driving disruption is business organizations' pursuit for profitability and the fact that large incumbents and smaller entrants face differing logics of profitability. Whereas large companies with the resources to scale up innovation have to deal with the considerable cost of maintaining the organization, smaller ones have less liabilities in this regard, but also harder time scaling. Therefore, the limited volumes and thin margins of disruptive innovation are not as attractive to the incumbents as they are for the entrants (Christensen and Raynor 2003).

Disruptive innovation may be brought to the market with different kinds of strategies. The innovation can offer a simpler way of meeting a need than the existing products or services, which are overserving some customer segments. On the other hand, disruptive innovation may create a new market and solve a problem for a segment or segments that have been previously neglected (Christensen and Raynor 2003). Moreover, research has also found that there are disruptive innovations that are not more cost-efficient than earlier products or services, but that make something simpler and introduce novel value attributes, and hence gain acceptance in specific niche markets (Govindarajan and Kopalle 2006).

As an illustrative example, Allal-Chérif and his colleagues (2023) found that the radical disruption strategy employed by Patagonia helped them to stay profitable despite higher costs and investments. Patagonia only serves niche customers with a strong appreciation for the quality of their products. Additionally, the company has created an ecosystem of suppliers, employees, customers, and investors all of whom are aligned with the ecological and sustainability values of the founder.

In many prevailing systems, incumbent organizations have the position and power to drive stability and to resist change especially when the power position of the incumbents is widely accepted and perceived as legitimate in civil society (Geels 2014). On the one hand, this enables gradual optimization of existing ways of operating, e.g. efficiency as well as predictability. A recent study found that German manufacturing and industrial companies undertake incremental changes such as recycling or remanufacturing before venturing into more disruptive changes (Neligan et al. 2023). On the other hand, it hinders progress along deviating performance targets, that may be critical for creating sustainable futures.

Some researchers call for clarity in the discussion on disruptive innovations and distinguish between disruptive technologies, disruptive product innovations, and disruptive business models (Kivimaa et al. 2021). From the viewpoint of this study, all of these types are relevant and actually it is the combinations of these that appear the most interesting.

Some disruptive innovations have created significant change in the market, which confirms the potential of disruptive innovation to create change at scale. Kuokkanen and colleagues (2019, p. 751) note that "*Firms and their activities can influence and shape sustainability transitions purposefully (*York and Venkataraman 2010*), by shifting the market they operate in and transforming their own business"* (Loorbach and Wijsman 2013).

Nevertheless, considering the scale of the challenges connected to creating more sustainable futures, disruptive innovation in its original scope seems insufficient. In addition, sustainability has only recently emerged as an innovation driver for businesses beyond the first-movers. As there is the need to achieve even broader-reaching change at scale, we need to develop better understanding of how the dynamics unfold. Therefore, we turn next to the sustainability transformation literature and zoom in on the social aspects of the transformations.

3 Sustainability Transformations

Socio-technical transitions unfold through interaction between three levels within the system: niche innovations, regime, and landscape. Different combinations of timing and the nature of influence yield different outcomes (Geels and Schot 2007). Moreover, transitions differ in the speed (rapid vs. gradual) and outcome (large vs. small) of change (Geels 2018).

Geels and Schot (2007) define transformation as one type of socio-technical transition pathway, where there is moderate pressure from the landscape level towards the regime level at a time when niche innovations remain insufficiently developed. In this kind of situation, the regime actors may react in ways that change the development path and thus create transformation. As niche innovations develop, they mount internal momentum and at the same time changes in on the landscape level create pressures towards the regime, which at time destabilizes the regime and opens a window of opportunity for transformation (Geels 2014). The timing of these pressures and the nature of interaction between the three levels determine what follows from the mounting pressures (Geels and Schot 2007).

Kivimaa et al. (2021) reviewed research on disruption in sustainability transitions and paint the picture of what is being disrupted and what are the prerequisites for rapid and deep enough change. Their review identifies four dimensions, where disruptions in the sustainability transformation context may occur (in addition to technology):

(1) markets and business models,
(2) regulations, policies and formal institutions,
(3) actors and networks, and
(4) behavior, practices and cultural models.

Kivimaa et al. (2021) propose a revised definition of disruption in the socio-technical transitions context outlining that a disruption is "a large magnitude of change on the system which addresses more than one system dimension, either gradually or rapidly" (Kivimaa et al. 2021, p.122; system dimensions referring to the four dimensions identified in their review, see the list above). It must be noted that the socio-technical transitions are not deterministic, but the trajectories can be influenced through action (Geels 2014). Especially in the case of sustainability transformations, it is critical to pay attention not only to innovation but to the dynamics, that is the alignments and co-evolution between technology and society (Geels 2018).

Transitions, however, are slow. In an attempt to understand the speed needed in relation to the current sustainability imperative, Gross et al. (2018) discovered that the path from an initial invention into widespread commercialization of new energy technologies is a multi-decade process. Moreover, Blomsma et al. (2023) remind us that historical evidence shows it is anywhere between 20 and 70 years for socio-technical paradigm shifts; the authors provide businesses and change agents with heightened awareness on the possibilities of actions that could be taken to move towards more sustainable futures.

4 Social Interactions

As noted in the previous section by Geels (2014), agency and action play a crucial role in how trajectories of socio-technical transitions unfold. When investigating technological change, science and technology studies (STS) have long posited that technology and society mutually shape each other. It is particularly at the heart of this mutual shaping where technology can hold great potential in responding to and solving some of the most pressing societal challenges such as the sustainability deficit.

While Christensen and Raynor's (2003) theory of disruptive innovation has focused on the technological innovation and business models, a STS approach to disruptive innovations would typically highlight the importance of context and the factors shaping technologies. Following MacKenzie and Wajcman (1985, p. 2), it is crucial to interrogate: "*What has shaped the technology that is having "effects?*" *What has caused and is causing the technological changes whose "impact" we are experiencing?*".

STS and particularly Social Construction of Technology (SCOT; Pinch and Bijker 1984, 1987) conceptualizes technologies as socially constructed meaning that the particular design of a technology is the result of negotiations between several social groups and networks. For the purposes of this study, three central concepts of SCOT are relevant. Called "*relevant social groups*" in SCOT, these represent not only the producers and users of technology but more generally diverse groups including journalists, politicians, or activists. Importantly, users' role has also shifted; users not only adopt technologies but instead pro-actively shape technologies through feedback or feature requests. Indeed, the latest developments in user-centric design have paid a great deal of attention to the needs of (potential) users, often to the detriment of other social groups, such as marginalized groups, who would bear the disproportionate impact of such technologies but would have limited input into their development. Relevant social groups hold alternative interpretations and meanings of the problem definition or of a given technology; the capacity of technologies to allow such divergent meanings is called "*interpretative flexibility*" in SCOT. As groups interact, e.g. in debating their interpretations and meanings, they gradually negotiate shared meanings engaging in a collective process of sense-making known as "*closure*" in SCOT. It is important to note that closure is impermanent. Even when the technology's meaning is stabilized, i.e. the relevant social groups have agreed on a dominant design for the technology, new social groups are likely to employ new interpretative flexibility and new meanings, opening new technological design pathways.

5 Discussion

Against the background presented above, it is important to conceptualize sustainable disruptive innovation in a way that connects the end user to the innovator(s) level and leaves room for the end users' agency, while ensuring sustainability at scale. Framing disruptive innovation as a novel value proposition or value capture strategy does exactly this (Kuokkanen et al. 2019). From the business' perspective disruption at scale might also include participating in different eco-systems where an orchestrated effort towards sustainability takes place.

Given the urgency of our current sustainability imperative, there is a pressing need for creating the understanding of how the processes and impacts of disruptive solutions unfold (Rutting et al. 2023). To this end, this study synthesizes insights from multiple theoretical streams. The focus is specifically on social shaping, which is a typically neglected area in relation to disruptive innovations. Scrutinizing the phenomenon from the social interaction angle is necessary to discover critical factors for driving or hindering the impact of disruptive solutions.

The insights highlighted above, namely that socio-technical transitions are not deterministic (Geels 2014), the co-evolution between technology and society is critical to sustainability transformation (Geels 2018) and that technologies are the result of negotiations between several

social groups (Social Construction of Technologies; Pinch and Bijker 1984, 1987) suggest that a consolidated effort to promote collaboration across society is needed to advance sustainable futures.

In addressing our research question, we identify three ways in which social interactions and collaborations can drive disruptive solution development in businesses (Fig. 1).

Firstly, by fostering increased interpretative flexibility, i.e. divergent meanings and interpretations of the technology. This can be achieved for instance by requesting, enabling, and implementing feedback channels from diverse groups; making the technology available for anyone to use and modify (e.g. as open source), showcasing how a particular technology has multiple applications across different contexts and domains.

Second, by disrupting at scale—with a particular focus on the role of businesses, markets, and (new) business models. This two-pronged approach involves both the breath of diffusion of the business' activities as well as the novelty of their products and services. In today's networked society, disruption at scale takes part in ecosystems, with different stakeholders attempting to bring their own contribution and value to solving pressing issues. It depends on the companies and their leaders to sense the opportunities and implement disruptive solutions to mitigate the sustainability deficit.

Third, by pushing sustainability transformation—given the present sustainability deficit, businesses have access to resources, including talent and technology to pursue mission-driven disruptive solutions. At the societal level, there is an acute need to lean into the sustainability transformation, to reconceptualize measurement indicators such as GDPs and start factoring social and environmental indicators for growth. Unless the urgency of this required shift is understood, internalized, and enacted by businesses, governments, and citizens alike, incremental, business-as-usual strategies will prevail.

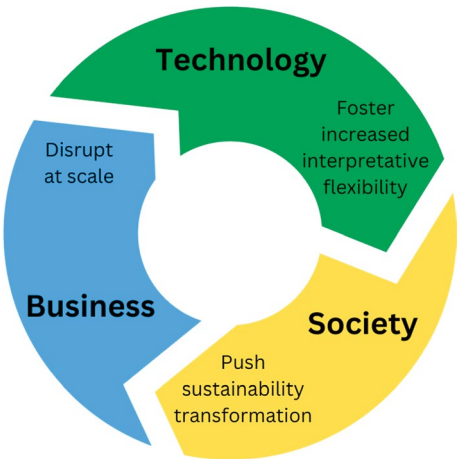


Fig. 1 Theory-based model for shaping disruptive solutions as part of sustainability transformations (*Source* The authors)

The proposed model will be now used to exemplify the social dimension in sustainability trajectories in food production. According to the World Economic Forum (2022), there is an urgent need for the food industry to address greenhouse gas emissions, employee safety, and consumer health; all these dimensions are part of the complex systemic challenges of food production sustainability. Digital solutions can address some of the aforementioned challenges, but their large-scale adoption is hampered by multiple barriers such as socio-cultural, institutional, economic as well as technological factors (Nyström and Giacometti 2022).

Increased interpretive flexibility offers a novel lens to overcome such challenges. Currently, there are unequal opportunities for companies in the food industry to access, apply, and scale digital technologies (Nyström and Giacometti 2022). Hence, a broader range of relevant stakeholders needs to be included in shaping the digital solutions. Stakeholders such as farmers, food producers, distributors, or vendors understand the challenges of the food system in depth and no off-the-shelf solution provided by technology companies will be adopted unless they are efficient in solving challenges while at the same time providing easy-to-use solutions. When a critical mass of stakeholders adopts solutions to tackle the sustainability deficit, the conditions are ripe for disruption. From the consumers' perspective, habits and food culture are factors affecting the sustainability transformation in the food industry (Valoppi et al. 2021). Businesses need to find ways of making the transition economically feasible, while addressing social and environmental goals. The increasing demand is critical for this, but so are societal incentives which can potentially play a big role.

All in all, the theoretical approaches reviewed in this study help us understand why and how some disruptions morph into a vicious cycle with negative unintended consequences (see for example Taneja and Maney 2022), while others morph into a virtuous cycle that reinforces sustainability benefits.

6 Conclusions

This study reviewed literature on disruptive innovation, sustainability transitions as a form of socio-technical transformation and drew insights from the social construction of technologies and social shaping of technologies. The overarching goal of this study has been to build a theory-based framework of disruptive solutions. We argued that disruptive solutions could tackle the sustainability deficiency and

outlined a theory-based model on how this change might be pursued.

This conceptual study has limitations that need to be acknowledged. As this is an initial understanding of conceptualizing disruptive solutions, the empirical validity of the framework remains for future research to examine. Moreover, different theoretical lenses and methods used would yield different perspectives into how disruption at scale, interpretative flexibility, and sustainability transformation become manifested in different contexts. This study is not intended to dissect the complexity of disruptive solutions, but rather to shed light on how stakeholders can actively engage in shaping sustainability transitions.

Further research could investigate ways how these phenomena play out in practice: how do businesses pursue sustainability-driven (technological) solutions at scale? What features do these solutions hold and what is their disruptive potential. By asking these questions, we invite fellow researchers to provide empirical evidence and test our model.

References

Allal-Chérif O, Climent JC, Berenguer KJU (2023) Born to be sustainable: how to combine strategic disruption, open innovation, and process digitization to create a sustainable business. J Bus Res 154. https://doi.org/10.1016/j.jbusres.2022.113379

Blomsma F, Bauwens T, Weissbrod I, Kirchherr J (2023) The 'need for speed': towards circular disruption—What it is, how to make it happen and how to know it's happening. Bus Strateg Environ 32(3):1010–1031. https://doi.org/10.1002/bse.3106

Christensen CM (1997) The Innovator's Dilemma. When new technologies cause great firms to fail. Harvard Business School Press, Boston, MA

Christensen CM, Raynor M (2003) The innovator's solution: creating and sustaining successful growth. Harvard Business School, Boston, MA

Christensen CM, McDonald R, Altman EJ, Palmer JE (2018) Disruptive innovation: an intellectual history and directions for future research. J Manage Stud 55:1043–1078. https://doi.org/10.1111/joms.12349

Foster RN (1986) Working the S-curve: assessing technological threats. Res Manage 29(4):17–20. https://doi.org/10.1080/00345334.1986.11756976

Geels FW (2004) From sectoral systems of innovation to socio-technical systems: insights about dynamics and change from sociology and institutional theory. Res Policy 33(6–7):897–920. https://doi.org/10.1016/j.respol.2004.01.015

Geels FW (2014) Regime resistance against low-carbon transitions: introducing politics and power into the multi-level perspective. Theory Cult Soc 31(5):21–40. https://doi.org/10.1177/0263276414531627

Geels FW (2018) Disruption and low-carbon system transformation: progress and new challenges in socio-technical transitions research and the multi-level perspective. Energy Res Soc Sci 37(2018):224–231. https://doi.org/10.1016/j.erss.2017.10.010

Geels FW, Schot J (2007) Typology of sociotechnical transition pathways. Res Policy 36(3):399–417. https://doi.org/10.1016/j.respol.2007.01.003

Govindarajan V, Kopalle PK (2006) The usefulness of measuring disruptiveness of innovations ex post in making ex ante predictions. J Prod Innov Manage 23:12–18. https://doi-org.ezproxy.utu.fi/10.1111/j.1540-5885.2005.00176.x

Gross R, Hanna R, Gambhir A, Heptonstall P, Speirs J (2018) How long does innovation and commercialisation in the energy sectors take? Historical case studies of the timescale from invention to widespread commercialisation in energy supply and end use technology. Energy Policy 123:682–699. https://doi.org/10.1016/j.enpol.2018.08.061

Hopster J (2021) What are socially disruptive technologies? Technol Soc 67:101750. https://doi.org/10.1016/j.techsoc.2021.101750

Kivimaa P, Laakso S, Lonkila A, Kaljonen M (2021) Moving beyond disruptive innovation: a review of disruption in sustainability transitions. Environ Innov Soc Trans 38(2021):110–126. https://doi.org/10.1016/j.eist.2020.12.001

Kuokkanen A, Uusitalo V, Koistinen K (2019) A framework of disruptive sustainable innovation: an example of the Finnish food system. Technol Anal Strat Manage 31(7):749–764. https://doi.org/10.1080/09537325.2018.1550254

Loorbach D, Wijsman K (2013) Business transition management: exploring a new role for business in sustainability transitions. J Clean Prod 45(April):20–28. https://doi.org/10.1016/j.jclepro.2012.11.002

Losacker S, Heiden S, Liefner I, Lucas H (2023) Rethinking bioeconomy innovation in sustainability transitions. Technol Soc 74:102291. https://doi.org/10.1016/j.techsoc.2023.102291

MacKenzie D, Wajcman J (eds) (1985) The social shaping of technology: how the refrigerator got its hum. Open University Press, Milton Keynes, England

Neligan A, Baumgartner RJ, Geissdoerfer M, Schöggl J-P (2023) Circular disruption: digitalisation as a driver of circular economy business models. Bus Strateg Environ 32(3):1175–1188. https://doi.org/10.1002/bse.3100

Nyström H, Giacometti A (2022) Discussion paper on Digitalisation of Food Systems. BioBaltic–project funded by Nordic Council of Ministers. http://nordregio.org/wp-content/uploads/2022/07/BioBaltic-Discussion-paper-on-Food-digitalisation.pdf. Accessed 15 Nov 2023

Pinch TJ, Bijker WE (1984) The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other. Soc Stud Sci 14(3):399–441. http://www.jstor.org/stable/285355

Pinch T, Bijker W (1987) The social construction of facts and artifacts: or how the sociology of science and the sociology of technology might benefit each other. In: Bijker W, Hughes T, Pinch T (eds) The social construction of technological systems: new directions in the sociology and history of technology. MIT Press, Cambridge, MA, pp 17–50

Rutting L, Vervoort J, Mees H et al (2023) Disruptive seeds: a scenario approach to explore power shifts in sustainability transformations. Sustain Sci 18:1117–1133. https://doi-org.ezproxy.utu.fi/10.1007/s11625-022-01251-7

Schuelke-Leech B (2018) A model for understanding the orders of magnitude of disruptive technologies. Technol Forecast Soc Change 129:261–274. https://doi.org/10.1016/j.techfore.2017.09.033

Scott D (2021) Sustainable tourism and the grand challenge of climate change. Sustainability 13(4):1966. https://doi.org/10.3390/su13041966

Taneja H, Maney K (2022) Intended consequences: how to build market-leading companies with responsible innovation. McGraw-Hill Education

Valoppi F, Agustin M, Abik F, Morais de Carvalho D, Sithole J, Bhattarai M, Varis JJ, Arzami ANAB, Pulkkinen E and Mikkonen KS (2021) Insight on current advances in food science and technology for feeding the world population. Front Sustain Food Syst 5:626227. https://doi.org/10.3389/fsufs.2021.626227

World Economic Forum (2022). https://www.weforum.org/agenda/2022/09/how-digital-technology-can-accelerate-food-sustainability/. Accessed 15 Nov 2023

York J, Venkataraman S (2010) The entrepreneur-environment nexus: uncertainty, innovation, and allocation. J Bus Ventur 25:449–463. https://doi.org/10.1016/j.jbusvent.2009.07.007

Unravelling Purchaser Retention: Exploring the Influence of Direct and Moderating Factors for Single and Multiple NFT Platform Purchasers

Chi Bo Wong and Yuqi Liang

Abstract

The market for non-fungible tokens (NFTs) has been experiencing significant growth recently, resulting in an increase in the number of platforms offering NFT services and intensifying competition amongst them. Despite the rapid adoption of NFTs and their crucial role in retaining purchasers, empirical investigation or understanding of NFT purchaser retention in the literature is scarce. Our study aimed to develop a comprehensive research framework that encompasses the direct effects of purchaser satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness on the retention of NFT purchasers towards their main NFT platforms to address this research gap. We also examined the moderating role of two strengthening moderators (trust and perceived usefulness) and two constraining moderators (switching costs and lack of alternative attractiveness) in the satisfaction–retention link. Furthermore, we aimed to identify purchaser groups (single and multiple NFT platform purchasers) and examine heterogeneity in the satisfaction–retention link in these two purchaser groups. Our approach will reveal the previously neglected effects on the retention of NFT purchasers towards their main NFT platforms. We conducted a large-scale online survey of NFT retail purchasers in Hong Kong. Results showed that the five direct effects of purchaser satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness have a significant and positive influence on purchaser retention. Additionally, the two strengthening and two constraining moderators have significant moderating effects on the satisfaction–retention link. In the examination of the heterogeneity between single and multiple NFT platform purchasers, the two strengthening moderators only play a significant moderating role in the satisfaction–retention link for single NFT platform purchasers. Meanwhile, the two constraining moderators only play a significant moderating role in the satisfaction–retention link for multiple NFT platform purchasers. This study concludes with a discussion of the practical and theoretical implications of the findings.

Keywords

Non-fungible tokens · Strengthening moderators · Constraining moderators · Purchaser satisfaction · Purchaser retention

1 Introduction

Non-fungible tokens (NFTs) have gained significant traction and emerged as a prominent asset in blockchain technology. Hayward (2023) asserted that the global user base for NFTs reached approximately 10.3 million in 2022 and is projected to grow to 19.31 million users by 2027. In 2022, the NFT market witnessed substantial trading activity, with an estimated trading volume of approximately US$24.7 billion across diverse blockchain platforms and marketplaces. The Hong Kong Special Administrative Region ranked second in terms of the highest number of Google searches for 'NFT' (surpassing Singapore but trailing behind Mainland China) from November 2021 to November 2022 (Statista Research Department 2023). Moreover, Hong Kong, along with Singapore, secured the third position in NFT ownership amongst adults, with 5%

C. B. Wong (✉) · Y. Liang
Department of Business Administration, Hong Kong Shue Yan University, Hong Kong, China
e-mail: cbwong@hksyu.edu

Y. Liang
e-mail: michaelyqliang@gmail.com

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_14

of the population owning an NFT as of September 2022 (De Best 2022). The NFT market is experiencing rapid growth due to advancements in blockchain technologies and the proliferation of NFT platforms, resulting in increased competition and raising concerns about purchaser retention.

Customer satisfaction has traditionally been regarded as a fundamental determinant of long-term consumer retention (Oliver 1980; Yi 1990). The more satisfied customers are, the greater is their retention. Accordingly, firms fundamentally strive to manage and increase customer satisfaction. However, satisfaction alone does not guarantee sustained retention (Lee et al. 2001; Ranaweera and Prabhu 2003; Wong 2011). Various factors, including trust, perceived usefulness, switching costs and lack of alternative attractiveness, also significantly influence retention, individually and in combination. Although the direct influence of these factors on retention is well-established in the literature, their role as moderating factors in the satisfaction–retention relationship has received limited attention. Furthermore, few studies have specifically categorised these moderators as either strengthening or constraining moderators. In this study, we not only consider trust and perceived usefulness as direct factors influencing retention but also recognise them as strengthening moderators within the satisfaction–retention link. Moreover, switching costs and lack of alternative attractiveness, in addition to their direct influence on retention, are acknowledged as constraining moderators within the satisfaction–retention link. In the context of purchaser retention in this study, the two strengthening moderators, namely, trust and perceived usefulness, can strengthen the influence of satisfaction on retention. Meanwhile, the two constraining moderators, namely, switching costs and lack of alternative attractiveness, weaken the relationship between satisfaction and retention.

The process of identifying purchaser groups and examining the heterogeneity in the satisfaction–retention link amongst these groups is crucial in facilitating targeted marketing, optimising resource allocation, enhancing purchaser retention, fostering product/service development and gaining a competitive advantage (Wong 2011). NFT platform providers can strategically tailor their approaches to effectively engage each specific group by comprehending the unique needs, preferences and behaviours exhibited by diverse purchaser groups. In this study, we have identified two key groups of purchasers: single and multiple NFT platform purchasers. Our objective is to explore the heterogeneity of the direct, strengthening moderating and constraining moderating factors influencing the satisfaction–retention link amongst these two groups.

Despite the widespread adoption of NFTs and the importance of purchaser retention for NFT platforms, empirical research and understanding regarding the direct and moderating factors influencing NFT purchasers' retention towards their main NFT platforms across different purchasers' groups are scarce. In this study, the main NFT platform is defined as the platform that a purchaser most frequently used to purchase NFTs at the time of the survey. The aim of this study is to develop a comprehensive research framework that investigates the factors influencing the retention of NFT purchasers towards their main NFT platforms. Specifically, the objectives of this study are to:

(a) investigate the simultaneous effects of the five direct factors on purchaser retention;
(b) examine the moderating role of the two strengthening and the two constraining moderators in the satisfaction–retention link; and
(c) explore the heterogeneity between single and multiple NFT platform purchaser groups in (a) and (b) above.

2 Research Framework and Hypotheses

In this study, the term 'purchaser' is used to distinguish individuals acquiring NFTs from traditional customers or consumers. Unlike customers or consumers who engage in transactions involving goods or services, the 'purchaser' emphasises the unique nature of acquiring digital assets, such as NFTs. NFTs are typically purchased for ownership, investment or collection purposes, rather than traditional consumption. Referring to individuals as 'purchasers' underscores their active role in acquiring NFTs and recognises the transactional aspect of participating in an NFT marketplace.

The effective management of purchase retention in the NFT platforms encounters significant challenges arising from three key factors. Firstly, market volatility introduces uncertainty as NFT prices fluctuate, potentially resulting in shifts in purchaser interest and disengagement. Secondly, intense competition amongst numerous platforms and artists poses challenges in differentiating and retaining purchasers within a highly competitive landscape. Lastly, the rapidly evolving trends and preferences within the NFT platforms make it challenging to accurately anticipate and cater to the changing tastes and interests of purchasers, thereby affecting retention efforts.

We develop a research framework linking purchaser satisfaction, trust, perceived usefulness, switching costs, and lack of alternative attractiveness to purchaser retention based on a review of the literature (Fig. 1). Our framework has three main features. Firstly, this framework examines the simultaneous effects of each of the five direct factors (purchaser satisfaction, trust, perceived usefulness,

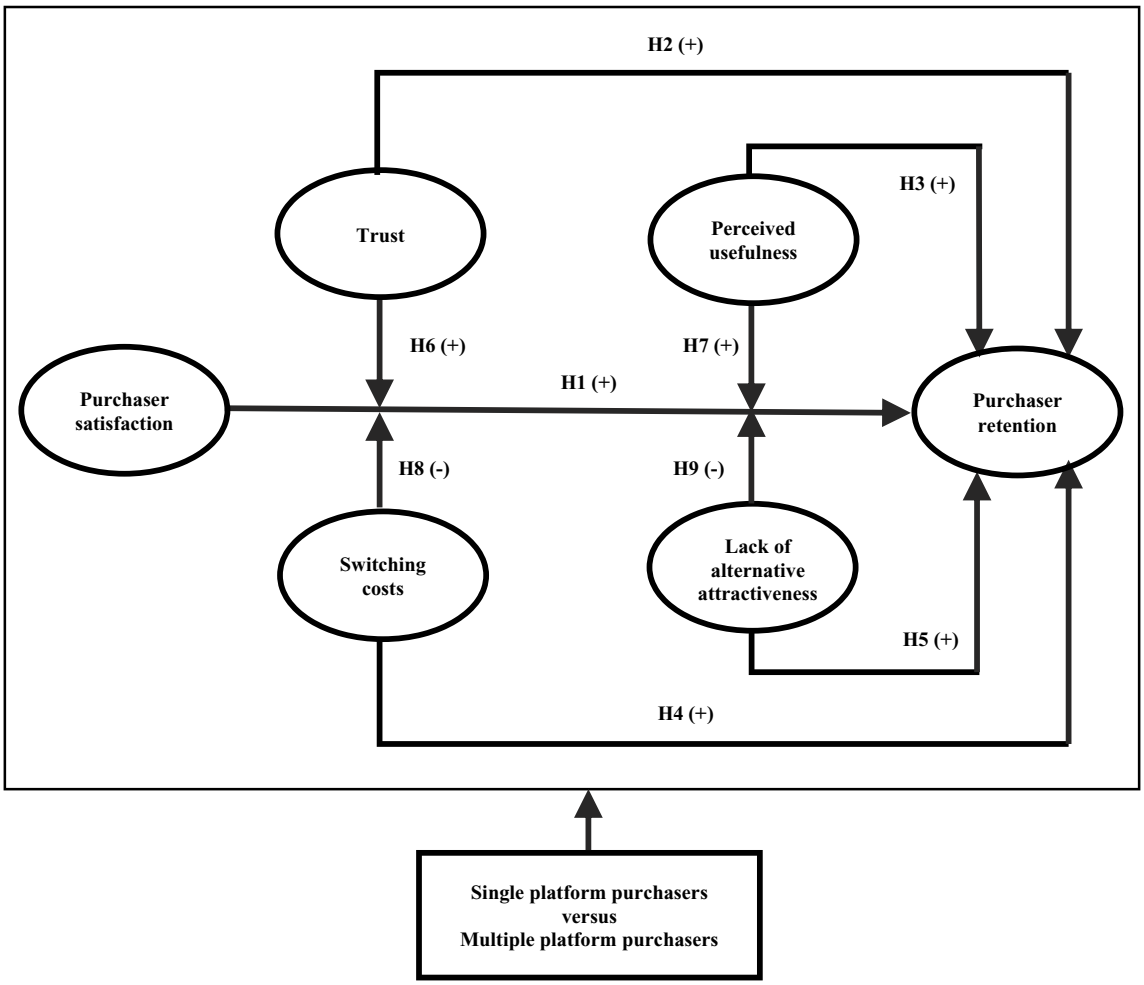


Fig. 1 Research framework

switching costs and lack of alternative attractiveness) on purchaser retention. Secondly, the framework examines the moderating effects of the two strengthening moderators (trust and perceived usefulness) and the two constraining moderators (switching costs and lack of alternative attractiveness) in the satisfaction–retention link. Thirdly, our framework examines the variation in the influence of the five direct factors, two strengthening moderators and two constraining moderators in the satisfaction–retention link in single and multiple NFT purchaser groups. We use this framework to develop our hypotheses as follows. This study aims to provide a comprehensive understanding of the factors influencing purchaser retention towards their main NFT platforms by considering these aspects, shedding light on the moderating effects and examining the divergence in different purchaser groups. This approach can uncover previously overlooked retention effects.

2.1 Purchaser Satisfaction as a Driver of Purchaser Retention

Numerous empirical studies have substantiated the theoretical proposition of a positive association between customer satisfaction and customer loyalty and retention. Satisfied customers exhibit heightened loyalty because their positive experiences engender emotional connections and diminish their inclination to switch to competing alternatives. Hadi et al. (2019) observed a significant elevation in loyalty amongst satisfied customers of cell phone service providers in Pakistan. Nguyen et al. (2020) identified a positive affiliation between satisfaction and loyalty in the context of electronic banking services in Vietnam. Furthermore, in the context of continuous purchasing, Ranaweera and Prabhu (2003) and Wong (2011) established significant and positive relationships between customer satisfaction and customer

retention amongst fixed-line telephone users in the UK and retail internet banking users in Hong Kong, respectively. These findings align with the fundamental tenets of the marketing concept, which emphasise the pivotal role of satisfying customer needs and wants in fostering customer retention (Kotler and Armstrong 2020). We propose the following hypothesis based on these previous research studies:

Hypothesis 1 (H1): Purchaser satisfaction is significantly and positively related to purchaser retention.

2.2 Trust as a Driver of Purchaser Retention

Extensive empirical evidence consistently supports a positive relationship between trust and customer retention. Trust assumes a pivotal role in establishing a sense of security and confidence within the customer–provider relationship, thereby fostering a heightened likelihood of customer repurchase intention and retention. This relationship has been observed across various contexts, such as online shopping in Indonesia and South Korea, where scholars, such as Sullivan and Kim (2018) and Wijaya and Astuti (2018), have identified the positive influence of trust on repurchase intention. In line with these prior research studies, we hypothesise that:

Hypothesis 2 (H2): Trust is significantly and positively related to purchaser retention.

2.3 Perceived Usefulness as a Driver of Purchaser Retention

The positive influence of the perceived usefulness of information systems on user retention is evident when user derive value from the system's features and functionalities, resulting in increased retention rates. Perceived usefulness, as proposed by the technological adoption model, is a key determinant of intention formation. This determinant has been initially developed in the organisational context and refers to the extent to which an individual's job performance is enhanced through the use of specific technology (Davis 1989). In the e-commerce literature, various factors contributing to usefulness have been identified, including faster search, improved search and buying processes, increased shopping productivity and enhanced shopping performance (Gefen and Straub 2000; Koufaris 2002). Wilson (2019) and Alshurideh et al. (2019) found significant positive effects of perceived usefulness on repurchase intentions in the electronic commerce industry in Indonesia and the electronic learning system in Jordan. In accordance with these previous research studies, we hypothesise that:

Hypothesis 3 (H3): Perceived usefulness is significantly and positively related to purchaser retention.

2.4 Switching Costs as a Driver of Purchaser Retention

Previous research has consistently demonstrated that high switching costs have a positive influence on customer retention, acting as deterrents and fostering loyalty and dependency on the current provider. Amoroso and Chen (2017) observed a positive relationship between switching costs and continuance intention amongst Chinese mobile consumers using financial apps in China. Lee et al. (2001) and Wong (2011) established significant and positive relationships between switching costs and customer retention in the contexts of mobile phone services in France and retail internet banking users in Hong Kong. We posit the following hypothesis in accordance with these prior research studies:

Hypothesis 4 (H4): Switching costs are significantly and positively related to purchaser retention.

2.5 Lack of Alternative Attractiveness as a Driver of Purchaser Retention

Prior studies have consistently demonstrated a positive relationship between the lack of attractive alternatives and customer retention. When customers perceive a scarcity of appealing options or unattractive alternatives in the market, their motivation to switch to alternative providers or products diminishes. Consequently, this reduced motivation strengthens customer loyalty and increases the likelihood of retaining customers with their current provider or product. Foroughi et al. (2023) found that the absence of appealing alternatives significantly contributes to user retention in the context of travel apps in Thailand. Koo et al. (2020) presented empirical evidence supporting the proposition that the lack of alternative attractiveness effectively retains hotel guests in South Korea. We propose the following hypothesis based on these prior research studies:

Hypothesis 5 (H5): Lack of alternative attractiveness is significantly and positively related to purchaser retention.

2.6 Moderating Effect of Trust in the Satisfaction–Retention Link

Trust plays a moderating effect in the satisfaction–retention link, acting as a protective mechanism that strengthens the positive influence of customer satisfaction on retention.

When customers have a high level of trust in a provider, product or service, their satisfaction is reinforced, resulting in a greater likelihood of continued patronage and retention. Additionally, trust can help mitigate the negative effects of potential dissatisfaction, particularly in cases of minor or occasional dissatisfaction. Minta (2018) and Ranaweera and Prabhu (2003) demonstrated the positive moderating effect of trust on the relationship between satisfaction and customer retention in the retail insurance industry in Mali and fixed-line telephone users in the UK. Thus, we propose the following hypothesis based on these earlier research studies:

Hypothesis 6 (H6): Trust has a significant and positive moderating effect on the relationship between purchaser satisfaction and purchaser retention.

2.7 Moderating Effect of Perceived Usefulness in the Satisfaction–Retention Link

The moderating effect of perceived usefulness plays a pivotal role in the satisfaction–retention connection. When perceived usefulness acts as a moderator, it affects the strength and direction of the relationship between customer satisfaction and retention. If customers perceive a product or service as highly useful, then it is likely to result in higher satisfaction levels, thereby increasing retention. When perceived usefulness is low, the influence of customer satisfaction on retention may be weakened. Accordingly, the moderating effect of perceived usefulness emphasises the significance of aligning customer perceptions of usefulness with satisfaction to cultivate long-term retention and loyalty. Shen et al. (2022) found that perceived usefulness moderates the influence of perceived benefits on sellers' willingness to retain. Consequently, the effect is stronger with the high perceived usefulness in the context of cross-border e-commerce sellers in China. Thus, we propose the following hypothesis based on these preceding research studies:

Hypothesis 7 (H7): Perceived usefulness has a significant and positive moderating effect on the relationship between purchaser satisfaction and purchaser retention.

2.8 Moderating Effect of Lack of Switching Costs in the Satisfaction–Retention Link

Customer satisfaction and switching costs are widely recognised as influential factors in determining customer retention. However, customer decisions to stay or switch when the switching costs are minimal are primarily driven by their level of satisfaction with the existing relationship. Grønhaug and Gilly (1991) asserted that dissatisfied customers may exhibit loyalty due to the significant barriers associated with switching. Lee et al. (2001) proposed that customer loyalty can arise from either satisfaction or dissatisfaction within a specific product category, particularly when the switching costs are relatively high, making it more challenging for customers to switch providers. Meanwhile, customer disloyalty can be a result of either dissatisfaction or satisfaction in a market where the switching costs are low, allowing customers to easily switch providers. In the context of continuous purchasing, Ranaweera and Prabhu (2003) and Wong (2011) have demonstrated significant and negative moderating effects of switching costs on the relationship between customer satisfaction and customer retention amongst fixed-line telephone users in the UK and retail internet banking users in Hong Kong. Thus, we propose the following hypothesis based on these existing scholarly studies:

Hypothesis 8 (H8): Switching costs have a significant and negative moderating effect on the relationship between purchaser satisfaction and purchaser retention.

2.9 Moderating Effect of Lack of Alternative Attractiveness in the Satisfaction–Retention Link

The lack of alternative attractiveness constrains the satisfaction–retention link. When attractive alternatives are scarce, customer satisfaction plays a more significant role in determining retention. When numerous appealing options are available, satisfaction has a limited influence on retention. The constraints imposed by the competitive landscape and the perceived attractiveness of alternatives must be considered when examining this moderating effect. When faced with dissatisfaction with a product or service, customers typically have the option to switch to a more suitable alternative. However, customers are compelled to remain in an unsatisfactory relationship when no alternative is available (Patterson and Smith 2003). In the context of mobile internet service in Malaysia, Chuah et al. (2017) revealed substantial and negative moderating effects of lack of alternative attractiveness on the association between customer satisfaction and customer retention. Kim et al. (2018) identified a negative moderating effect of lack of alternative attractiveness on the relationship between customer satisfaction and customer retention amongst users of freemium software applications in Korea. The following hypotheses are proposed in accordance with these past research studies.

Hypothesis 9 (H9): Lack of alternative attractiveness has a significant and negative moderating effect on the relationship between purchaser satisfaction and purchaser retention.

3 Methodology

This study was conducted in Hong Kong with the objective of developing a research paradigm that aligns with the local context. The research process encompassed four distinct stages: literature review, qualitative-focused interviews, pre-test questionnaire and main survey. In the first stage, the literature review involved an analysis of existing literature on factors influencing purchaser retention, including satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness. This comprehensive review served as the foundation for developing a theoretical research framework and formulating hypotheses.

In the second stage, a survey questionnaire was developed based on insights from the literature review, ensuring linguistic equivalence by translating it into English and Chinese versions (Brislin 1980). Thirty experts in NFTs participated in qualitative-focused interviews conducted via Zoom on 3 July 2023. The experts consisted of 10 academic lecturers specialising in digital marketing, 10 academic lecturers specialising in information technology and 10 experienced NFT purchasers. Convenience sampling was employed for this stage, and the participants were selected from the researcher's acquaintances, including classmates, friends and relatives. The interviews followed a question-by-question approach to ensure consistent comprehension and reliable responses. The participants provided valuable feedback, leading to the rephrasing of certain questions for improved clarity. The interviews took approximately 2 h to complete. The constructs of purchaser retention, satisfaction and trust were measured using three items each. Meanwhile, perceived usefulness and switching costs were measured using five items. The lack of alternative attractiveness was measured using four items. All items were adapted from previous studies and rated on a seven-point Likert scale, ranging from '1 = strongly disagree' to '7 = strongly agree'.

In the third stage, a pre-test questionnaire was administered to evaluate the reliability and validity of the measurement instruments. A total of 38 NFT purchasers completed the questionnaire on Zoom on 10 July 2023. The participants included 20 males and 18 females. 20 participants (52.63%) were 34 years old or below, 13 participants (34.21%) were 34–54 years old, and 5 participants (13.16%) were above 54 years old. Convenience sampling was employed, similar to the second stage. Internal consistency reliability was assessed using Cronbach's alpha coefficients. Higher Cronbach's alpha values, typically above 0.70 as recommended by Nunnally (1978), indicated stronger internal consistency. The pre-test results showed favourable Cronbach's alpha values for each construct: purchaser retention (0.940), purchaser satisfaction (0.919), trust (0.820), perceived usefulness (0.912), switching costs (0.903) and lack of alternative attractiveness (0.906). These values exceeded the recommended threshold of 0.7 for Cronbach's alpha. Consequently, the pre-test findings were deemed satisfactory as all constructs demonstrated strong internal consistency, as indicated by the Cronbach's alpha values.

In the last stage, the final version of the questionnaire was developed using the Google Forms platform after completing the pre-test questionnaire in the third stage. The questionnaire was published on Google on 16 July 2023, and the data collection took place for a period of 15 days, concluding on 30 July 2023. A cover message was prepared to clarify the research objectives and assure respondents of their anonymity. Convenience sampling was employed, similar to stages two and three. The online questionnaire was distributed to the researcher's acquaintances, including classmates, friends and relatives, through instant messaging platforms, such as Email, WhatsApp, Signal and WeChat. The participants were also encouraged to share the online questionnaire's URL with others to expand the participant pool for the study. The online questionnaire aimed to examine the direct and moderating effects of trust, perceived usefulness, switching costs and lack of alternativeness in the satisfaction–retention link. The data analysis was performed using Smart partial least squares (SmartPLS) software. The target population for the main survey consisted of individuals aged 18 years old or older who had purchased at least one NFT within the preceding 12 months. Apart from these two criteria, no additional requirements were imposed on the target population. Given that the research was conducted in Hong Kong, majority of the respondents would be Hong Kong Chinese. Regarding the sample size required for structural equation modelling (SEM), Hair et al. (2011) proposed that determination of the ideal sample size does not follow a fixed rule. Instead, researchers should consider factors, such as the model's complexity, desired statistical power and expected effect size. They emphasised the significance of having a sufficient sample size to ensure reliable estimation and statistical power. They recommended a minimum sample size of 100–200 participants for exploratory research, whilst confirmatory research would benefit from a larger sample size of 200–400 participants. Despite the exploratory nature of our study, our objective was to surpass 400 valid respondents for the main survey. The questionnaire for the main survey was designed in three sections. The first part consisted of a cover letter explaining the survey's purpose and guaranteeing respondent anonymity. The

second part included items and their definitions related to the constructs under investigation (as summarised in Appendix 1). The final section encompassed demographic questions regarding the respondents' personal profiles.

4 Findings

A total of 442 questionnaires were returned, but 24 (5.4%) of them were completed by individuals who were not part of our target group: Hong Kong residents. Finally, 418 out of the 442 questionnaires were included in the study. The sample consisted of 255 males (61%) and 163 females (39%). Amongst the respondents, 198 (47.37%) were 34 years old or below, 206 (49.28%) were 34–54 years old, and 14 (3.35%) were above 54 years old. In terms of monthly income, 138 respondents (33.01%) had a personal monthly income below HK$15,000, 243 (58.13%) reported an income level between HK$15,000 and 29,999, and 37 (8.86%) had an income level exceeding HK$29,999. Additionally, 47 participants (11.24%) had a secondary education level or lower, 66 (15.79%) had completed postsecondary education or obtained a diploma, and 305 (72.97%) had a tertiary/university level of education or higher.

4.1 Common Method Variance

After the survey, the 418 selected data points were analysed using Harman's single factor test (1976) to assess the common method variance before proceeding with further data analysis. Harman's single-factor test involves conducting a factor analysis to determine if a dominant factor explains a significant portion of the variance amongst variables, indicating the presence of common method bias. The generally accepted criterion for the single factor extracted is below 50% (Podsakoff et al. 2003). In this study, the computed single factor was 42.56% in SPSS 28, indicating that no single factor significantly influenced the majority of the variance from the single data source.

4.2 Examination of the Underlying Structure of Items

The initial data examination aimed to determine the underlying structure of items, eliminating insignificant ones for simplicity (Hair et al. 2009). This examination involved three steps. Firstly, the Bartlett test of sphericity and KMO test were conducted, resulting in a KMO value of 0.937 ($p < 0.001$), indicating sufficient data adequacy for factor analysis. Secondly, principal component analysis was performed to identify data relationships and extract factors with eigenvalues greater than one. The analysis revealed six constructs with corresponding cross-loadings and factor loadings, explaining a total variance of 82.96%. Lastly, the skewness and kurtosis values were computed to assess the departure from normality for each construct. The recommended acceptable range for skewness and kurtosis was an absolute value not exceeding 3 and 10, respectively (Kline 2011). The data did not exhibit significant deviations from normality because all skewness and kurtosis values for the items fell within these acceptable limits.

4.3 Partial Least Squares Analysis

A partial least squares (PLS) model comprises an outer model and an inner model, with the outer model examining relationships between latent and observed variables, whilst the inner model focuses on the hypothesised relationships amongst latent variables (Chin 1998). PLS, similar to the other structural equation modelling (SEM) approaches, provides statistical outputs, such as factor loadings, average variance explained and composite reliabilities (CRs), to evaluate the model validity and reliability (Fornell and Cha 1994). Bootstrap techniques are employed to calculate the t-statistic and standard error for each parameter and address the absence of formal significance tests in nonparametric methods (Chin 1998). In this study, a random bootstrap set of 3000 samples was utilised to ensure reliable standard errors and minimise discrepancies between overall sample estimates and subsample means (Leger et al. 1992).

Evaluation of the Measurement Outer Model

Various assessments were conducted, including the examination of component loadings for each item and tests to evaluate reliability, convergent validity and discriminant validity of the six constructs, to determine the appropriateness of the measurement model (Hair et al. 2014). The factor loading of each item in PLS was assessed. The findings revealed that all 23 items had factor loading measures exceeding 0.70, and the average variance extracted (AVE) for each construct exceeded 0.50. These results indicate compliance with the recommended criteria for convergent validity (Hair et al. 2011) and confirm the measurement model's satisfactory convergent validity.

The Cronbach's alpha and CRs were utilised to assess reliability. The Cronbach's alpha was estimated using the reliability analysis procedure in SPSS, whilst the composite analysis was derived from the principal component analysis procedure in PLS (Chin 1998). Table 1 presents the results, demonstrating that the reliability measures for all constructs surpassed 0.70, meeting the suggested criteria for reliability (Nunnally 1978). This result further suggests that the

Table 1 Quality criteria of the constructs

Latent variables	Items	Mean	Standard deviation	Standardised outer loadings	*p*-value	AVE	CR	Cronbach's alpha
Perceived Satisfaction (PS)	PS1	3.72	0.923	0.868	0.000	0.790	0.919	0.867
	PS2	3.66	1.050	0.896	0.000			
	PS3	3.66	0.990	0.903	0.000			
Trust (TR)	TR1	3.79	0.891	0.893	0.000	0.786	0.917	0.864
	TR2	3.88	0.936	0.885	0.000			
	TR3	3.76	0.989	0.881	0.000			
Perceived Usefulness (PU)	PU1	3.80	1.020	0.840	0.000	0.743	0.935	0.914
	PU2	3.63	1.031	0.879	0.000			
	PU3	3.66	1.068	0.891	0.000			
	PU4	3.67	1.090	0.866	0.000			
	PU5	3.90	1.047	0.834	0.000			
Switching Costs (SC)	SC1	4.05	1.115	0.888	0.000	0.808	0.955	0.941
	SC2	3.92	1.199	0.891	0.000			
	SC3	4.02	1.254	0.898	0.000			
	SC4	4.01	1.243	0.903	0.000			
	SC5	4.06	1.306	0.914	0.000			
Lack of Alternative Attractiveness (LAA)	LAA1	4.28	1.281	0.917	0.000	0.855	0.959	0.943
	LAA2	4.12	1.288	0.927	0.000			
	LAA3	4.02	1.394	0.941	0.000			
	LAA4	4.06	1.195	0.912	0.000			
Purchaser Retention (PR)	PR1	4.53	1.099	0.920	0.000	0.861	0.949	0.919
	PR2	4.53	1.198	0.927	0.000			
	PR3	4.64	1.163	0.936	0.000			

measurement model exhibits favourable internal consistency reliability. All reflective indicators are significant at a 1% level, based on the significance levels (*p*-value). Thus, all indicators in the reflective constructs are retained. The satisfactory levels of quality are exhibited by all reflective constructs, allowing for the evaluation of the structural model to proceed. A summary of these findings can be found in Table 1.

Discriminant validity was evaluated through the analysis of cross-loadings of the indicators and the application of the Fornell–Larcker criterion (Fornell and Larcker 1981). Firstly, if an item's factor loadings on its designated construct are greater than its cross-loadings on other constructs, then this provides evidence of discriminant validity. The range of factor loading and cross-loading for each construct is summarised in Table 2. Each item's measurement was aligned with the construct it aimed to represent and did not exhibit a stronger association with any other construct. Secondly, the Fornell–Larcker criterion compares the square roots of the AVE for reflective constructs with the correlations between the constructs, thus confirming the discriminant validity of the measurement model. Table 3 presents the square root of AVE for each construct, which exceeds the inter-construct correlations. Consequently, this measurement model demonstrates satisfactory discriminant validity.

Evaluation of the Structural Inner Model

The structural model was systematically validated in three steps (Hair et al. 2014): (1) conducting a collinearity assessment, (2) evaluating the significance and relevance of the relationships within the structural model and (3) examining the predictive power of the model using R^2. In the first step, collinearity was assessed using the variance inflation factor (VIF), which is the inverse of the tolerance. The VIF values for trust, perceived usefulness, switching costs, lack of alternative attractiveness and purchaser satisfaction as predictors of purchaser retention were 1.942, 2.479, 1.169, 1.267 and 2.768, respectively. All VIF values were below the threshold of five (Hair et al. 2011), indicating the absence of collinearity issues amongst the predictor variables.

The second step of the study involved investigating the relationships between the constructs of the structural

Table 2 Examination of cross-loadings and factor loadings

	LAA	PU	PR	PS	SC	TR
LAA1	**0.917**	0.428	0.696	0.321	0.220	0.312
LAA2	**0.928**	0.381	0.618	0.245	0.120	0.249
LAA3	**0.941**	0.420	0.636	0.315	0.013	0.260
LAA4	**0.912**	0.439	0.629	0.349	−0.011	0.292
PR1	0.609	0.768	**0.921**	0.684	0.373	0.685
PR2	0.656	0.732	**0.927**	0.601	0.430	0.640
PR3	0.681	0.747	**0.936**	0.572	0.422	0.589
PS1	0.264	0.640	0.580	**0.868**	0.094	0.571
PS2	0.301	0.659	0.597	**0.896**	0.116	0.563
PS3	0.323	0.644	0.602	**0.903**	0.118	0.593
PU1	0.354	**0.840**	0.697	0.610	0.232	0.478
PU2	0.457	**0.879**	0.728	0.644	0.163	0.481
PU3	0.478	**0.891**	0.734	0.606	0.195	0.429
PU4	0.343	**0.866**	0.666	0.640	0.197	0.466
PU5	0.299	**0.834**	0.647	0.644	0.215	0.462
SC1	0.204	0.277	0.481	0.165	**0.888**	0.329
SC2	0.028	0.207	0.362	0.101	**0.891**	0.276
SC3	0.076	0.136	0.353	0.041	**0.898**	0.248
SC4	−0.034	0.144	0.318	0.047	**0.903**	0.263
SC5	0.101	0.239	0.420	0.164	**0.914**	0.303
TR1	0.261	0.494	0.621	0.619	0.256	**0.893**
TR2	0.306	0.409	0.599	0.479	0.331	**0.885**
TR3	0.237	0.523	0.608	0.622	0.267	**0.881**

Table 3 Discriminant validity

	TR	PU	SC	LAA	PS	PR
TR	***0.887***					
PU	0.537	***0.862***				
SC	0.321	0.231	***0.899***			
LAA	0.302	0.451	0.096	***0.924***		
PS	0.648	0.728	0.123	0.333	***0.889***	
PR	0.688	0.807	0.440	0.699	0.667	***0.928***

Note The diagonal italic elements are square roots of AVE. According to the Fornell–Larcker criterion, the assessment is to compare the square roots of the reflective construct's AVE with the correlations between the constructs

model using the bootstrapping method. We employed the bootstrapping procedure, utilising 3000 samples and without sign changes. The analysis yielded three main findings related to the nine hypotheses under investigation (Table 4). Firstly, concerning the direct effects of purchaser satisfaction (PS), trust (TR), perceived usefulness (PU), switching costs (SC) and lack of alternative attractiveness (LAA) on purchaser retention (PR), the results indicate that all five direct effects are positive and statistically significant (purchaser satisfaction: $\beta = 0.122$, $p < 0.01$; trust: $\beta = 0.140$, $p < 0.01$; perceived usefulness: $\beta = 0.247$, $p < 0.01$; switching costs: $\beta = 0.227$, $p < 0.01$; lack of alternative attractiveness: $\beta = 0.475$, $p < 0.01$), thereby confirming hypotheses H1, H2, H3, H4 and H5. Additionally, the findings indicate that lack of alternative attractiveness has a stronger influence on purchaser retention compared with the other four direct effects. Secondly, regarding the moderating effects of the two strengthening moderators of trust and perceived usefulness in the satisfaction–retention link, the results provide support for hypotheses H6 and H7. Specifically, the moderating effects of trust ($\beta = 0.084$, $p < 0.01$) and perceived usefulness ($\beta = 0.118$, $p < 0.01$) are positive and significant,

Table 4 Results of hypothesis testing

Hypothesis	Path	Standard path coefficient (β)	Standard error	*t*-value	*p*-value	Sign	Result
H1	PS → PR	0.122	0.030	4.086	0.000	**	Supported
H2	TR → PR	0.140	0.029	4.889	0.000	**	Supported
H3	PU → PR	0.247	0.046	5.356	0.000	**	Supported
H4	SC → PR	0.227	0.028	8.115	0.000	**	Supported
H5	LAA → PR	0.475	0.026	17.986	0.000	**	Supported
H6	TR × PS → PR	0.084	0.018	4.583	0.000	**	Supported
H7	PU × PS → PR	0.118	0.028	4.283	0.000	**	Supported
H8	SC × PS → PR	−0.061	0.029	2.137	0.034	*	Supported
H9	LAA × PS → PR	−0.066	0.021	3.224	0.001	*	Supported

Note Sign. = significance (*$p<0.05$, **$p<0.01$)

aligning with the hypothesised relationships. This finding suggests that trust and perceived usefulness act as complements to satisfaction, further enhancing purchaser retention. Consequently, in the absence of trust and perceived usefulness, the influence of satisfaction on retention is likely to be diminished. Thirdly, concerning the moderating effects of the two constraining moderators of switching costs and lack of alternative attractiveness in the satisfaction–retention link, the findings support hypotheses H8 and H9. Specifically, the moderating effects of switching costs ($\beta = -0.061$, $p<0.05$) and lack of alternative attractiveness ($\beta = -0.066$, $p<0.01$) are negative and significant, consistent with the hypothesised relationships. This finding suggests that switching costs and lack of alternative attractiveness act as constraints, preventing dissatisfied purchasers from leaving their main NFT platform. Thus, under appropriate circumstances, switching costs and lack of alternative attractiveness can be effective alternative means of improving purchaser retention.

The third step involved evaluating the explanatory power of the structural model by using the R^2 value of the dependent variables (Hair et al. 2014). The R^2 value of the dependent variable of retention was 0.562, indicating that 56.2% of the variance in customer retention could be explained using the five direct factors, the two strengthening moderators and the two constraining moderators. This value demonstrates the moderated explanatory power of the model used in this study (Chin 1998). Figure 2 displays the standardised outer loadings of items within each construct, along with their respective *p*-values. Additionally, the aforementioned figure presents the β values and corresponding *p*-values for each hypothesis.

Group Analysis

Our findings indicate that purchaser satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness have a significant and positive influence on purchaser retention for the single and multiple NFT platform purchaser groups. In the investigation of heterogeneity between single and multiple NFT platform purchasers, amongst single NFT platform purchasers, only the two strengthening moderators of trust ($\beta = 0.1488$, $p<0.01$) and perceived usefulness ($\beta = 0.1794$, $p<0.01$) exclusively display a significant moderating effect in the satisfaction–retention relationship. Meanwhile, only the two constraining moderators of switching costs ($\beta = -0.2137$, $p<0.01$) and lack of alternative attractiveness ($\beta = -0.1083$, $p<0.05$) solely demonstrate a significant moderating effect on the satisfaction–retention relationship amongst multiple NFT platform purchasers. The overall results of group analysis indicate that the moderating effects of two strengthening moderators and two constraining moderators on the relationship between satisfaction and retention vary between single and multiple NFT platform purchasers. The results are depicted in Table 5.

The findings suggest that NFT platforms and similar online shopping platforms should prioritise several key factors, including purchaser satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness. These factors hold significance for single and multiple NFT platform purchasers. In single NFT platform purchasers, trust and perceived usefulness emerge as crucial factors that contribute to strengthening the satisfaction–retention relationship. Meanwhile, switching costs and lack of alternative attractiveness have a nonsignificant influence on constraining the satisfaction–retention link amongst this group. In multiple NFT platform purchasers, trust and perceived usefulness are not considered crucial factors in strengthening the satisfaction–retention relationship. However, switching costs and lack of alternative attractiveness play a significant role in constraining the satisfaction–retention link within this group. Platforms should tailor their strategies based on the specific behaviours and characteristics exhibited by both single and multiple NFT platform purchasers to effectively retain customers and cultivate long-term relationships.

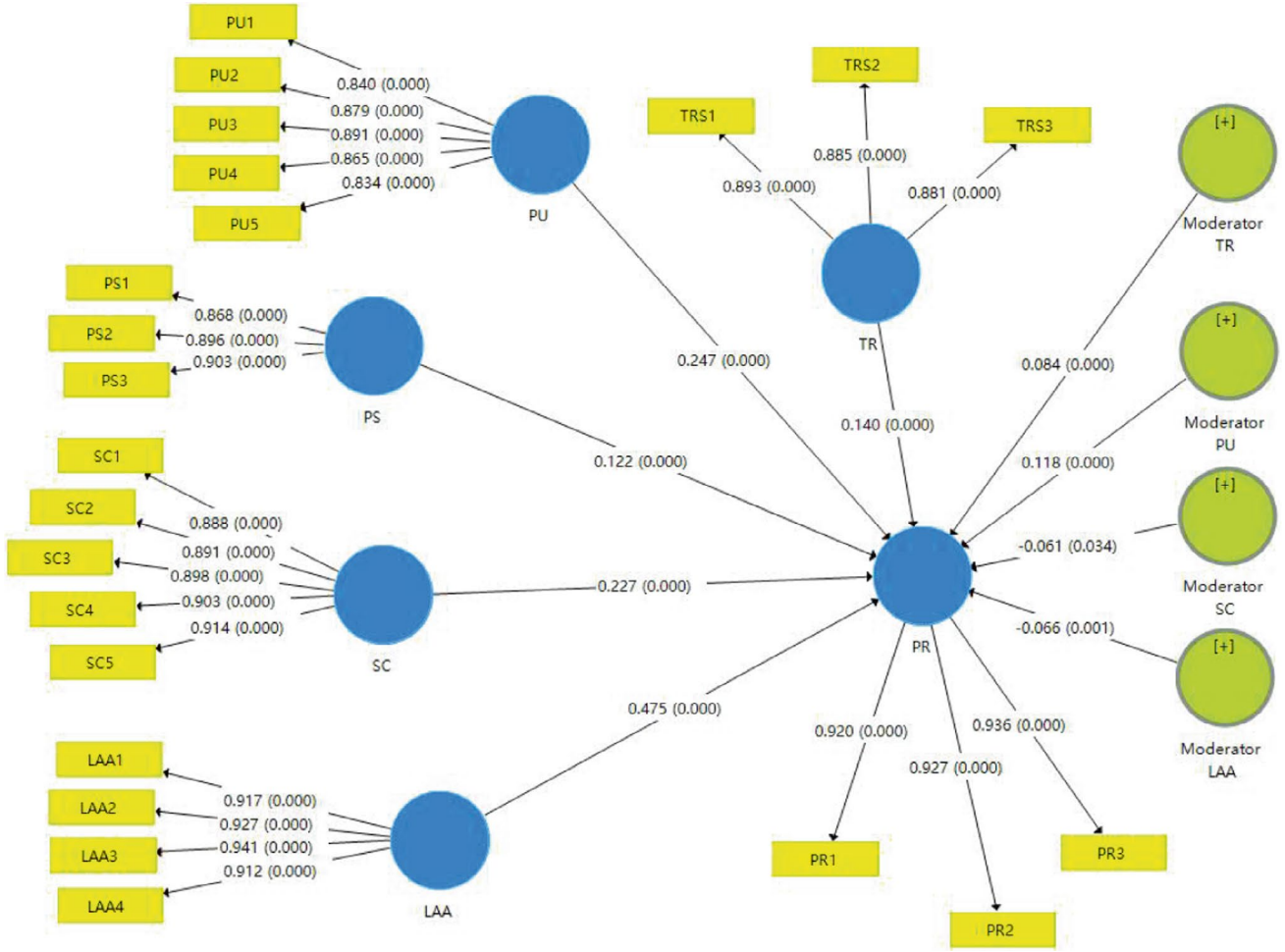


Fig. 2 Results of the research framework (H1–H9)

Table 5 Significance of the testing results of the single and multiple platform purchasers

Hypothesis	Path	Single platform purchasers ($n = 120$)			Multiple platform purchasers ($n = 298$)		
		Std. beta	*t*-value	*p*-value	Std. beta	*t*-value	*p*-value
H1	PS → PR	0.177	2.095	0.036	0.174	2.891	0.004
H2	TRS → PR	0.213	3.851	0.000	0.071	2.253	0.024
H3	PU → PR	0.402	5.405	0.000	0.112	2.466	0.014
H4	SC → PR	0.170	3.036	0.002	0.379	7.201	0.000
H5	LAA → PR	0.120	2.096	0.036	0.851	16.541	0.000
H6	TRS × PS → PR	0.149	2.966	0.003	0.030	0.871	0.384
H7	PU × PS → PR	0.179	3.457	0.001	−0.045	1.002	0.316
H8	SC × PS → PR	−0.044	1.069	0.285	−0.214	3.175	0.002
H9	LAA × PS → PR	−0.074	1.040	0.299	−0.108	2.249	0.025

5 Conclusions

5.1 Discussion

This study makes a significant contribution by developing an innovative research framework that investigates the direct effects of purchaser satisfaction, trust, perceived usefulness, switching costs and lack of alternative attractiveness on NFT purchaser retention (H1–H5). Additionally, this study explores the moderating roles of trust and perceived usefulness in the satisfaction–retention link (H6 and H7) and the moderating roles of switching costs and lack of alternative attractiveness (H8 and H9). The choice of Hong Kong as the research location is justified by its status as a global financial hub, advanced technological infrastructure and a notable prevalence of NFT ownership amongst adults. De Best (2022) reported that approximately 5% of the population in Hong Kong owned an NFT as of September 2022. The results of the nine hypotheses are summarised in Table 4. Furthermore, this study examines the heterogeneity between single and multiple NFT platform purchaser groups in relation to the nine hypotheses. The findings reveal that the two strengthening moderators are significant only for single NFT platform purchasers, whilst the two constraining moderators are significant only for multiple NFT platform purchasers. These findings provide valuable insights into the complex dynamics of purchaser retention in NFT platforms, emphasising the need to consider different user segments when formulating and optimising retention strategies.

5.2 Implications of this Study

In academia, this study moves beyond traditional Internet research, whose focus is mainly on factors that encourage adoption and increase. This study argues that once the Internet system has been adopted and used, the next logical inquiry for Internet system research is how to retain users. The findings of this study urge Internet researchers to extend their emphasis on direct and moderating factors leading to effective strategies to retain purchasers. NFT service providers should develop well-designed purchaser satisfaction programs to increase purchaser retention. The findings of this study provide evidence that purchaser satisfaction programs should be accompanied by trust, perceived usefulness, switching costs and lack of alternative attractiveness strategies for early detection and prevention of switching behaviour. NFT service providers can take advantage of structuring retention strategies in different purchaser groups by understanding the moderating role of the strengthening and constraining moderating effects in affecting the relationship between satisfaction and retention.

5.3 Limitations and Suggested Future Research

This study has four main limitations that should be considered: reliance on self-reported measures, a generic definition for multiple platform purchasers, a focus on a specific population and the variance explained in purchaser retention. Firstly, a limitation is that the intention to stay using the main NFT platform may not entirely align with actual retention behaviours in this study, despite previous research predominantly using it as a surrogate measure for actual retention. This limitation emphasises the need for future research to directly investigate and measure the actual retention of purchasers in NFT platforms, rather than relying solely on intentions or self-reported measures. Secondly, a limitation is that the definition of multiple platform purchasers in this study was considerably generic because it did not capture the specific types of NFTs purchased and the intentions behind the purchases. Future research should refine the categorisation of multiple platform purchasers by allowing respondents to provide detailed information on the specific NFT platforms that they used and the types of NFTs that they engaged with. This approach would provide insights into the nuances of purchaser behaviour and preferences within the context of multiple platform usage. Thirdly, a limitation is that the study focused solely on Hong Kong Chinese participants, potentially introducing selection bias and limiting the generalisability of the findings. This study must be replicated with samples from different regions of the world to address this limitation. Researchers can assess the generalizability of the findings and identify potential variations in retention patterns across populations by including participants from diverse cultural backgrounds and geographic locations. Lastly, a limitation is that the five factors examined in this study accounted for only 56.2% of the variance in purchaser retention. This notion suggests that additional factors influencing purchaser retention beyond those investigated in this research. Therefore, future studies should aim to incorporate more direct and moderating factors to increase the amount of variance explained in purchaser retention.

Appendix 1: Measures and Definitions of Constructs

Constructs (source)	Definitions	Code	Items used
Purchaser Retention (PR) (Zeithaml et al. 1996)	We conceptualise purchaser retention as the future propensity of an NFT purchaser who remains loyal to their main NFT platform	PR1	How likely are you to switch to a different main NFT platform within the next 3 months?
		PR2	How likely are you to switch to a different main NFT platform within the next 6 months?
		PR3	How likely are you to switch to a different main NFT platform within the next 12 months?
Purchaser Satisfaction (PS) (Cronin et al. 2000)	We conceptualise purchaser satisfaction as a subjective evaluation of emotions, reflecting the extent to which an NFT purchaser believes that their main NFT platform elicits positive feelings	PS1	I am satisfied with the services provided by my main NFT platform
		PS2	I feel happy with the services provided by my main NFT platform
		PS3	My decision to use my main NFT platform was a wise decision
Trust (TR) (Morgan and Hunt 1994)	We conceptualise trust as the confidence and belief that an NFT purchaser has confidence in their main NFT platform's reliability and integrity	TR1	My main NFT platform can be trusted
		TR2	My main NFT platform can keep its promise
		TR3	My main NFT platform always pays attention to the interests of purchasers
Perceived Usefulness (PU) (Davis 1989)	We conceptualise perceived usefulness as the degree to which an NFT purchaser perceives their main NFT platform to enhance their performance in searching for and purchasing NFTs	PU1	My main NFT platform improves my performance in searching for and buying NFTs
		PU2	My main NFT platform enables me to search for and buy NFTs faster
		PU3	My main NFT platform enhances my effectiveness in searching for and buying NFT
		PU4	My main NFT platform makes it easier for me to search for and purchase NFTs
		PU5	My main NFT platform increases my productivity in searching for and buying NFTs
Switching Costs (SC) (Porter 1980)	We conceptualise switching costs as the perception of the magnitude of the additional costs required for an NFT purchaser to terminate the current relationship with their main NFT platforms and secure an alternative	SC1	It would take me a lot of time to change my main NFT platform
		SC2	It would take me a lot of effort to change my main NFT platform
		SC3	It is risky to change my main NFT platform, as the new platform may not provide good services
		SC4	I would feel frustrated if I stopped using my current main NFT platform
		SC5	Considering everything, the cost of stopping the use of my main NFT platform and switching to a new main NFT platform would be high

Constructs (source)	Definitions	Code	Items used
Lack of Alternative Attractiveness (LAA) (Jones et al. 2000)	We conceptualise lack of alternative attractiveness as purchaser perceptions regarding the extent to which viable competing alternative NFT platforms are not available in the marketplace	LAA1	If I needed to change my main NFT platform, there are very few good platforms to choose from
		LAA2	I would probably be happy with the products and services of another NFT platform (Reverse coded)
		LAA3	Compared to my main NFT platform, there are other platforms with which I would probably be equally or more satisfied (Reverse coded)
		LAA4	Compared to my main NFT platform, there are not many other platforms with which I could be satisfied

References

Alshurideh M, Al Kurdi B, Salloum SA (2019) Examining the main mobile learning system drivers' effects: a mix empirical examination of both the Expectation-Confirmation Model (ECM) and the Technology Acceptance Model (TAM). In: International conference on advanced intelligent systems and informatics, pp 406–417

Amoroso DL, Chen Y (2017) Constructs affecting continuance intention in consumers with mobile financial apps: a dual factor approach. Journal of Information Technology Management 28(3):1–24

Brislin RW (1980) Translation and content analysis of oral and written materials. Methodology, 389–444

Chin WW (1998) The partial least squares approach to structural equation modeling. Mod Methods Bus Res 295(2):295–336

Chuah SHW, Marimuthu M, Kandampully J, Bilgihan A (2017) What drives Gen Y loyalty? Understanding the mediated moderating roles of switching costs and alternative attractiveness in the value-satisfaction-loyalty chain. J Retail Consum Serv 36:124–136

Cronin JJ Jr, Brady MK, Hult GTM (2000) Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. J Retail 76(2):193–218

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 319–340

De Best R (2022) NFT adoption in 26 different countries worldwide 2022. Statista, November 29. https://www-statista-com.hksyu.idm.oclc.org/statistics/1278047/global-nft-adoption-by-country/

Fornell C, Cha J (1994) Partial least squares. In: Bagozzi R (ed) Advanced methods of marketing research. Blackwell, pp 52–87

Fornell C, Larcker DF (1981) Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 18(3):382–388

Foroughi B, Nhan PV, Iranmanesh M, Ghobakhloo M, Nilashi M, Yadegaridehkordi E (2023) Determinants of intention to use autonomous vehicles: findings from PLS-SEM and ANFIS. J Retail Consum Serv 70:103158

Gefen D, Straub DW (2000) The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption. J Assoc Inf Syst 1(1):8

Grønhaug K, Gilly MC (1991) A transaction cost approach to consumer dissatisfaction and complaint actions. J Econ Psychol 12(1):165–183

Hadi NU, Aslam N, Gulzar A (2019) Sustainable service quality and customer loyalty: the role of customer satisfaction and switching costs in the Pakistan cellphone industry. Sustainability 11(8):2408

Hair JF, Ringle CM, Sarstedt M (2011) PLS-SEM: Indeed a silver bullet. J Mark Theory Pract 19(2):139–152

Hair JF, Hult T, Ringle C, Sarstedt M (2014) A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publications, Inc.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2009) Multivariate data analysis: global edition, 7th edn. Pearson Education

Harman HH (1976) Modern factor analysis. University of Chicago Press

Hayward A (2023) NFT Sales in 2022 Nearly matched the 2021 boom, despite market crash. Decrypt, January 5. https://decrypt.co/118438/2022-versus-2021-nft-sales

Jones MA, Mothersbaugh DL, Beatty SE (2000) Switching barriers and repurchase intentions in services. J Retail 76(2):259–274

Kim J, Lee J, Zo H (2018) Toward sustainable Freemium software: the role of user satisfaction and use context. J Electron Commer Res 19(3)

Kline RB (2011) Principles and practice of structural equation modeling. Guilford Press

Koo B, Yu J, Han H (2020) The role of loyalty programs in boosting hotel guest loyalty: impact of switching barriers. Int J Hosp Manag 84:102328

Kotler P, Armstrong G (2020) Principles of marketing, 18th edn. Pearson

Koufaris M (2002) Applying the technology acceptance model and flow theory to online consumer behavior. Inf Syst Res 13(2):205–223

Lee J, Lee J, Feick L (2001) The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France. J Serv Mark 15(1):35–48

Leger C, Politis DN, Romano OP (1992) Bootstrap technology and applications. Technometrics 34(4):378–398

Minta Y (2018) Link between satisfaction and customer loyalty in the insurance industry: moderating effect of trust and commitment. J Mark Manag 6(2):25–33

Morgan RM, Hunt SD (1994) The commitment-trust theory of relationship marketing. J Mark 58(3):20–38

Nguyen DT, Pham VT, Tran DM, Pham DBT (2020) Impact of service quality, customer satisfaction and switching costs on customer loyalty. J Asian Financ, Econ Bus 7(8):395–405

Nunnally JC (1978) Psychometric theory, 2nd edn. McGraw-Hill

Oliver RL (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 17(4):460–469

Patterson PG, Smith T (2003) A cross-cultural study of switching barriers and propensity to stay with service providers. J Retail 79(2):107–120

Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 88(5):879–903

Porter ME (1980) Industry structure and competitive strategy: keys to profitability. Financ Anal J 36(4):30–41

Ranaweera C, Prabhu J (2003) The influence of satisfaction, trust and switching barriers on customer retention in a continuous purchasing setting. Int J Serv Ind Manag 14(4):374–395

Shen H, Yu J, Zhang H, Gou J, Zhang X (2022) How does social support affect the retention willingness of cross-border e-commerce sellers? Front Psychol 12:797035

Statista Research Department (2023) Monthly consumer searches for "NFT" on Google in 206 countries Worldwide 2021–2022. Statista, August 21. https://www-statista-com.hksyu.idm.oclc.org/statistics/1265980/nft-online-search-interest-country/

Sullivan YW, Kim DJ (2018) Assessing the effects of consumers' product evaluations and trust on repurchase intention in e-commerce environments. Int J Inf Manage 39:199–219

Wijaya HR, Astuti ST (2018) The effect of trust and brand image to repurchase intention in online shopping. KnE Soc Sci 915–928

Wilson N (2019) The impact of perceived usefulness and perceived ease-of-use toward repurchase intention in the Indonesian e-commerce industry. Jurnal Manajemen Indonesia 19(3):241–249

Wong CB (2011) The influence of customer satisfaction and switching costs on customer retention: retail internet banking services. Glob Econ Financ J 4(1):1–18

Yi Y (1990) A critical review of consumer satisfaction. Rev Mark 4(1):68–123

Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. J Mark 60(2):31–46

Exploring Determining Factors for SMEs' Access to Alternative Financing Through the Technology-Organization-Environment (TOE) Framework

Shazia Shah and Husam-Aldin Al-Malkawi

Abstract

Small and medium-sized enterprises (SMEs) are vital for a country's economic growth. However, they often struggle with a persistent issue, namely, a funding gap. Alternative finance presents a solution to this problem but remains underutilized due to limited SME understanding and access. This research paper aims to explore the factors that influence small and medium-sized enterprises (SMEs) to access innovative alternative financing models. Through a literature review and conceptual framework, this study identifies the influential factors affecting SMEs' access to alternative financing options. The research adopts the Technology-Organization-Environment (TOE) model as a theoretical framework to understand the factors that influence SMEs' access to alternative financing. The TOE framework explains the process of adopting and implementing technological innovations in organizations and describes how it is influenced by technological, organizational, and environmental contexts. This study not only tests the TOE framework for SMEs but also extends it by adding the individual context, specifically the financial literacy of SME owners/managers, as a determining factor. The comprehensive conceptual model presented in this study offers a holistic perspective on the factors influencing SMEs' access to alternative financing and delineates the dimensions of these factors. The model also provides avenues for future research to test and enhance it. This study makes substantial contributions to theory and practice by exploring emerging business models empowered by disruptive technology. This highlights the crucial role of alternative finance in fostering SME growth and underscores the significance of harnessing disruptive technology in creating innovative business models. The findings of this research are relevant to policymakers, regulators, and SME leaders as they provide valuable insights into fostering SME development and ultimately promoting economic prosperity.

Keywords

Alternative financing · Determining factors · SMEs · Disruptive technologies · TOE framework · Conceptual model

1 Introduction

The Organization for Economic Co-operation and Development emphasizes the paramount importance of small and medium-sized enterprises (SMEs) in advancing the United Nations' Sustainable Development Goals, which encompass job creation, sustainable industrial growth, innovation promotion, and reducing income inequality. Bartolacci et al. (2020) showed the significance of SMEs' financial stability and sustainability in attaining these objectives. However, SMEs face a multitude of challenges, with access to finance being a significant hindrance to their inclusive growth in the global economy. This challenge is substantiated by the growing body of literature (Ang 1991; Baker et al. 2020; Carpenter and Petersen, 2002; Chittenden et al. 1996; Cosh and Hughes 1993; Fazzari et al. 1987; Guariglia et al. 2011; Hutchinson and Xavier 2006; Moscalu et al. 2020).

In the small and medium-sized enterprise (SME) sector, financial constraints often manifest as a scarcity of access to

S. Shah (✉)
Faculty Business, Higher Colleges of Technology (HCT), Abu Dhabi, UAE
e-mail: sshah@hct.ac.ae

H.-A. Al-Malkawi
Faculty of Business and Law, The British University in Dubai, Dubai, UAE

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_15

and utilization of financial resources (Fazzari et al. 1987). The credit rationing model proposed by Stiglitz and Weiss (1981) explains how some borrowers can secure funding while others are denied, due to the dynamics of demand and supply. This results in a substantial financing gap for SMEs, as they may either be denied credit by banks and financial institutions or receive loans that are not aligned with their specific needs, thus hindering viable business projects and profitability. Numerous empirical studies have explored these financing challenges faced by SMEs, with banks serving as the primary source of capital for SMEs in both developed and developing countries (Carey and Flynn 2005; Ono and Uesugi 2009; Vera and Onji 2010; Wu et al. 2008; Zhou 2009). However, SME financing poses risks to banks' liquidity and profitability, particularly in cases of business defaults, which leads to a contraction of credit supply and the emergence of a financing gap for SMEs.

According to Levine (2005), a well-developed financial system can alleviate external finance constraints and foster economic growth. Ghassibe et al. (2019) suggests that alternative finance mechanisms offer individuals and businesses access to affordable financial products and services that meet their needs responsibly and sustainably. These alternative financing options present a new opportunity for small and medium enterprises (SMEs) in emerging markets to bridge the financing gap. The primary objective of this research paper is to develop a unified model that examines the crucial factors that determine SMEs' inclination towards alternative financing options. In the Global Alternative Finance Market Benchmarking report published by Ziegler et al. (2021), it was disclosed that the alternative finance market has experienced substantial growth globally, with a market size of approximately $150 billion in 2015. The report also highlighted the predominant utilization of alternative financing by several countries, including the USA, Canada, the UK, France, Germany, China, Japan, Australia, and New Zealand. Furthermore, the report shed light on the potential for alternative financing in countries such as the Association of Southeast Asian Nations (ASEAN) region, Africa, and the Middle East. However, despite the growth of alternative finance and the importance of SME financing, research gaps still exist, particularly in the emergence of alternative finance channels in the digital economies of developing countries (Bazarbash and Beaton 2020).

Within the context of SME financing, it is imperative to acknowledge the existence of research gaps. Previous studies have explored the challenges and opportunities in this domain, but there is a paucity of investigations into the emergence of alternative finance channels in the digital economies of developing countries. The existing literature primarily focuses on financial practices within the traditional regulated banking and capital market system and often overlooks the practices occurring outside this system within the alternative finance space of developing economies. Furthermore, no formal paper has previously examined the factors that influence SMEs' access to alternative finance in emerging economies, which motivated the selection of our research topic.

Our paper identifies several deficiencies in the existing research landscape. Although prior studies have recognized the significance and challenges of financial inclusion for SMEs, they have not delved into the concrete factors affecting SMEs' willingness to access alternative financing. For example, while Baker et al. (2020) and Maiti (2018) explored financing preferences and practices in Indian SMEs, they did not investigate how such infrastructure could be effectively utilized or the factors that influence SMEs' decisions to seek funds beyond this infrastructure. Additionally, research studies such as those by Eniola and Entebang (2017), Hakim et al. (2018), and Hossain et al. (2023) have investigated the role of managers' financial literacy in business performance. However, there is a significant research gap in examining the relationship between the financial literacy of SME owners/managers and their financial decision-making regarding various sources of finance.

In prior research, Adegboye and Iweriebor (2018) demonstrated the significance of innovation in accessing finance for firms, particularly in the context of banks not accepting innovative products as collateral. However, they did not provide strategies for alternative financing mechanisms. Similarly, Hervé and Schwienbacher (2019) acknowledged the potential of alternative financing for SMEs' innovative projects but lacked empirical evidence to establish the relationship between alternative financing and SMEs' willingness to access it. Moreover, Block et al. (2017) reviewed evidence on innovative entrepreneurship in developed countries but did not address the gap in research regarding how alternative forms of finance can complement traditional finance for small firms in emerging markets. Dara (2018) highlighted the lack of adequate access to finance for innovative projects among many SMEs in emerging markets but did not explore the factors responsible for this lack of access.

Our proposed model takes a comprehensive approach by examining several factors simultaneously—including alternative finance complexity, perceived trust in alternative finance, perceived risk of alternative finance, alternative finance credit terms, alternative finance visibility, social influence, innovation, and financial literacy—to determine SMEs' willingness to access alternative financing, which is a unique characteristic of this research paper. However, prior research has primarily focused on one or two variables to determine the relationship between financing sources and SMEs' willingness to access these resources.

2 Literature Review

The literature review is divided into three distinct areas. In the first area, we provide a clear definition of alternative financing and its distinction from traditional financing methods. In the second area, we investigate the significance of alternative financing in addressing the financing gap for SMEs. Finally, in the third area, we delve into the various factors that influence SMEs' decisions to opt for alternative financing options.

The concept of alternative finance, introduced by the United Nations Development Programme (UNDP), encompasses a diverse range of funding sources that have emerged outside conventional banking systems and traditional capital markets. This umbrella term includes financial solutions such as crowdfunding, peer-to-peer lending, invoice trading, and venture capital. Kauffmann (2005) and Levine (2005) argue that a well-developed financial system can ease external finance constraints while fostering business and economic growth. In line with Levine's (2005) empirical research evaluation, alternative finance mechanisms are crucial ways through which individuals and businesses gain access to affordable financial products and services, meeting their needs responsibly and sustainably (Ghassibe et al. 2019). Anwar et al. (2022) suggest that recognizing opportunities is crucial for the success of ventures in emerging markets, with innovation and growth being key opportunities for SMEs.

Regarding alternative financing mechanisms for SMEs, we conducted a comprehensive literature review to identify various non-bank financing options. The first option is family, friends, and social networks, leveraging social capital (Lee and Persson 2016). Another option is angel investors, affluent individuals willing to invest in early-stage businesses (Sohl 2012). Studies such as Mondal and Shrivastava (2016) and Pedchenko et al. (2018) have explored the potential of angel investors as an alternative source of financial support for small businesses through empirical and statistical analysis. Venture capital, where financial intermediaries raise funds from investors to invest in high-risk young or startup firms, is another alternative financing source (Potter and Porto 2007). Venture capitalists and angel investors can provide value to SMEs by assisting with accurate forecasts, focused business ideas, and clear business models that small businesses may lack (Macht and Weatherston 2014). Trade credit, which is used for ordinary business transactions when a company purchases goods and services on deferred payment, is the primary external financing source for SMEs in most developed and developing countries (Berger and Udell 2006; Demirgüç-Kunt and Maksimovic 2002; Pepur et al. 2020). Crowdfunding, an emerging form of alternative financing, has been discussed across various disciplines (Agrawal et al. 2014; Zhang and Chen 2019).

The mechanism of peer-to-peer lending, an online form of financing connecting lenders and borrowers, has attracted interest from both individual and institutional investors, as well as small businesses. Investors grant loans, and interest is determined by the loan term and the borrower's risk of default. Research into the demand for alternative financing among small and medium-sized enterprises (SMEs) and their access to financial resources is a growing area of interest for researchers. Factors affecting SMEs' financing behavior, including firm-specific, owner-manager, environmental, social, and psychological factors, have been discussed in the entrepreneurship literature. A review of existing literature on SME financing reveals various factors influencing their adoption of alternative funding sources, such as perceptions of complexity, trust requirements for using alternative financing mechanisms, and concerns about business secrecy.

In addition to the aforementioned studies, the works of including Adegboye and Iweriebor (2018), Audretsch et al. (2008), Hochberg et al. (2018), Mbizi et al. (2013), Paul et al. (2017), Zarrouk et al. (2020) have identified the crucial role of innovation in shaping SMEs' willingness to access alternative financing. Additionally, credit terms offered by alternative financing mechanisms (Mambula 2002; Rahman et al. 2017); visibility and awareness of available alternative financing options (Hester 2011; Venkatesh et al. 2003), and social behavior (Colombo et al. 2015; Luu and Nguyen 2013; Mollick 2014; Rauwerda and Graaf 2021) have been found to impact SMEs' borrowing behavior. Furthermore, financial literacy of SME owners/managers (Hussain et al. 2018; Wise 2013; Carbo-Valverde et al. 2016; Eniola and Entebang 2017; Fatoki 2014; Lusardi and Mitchell 2014; Okello Candiya Bongomin et al. 2017; Adomako et al. 2016) is a crucial skill for making informed financial decisions. It is important to consider the aforementioned factors when exploring alternative financing options.

3 Theoretical Framework: TOE (Technological, Organizational, and Environmental) Framework

Several theories have been developed to understand technology adoption dynamics at the firm level, with the Theory of Planned Behavior (TPB) being a widely used framework in this field (Davis 1989). However, TPB is limited in explaining technology adoption by firms since it relies on individual-level data. In contrast, the Technology-Organization-Environment (TOE) framework, proposed

by Tornatzky et al. (1990), offers a more comprehensive understanding of technology adoption at the firm level. The TOE framework suggests that technology adoption by firms is influenced by three key contexts: technological, organizational, and environmental. The technological context involves technology characteristics, while the organizational context includes aspects such as organizational structure, innovation-enabling processes, and resource availability. The environmental context considers factors like market structure and regulations. According to Hue (2019), these three factors within the TOE model determine a firm's willingness and process of adopting and accepting innovative models. Empirical studies consistently find that these three contexts are major determinants of firms' motivation to adopt new technology (Chiu et al. 2017; Chau and Tam 1997; Lippert and Govindrajulu 2006; Wan Muhammad 2016; Zhu et al. 2004).

This research paper aims to contribute to the existing literature by empirically examining the factors affecting small and medium-sized enterprises (SMEs)' access to alternative financing, using the Technology-Organization-Environment (TOE) framework as a theoretical foundation. Furthermore, an extension to the TOE framework is proposed, incorporating the individual context of owner/manager financial literacy. It is argued that financially literate managers are more likely to choose appropriate financing products and resources.

Previous research by Adomako et al. (2016) and Hussain et al. (2018) has explored the effects of financial literacy on entrepreneurs' financial decisions and access to resources. In this paper, several key factors were identified from a comprehensive literature review and subsequently mapped to the TOE framework. The validity of this mapping is demonstrated in Table 1.

4 Conceptual Model

Figure 1 depicts the original TOE framework model, while Fig. 2 presents the proposed conceptual model. This research paper, inspired by the TOE framework (Technological, Organizational, and Environmental), integrates eight factors identified through an extensive literature review in the previous section. These factors are aligned with the foundational elements of the original TOE model, with the addition of a novel extension related to individual context. This individual context construct measures the financial literacy of SME owners and managers, making it a unique contribution to the theoretical framework.

4.1 Technological Context

In the context of alternative financing, essential technological elements and innovations include blockchain, peer-to-peer lending platforms, crowdfunding platforms, and fintech. It's crucial to understand the perceived complexity, trust, and risk associated with these alternative financing models, as these factors significantly influence firms' willingness to utilize these resources. According to Tornatzky et al. (1990), technology characteristics play a critical role in facilitating firms' adoption of new models. In the case of alternative financing models, technological innovation can significantly influence a firm's intention to use these resources based on their perceived advantages. Awa et al. (2017) have offered a comprehensive perspective on the technological context as a multifaceted determinant of adoption, which includes factors such as availability, perceived advantages, compatibility, complexity, pilot testing, and visibility. Within this framework, complexity, perceived trust, and perceived risk emerge as critical factors that influence a small and medium-sized enterprise's (SME's) access to alternative financing.

In the proposed conceptual model, complexity is defined as the extent to which SMEs perceive alternative financing options as understandable and manageable. Perceived complexity of these models significantly impacts SMEs' financing decisions. Their confidence in dealing with complex financial products and services, coupled with their ability to comprehend and navigate alternative finance mechanisms, plays a substantial role in their choice between innovative financing models and conventional options. The level of trust that SMEs place in alternative financing instruments and their perception of associated risks are crucial in shaping their borrowing preferences.

Understanding these perceptions is critical for comprehending why firms choose certain financing models over others and how this impacts their financial strategies and business growth. Small and medium-sized enterprises (SMEs) place trust in the reliability and security of various alternative financing instruments. This concept of trust, as articulated by Islam and Khan (2021), encompasses beliefs in the reliability, trustworthiness, integrity, stability, and technological appropriateness of finance sources. This trust positively influences users' demand for and participation in these financing alternatives.

Mitchell (1992) argues that risk remains a persistent factor throughout all stages of purchasing decisions and significantly influences consumer behavior. Building on this premise, Islam and Khan (2021) assert that major risks

Table 1 Mapping the TOE framework (Tornatzky et al. 1990) with the proposed framework

Dimension	TOE framework, Tornatzky et al. (1990)	Proposed TOE framework	References
Technological context	Quality, functionality, complexity, compatibility, and user friendliness with technology	Constructs mapped: – Complexity – Perceived trust – Perceived risk	Awa et al. (2017), Venkatesh et al. (2003), Yeh (2020), Fu et al. (2022), Thong (1999)
Organizational context	Organization's size, structure, culture, and available resources	Construct mapped: – Innovation	Adegboye and Iweriebor (2018), Calantone et al. (2002)
Environmental context	Market conditions, regulatory demands, societal, and cultural norms	Constructs mapped: – Credit terms – Visibility – Social influence	Awa et al. (2017)
Individual context		Individual context (added context to TOE framework proposed in conceptual model of the research paper) SME owner/manager financial literacy	Awa et al. (2017)

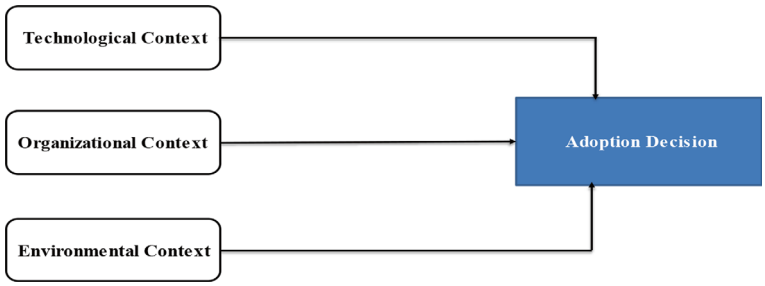


Fig. 1 TOE framework with traditional attributes. *Source* Tornatzky et al. (1990, p. 76)

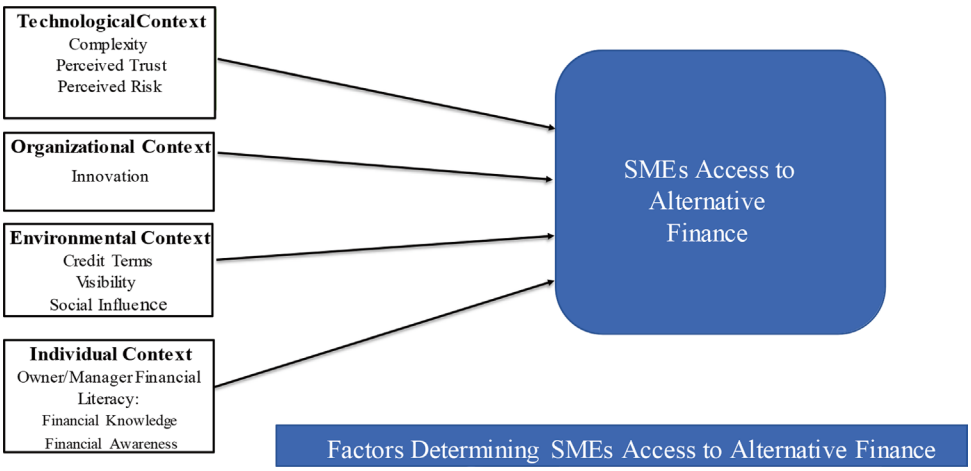


Fig. 2 Conceptual model

associated with alternative finance sources include the leakage of business privacy, performance risk, time consumption risk, psychological pressure, imitation risk concerning new business ideas, and the risk of disclosing sensitive business information.

Researching perceptions of complexity, trust, and risk in the realm of alternative finance is essential for understanding firms' willingness to embrace these resources. These factors collectively shape SMEs' decisions regarding the adoption of innovative financing models, ultimately influencing the trajectory of their financial strategies and business growth.

4.2 Organizational Context

This paper utilizes the innovation construct within the Technology-Organization-Environment (TOE) model to assess the intentions of small and medium-sized enterprises (SMEs) toward alternative financing options. The integration of emerging technology and innovation can enhance SMEs' access to finance, particularly when owners have financial literacy and appropriate regulatory support exists within the infrastructure. Financial innovation, as noted by Beck and Demirguc-Kunt (2006), can aid SMEs in accessing external financing alternatives, such as leasing and factoring, in cases where bank loans are unavailable. The paper examines the relationship between innovativeness and the intention to access alternative financing. A firm's innovative capabilities, encompassing its ability to create, adapt, and apply innovative solutions to address challenges and opportunities, play a crucial role in determining its readiness and inclination to utilize alternative financing options. Firms with strong innovative capabilities tend to be more inclined to explore and utilize alternative financing options. This paper investigates the innovative capabilities of firms to determine their receptiveness to non-conventional funding sources, which can reveal their intention and suitability for accessing alternative finance. Researching a firm's innovative capabilities is essential in determining its intention and suitability for accessing alternative finance, as these capabilities influence the firm's mindset, risk tolerance, value proposition, strategic fit, and long-term viability, all critical considerations when exploring non-traditional financing avenues.

4.3 Environmental Context

Based on the empirical findings of Awa et al. (2017), this research delves into three critical factors that influence the intention of small and medium enterprises (SMEs) to access alternative finance: the terms of credit associated with alternative financing, the visibility of funding sources outside

traditional channels, and the impact of social influences on managers' decision-making processes.

The terms and conditions of credit attached to alternative financing options have a significant impact on the borrowing decisions of SMEs. Favorable credit terms, such as those offered by alternative financing options, can be a compelling incentive for SMEs to explore these options. Conversely, stringent conditions may discourage SMEs from considering alternative financing. Bond and Meghir (1994) underscore the crucial role of credit terms in determining SMEs' access to financial resources. When credit terms are favorable, SMEs are more likely to embrace the opportunity to secure the funds they need.

The visibility of alternative financing sources is also a crucial factor in influencing entrepreneurs' decisions. Hester (2011) notes that the degree of visibility of alternative financing options can significantly impact entrepreneurs and their willingness to engage with these sources. Awareness and comprehension of alternative finance within the business environment are essential factors that shape SMEs' intentions to pursue or overlook these funding opportunities. This notion aligns with the theory of network externalities (Katz and Shapiro 1985), which posits that the value of using certain technologies or financial instruments increases as more individuals and businesses become aware of and adopt them. In the context of alternative finance, widespread adoption among numerous SMEs can enhance its perceived value and exert pressure on others to embrace these funding sources.

Finally, the impact of social influences on managers' decision-making processes cannot be overlooked. The influence of family, friends, and other stakeholders can be significant in shaping SME managers' perceptions of alternative finance and their willingness to pursue these options. According to Shao and Sun (2021), social networks play a critical role in the process of entrepreneurs' decision-making, including their selection of financing sources. Understanding the importance of social networks in the decision-making processes of SME managers is essential in developing effective strategies to promote alternative finance options.

It is imperative to note that the influence of social networks on financial decision-making among small and medium-sized enterprise (SME) owner-managers is a significant factor. Managers commonly rely on the guidance and opinions of close associates and contacts when making financial decisions, and these relationships play a crucial role in the decision-making process, particularly concerning financing and project completion. In the context of alternative finance, social influence holds substantial sway over SMEs' financial choices. It helps bridge information gaps and reduces uncertainties related to potential sources of finance, thereby emerging as a crucial determinant for SMEs and significantly shaping their financial decisions.

Research into the credit terms associated with alternative finance, the visibility of funding sources, and the impact of social influences on managers is essential for understanding and predicting SMEs' intentions to access nontraditional financing. These factors collectively influence the attractiveness of alternative financing options, the level of awareness and comprehension within the business ecosystem, and the interpersonal dynamics that guide financial decision-making.

4.4 Individual Context

The current paper aims to contribute to the technology-organization-environment (TOE) model by incorporating the individual context, which has a significant influence on the adoption of new technologies by small and medium enterprises (SMEs). Prior research (Erdogan 2018; Wong et al. 2018) has emphasized that owners/managers with a comprehensive understanding of alternative financing mechanisms are more likely to effectively approach and comprehend them. As alternative financing mechanisms play a vital role for SMEs, it is essential to examine the role of financial literacy and awareness in accessing finances.

This paper seeks to address a gap in the existing literature by proposing the measurement of financial literacy in two dimensions: financial knowledge and financial awareness. Financial knowledge refers to actual proficiency in financial matters, while financial awareness delves into a deeper understanding of how financial knowledge can be practically applied within the context of the business environment.

Both financial knowledge and awareness are of utmost importance to SME owner-managers as they strive to make informed financial decisions and navigate the intricacies of managing their businesses' finances. This two-dimensional approach aims to enhance our understanding of financial literacy among SME owner-managers by providing valuable insights into their capacity to effectively access and navigate alternative financing options.

This paper emphasizes the importance of examining the connection between owner-manager financial literacy and their inclination to seek out financing sources. Through an in-depth examination of financial knowledge and understanding, this research aims to illustrate the paramount significance of financial literacy in small and medium-sized enterprises' capacity to leverage alternative financing mechanisms to foster growth and development.

5 Conclusion

Small and medium-sized enterprises (SMEs) face a range of challenges, particularly in accessing finance, which is a significant obstacle to their inclusive growth in the global economy. The lack of access to appropriate credit from traditional financial institutions results in a financing gap, which poses risks to both SMEs and traditional banks. To address these external finance constraints and foster economic growth, this research paper examines the emergence of alternative finance mechanisms. These mechanisms provide new opportunities for SMEs, particularly in emerging markets, to bridge the financing gap. The primary objective of this research paper is to develop a unified model that examines the crucial factors that determine small and medium enterprises (SMEs) inclination towards alternative financing options.

The literature review in this research paper not only identifies various forms of alternative finance but also uncovers the complex factors affecting SMEs' decisions to opt for alternative financing options. This comprehensive analysis reveals research gaps and unexplored areas within the current literature, paving the way for further exploration and inquiry. Additionally, this research paper extends the well-established TOE (Technological, Organizational, and Environmental) framework by adding an individual context—the financial literacy of SME owners/managers. This innovation has the potential to enhance our understanding of how financial literacy influences SMEs' willingness to access alternative finance.

From a professional perspective, the utilization of the conceptual model can provide valuable insights for small and medium-sized enterprises (SMEs), non-bank financing companies, regulators, and policymakers. The research findings derived from the application of the conceptual model can offer SMEs a more comprehensive understanding of the factors that influence their borrowing decisions, enabling non-bank financing companies to develop more effective strategies for meeting the financing needs of SMEs. Additionally, the knowledge gained from the relationship between the independent and dependent variables can be leveraged by regulators and policymakers in the development of frameworks that promote alternative financing, ultimately contributing to the growth and stability of the economy.

This research not only underscores the importance of SMEs and the challenges they face but also lays the foundation for a deeper understanding of the role of alternative financing mechanisms in addressing these challenges. It is anticipated that this research will stimulate further exploration in the field, ultimately leading to the strengthening and diversification of financial markets, with a specific focus on revitalizing the SME sector. Recognizing the importance of these factors can assist policymakers, financial institutions, and entrepreneurs in developing strategies to facilitate SMEs' access to alternative financial resources and promote their growth. It is imperative to address the need for further research in these areas to support the development of effective policies and strategies that can help SMEs access alternative financial resources and promote their growth. The next step in our research shall involve the gathering of data, followed by its analysis and the validation of our conceptual model.

6 Recommendations for Future Research

In light of our previous discussions, we offer the following recommendations for future research to further explore and expand upon the identified areas of interest. These recommendations are intended to provide a more comprehensive understanding of the relationship between the individual context, financial literacy, and SMEs' financing intentions, as well as to identify potential policy implications.

(i) Cross-sectional correlational surveys: This method of research, which involves collecting data from a large, diverse sample simultaneously, is useful for hypothesis testing and allows researchers to measure the strength of relationships between variables.
(ii) Mediation and Moderation Analyses: It is recommended to examine whether technological, environmental, and organizational variables mediate or moderate the relationship between the individual context, financial literacy, and SMEs' financing intentions. Investigating these interactions can provide deeper insight into the mechanisms driving SME financing decisions.
(iii) Cross-Country Comparative Studies: To test the universality of the model and identify country-specific factors, studies may be conducted in different economic contexts. Comparative research could strengthen these results and provide valuable insights into how SME financing varies across different regions.
(iv) Mixed-Methods Approach: Combining quantitative and qualitative research, plus case studies, can provide a more comprehensive understanding of the relationship between variables. Qualitative insights can explain and enhance the quantitative results, while case studies can offer more in-depth, contextual information.
(v) Big data and advanced analytics can be used to analyze large datasets for SME financing decisions. Machine learning can identify patterns and predict financing intentions from a data-driven viewpoint, which may provide valuable insights into SME financing patterns and trends.

(vi) Policy Implications: It is recommended to explore how policymakers can use these insights to foster SME financing. Investigating whether policy interventions can address the challenges identified in the research may provide useful information for policymakers, financial institutions, and SME owners and managers. This research can help SMEs access financing and grow, benefiting the SME sector and economy in general.

References

Adegboye AC, Iweriebor S (2018) Does access to finance enhance SME innovation and productivity in Nigeria? Evidence from the World Bank Enterprise Survey. Afr Dev Rev 30(4):449–461

Adomako S, Danso A, Ofori Damoah J (2016) The moderating influence of financial literacy on the relationship between access to finance and firm growth in Ghana. Ventur Cap 18(1):43–61

Agrawal A, Catalini C, Goldfarb A (2014) Some simple economics of crowdfunding. Innov Policy Econ 14(1):63–97

Ang JS (1991) Small business uniqueness and the theory of financial management. J Small Bus Financ 1(1):1–13

Anwar M, Clauss T, Issah WB (2022) Entrepreneurial orientation and new venture performance in emerging markets: the mediating role of opportunity recognition. RMS 16(3):769–796

Audretsch DB, Bönte W, Keilbach M (2008) Entrepreneurship capital and its impact on knowledge diffusion and economic performance. J Bus Ventur 23(6):687–698

Awa HO, Ukoha O, Igwe SR (2017) Revisiting technology-organization-environment (TOE) theory for enriched applicability. Bottom Line 30(01):2–22

Baker HK, Kumar S, Rao P (2020) Financing preferences and practices of Indian SMEs. Glob Financ J 43:100388

Bartolacci F, Caputo A, Soverchia M (2020) Sustainability and financial performance of small and medium-sized enterprises: a bibliometric and systematic literature review. Bus Strateg Environ 29(3):1297–1309

Bazarbash M, Beaton K (2020) Filling the gap: Digital credit and financial inclusion. IMF Working Papers, 2020(150)

Beck T, Demirguc-Kunt A (2006) Small and medium-size enterprises: access to finance as a growth constraint. J Bank Finance 30(11):2931–2943

Berger AN, Udell GF (2006) A more complete conceptual framework for SME finance. J Bank Financ 30(11):2945–2966

Block JH, Fisch CO, Van Praag M (2017) The Schumpeterian entrepreneur: a review of the empirical evidence on the antecedents, behaviour and consequences of innovative entrepreneurship. Ind Innov 24(1):61–95

Bond S, Meghir C (1994) Financial constraints and company investment. Fisc Stud 15(2):1–18

Calantone RJ, Cavusgil ST, Zhao Y (2002) Learning orientation, firm innovation capability, and firm performance. Ind Mark Manage 31(6):515–524

Carbo-Valverde S, Rodriguez-Fernandez F, Udell GF (2016) Trade credit, the financial crisis, and SME access to finance. J Money, Credit, Bank 48(1):113–143

Carey D, Flynn A (2005) Is bank finances the achilles of Irish SMEs? J Eur Indus Train 29(9)

Carpenter RE, Petersen BC (2002) Capital market imperfections, high-tech investment, and new equity financing. Econ J 112(477):F54–F72

Chau PY, Tam KY (1997) Factors affecting the adoption of open systems: an exploratory paper. MIS Quart 1–24

Chittenden F, Hall G, Hutchinson P (1996) Small firm growth, access to capital markets, and financial structure: review of issues and an empirical investigation. Small Bus Econ 8:59–67

Chiu CY, Chen S, Chen CL (2017) An integrated perspective of TOE framework and innovation diffusion in broadband mobile applications adoption by enterprises. Int J Manage Econ Soc Sci (IJMESS) 6(1):14–39

Colombo MG, Franzoni C, Rossi-Lamastra C (2015) Internal social capital and the attraction of early contributions in crowdfunding. Entrep Theory Pract 39(1):75–100

Cosh AD, Hughes A (1993) Size, financial structure, and profitability: UK companies in the 1980s. Department of Applied Economics, Small Business Research Centre, University of Cambridge

Dara NR (2018) The global digital financial services: a critical review to achieve for digital economy in emerging markets. Int Res J Hum Resourc Soc Sci 5(1):141–163

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart 319–340

Demirgüç-Kunt A, Maksimovic V (2002) Funding growth in bank-based and market-based financial systems: evidence from firm-level data. J Financ Econ 65(3):337–363

Erdogan AI (2018) Factors affecting SME access to bank financing: an interview study with Turkish bankers. Small Enterp Res 25(1):23–35

Eniola AA, Entebang H (2017) SME managers and financial literacy. Glob Bus Rev 18(3):559–576

Fatoki O (2014) The financial literacy of micro entrepreneurs in South Africa. J Soc Sci 40(2):151–158

Fazzari S, Hubbard RG, Petersen BC (1987) Financing constraints and corporate investment (No. w2387). National Bureau of Economic Research

Fu HP, Chang TS, Wang CN, Hsu HP, Liu CH, Yeh CY (2022) Critical factors affecting the introduction of mobile payment tools by micro retailers. Technol Forecast Soc Chang 175:121319

Ghassibe M, Appendino M, Mahmoudi SE (2019) SME financial inclusion for sustained growth in the Middle East and Central Asia. International Monetary Fund

Guariglia A, Liu X, Song L (2011) Internal finance and growth: micro econometric evidence on Chinese firms. J Dev Econ 96(1):79–94

Hakim MS, Oktavianti V, Gunarta IK (2018) Determining factors that contribute to financial literacy for small and medium enterprises. In: IOP conference series: materials science and engineering, vol 337, no 1, p 012064. IOP Publishing

Hervé F, Schwienbacher A (2019) Crowdfunding and innovation. In: Contemporary topics in finance: a collection of literature surveys, pp 331–349

Hester AJ (2011) A comparative analysis of the usage and infusion of wiki and non-wiki-based knowledge management systems. Inf Technol Manage 12:335–355

Hochberg YV, Serrano CJ, Ziedonis RH (2018) Patent collateral, investor commitment, and the market for venture lending. J Financ Econ 130(1):74–94

Hossain MM, Ibrahim Y, Uddin MM (2023) Finance, financial literacy and small firm financial growth in Bangladesh: the effectiveness of government support. J Small Bus Entrep 35(3):336–361

Hue TT (2019) The determinants of innovation in Vietnamese manufacturing firms: an empirical analysis using a technology–organization–environment framework. Eurasian Bus Rev 9(3):247–267

Hussain J, Salia S, Karim A (2018) Is knowledge that powerful? Financial literacy and access to finance. J Small Bus Enterp Dev 25(6)

Hutchinson J, Xavier A (2006) Comparing the impact of credit constraints on the growth of SMEs in a transition country with an established market economy. Small Bus Econ 27:169–179
Islam MT, Khan MTA (2021) Factors influencing the adoption of crowdfunding in Bangladesh: a paper of start-up entrepreneurs. Inf Dev 37(1):72–89
Kauffmann C (2005) Financing SMEs in Africa. OECD
Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. Am Econ Rev 75(3):424–440
Lee S, Persson P (2016) Financing from family and friends. The Review of Financial Studies 29(9):2341–2386
Levine R (2005) Finance and growth: theory and evidence. Handbook of economic growth, vol 1, pp 865–934
Lippert SK, Govindarajulu C (2006) Technological, organizational, and environmental antecedents to web services adoption. Commun IIMA 6(1):14
Lusardi A, Mitchell OS (2014) The economic importance of financial literacy: theory and evidence. J Econ Literat 52(1):5–44
Luu N, Nguyen N (2013) Determinants of financing pattern and access to formal-informal credit: the case of small and medium-sized enterprises in Vietnam. J Manag Res 5(2):240–259
Macht SA, Weatherston J (2014) The benefits of online crowdfunding for fund-seeking business ventures. Strateg Chang 23(1–2):1–14
Maiti M (2018) Scope for alternative avenues to promote financial access to MSMEs in developing nations: evidence from India. Int J Law Manag 60(5)
Mambula C (2002) Perceptions of SME growth constraints in Nigeria. J Small Bus Manage 40(1):58–65
Mbizi R, Hove L, Thondhlana A, Kakava N (2013) Innovation in SMEs: a review of its role in organizational performance and SMEs' operational sustainability. Interdisc J Contemp Res Bus 4(11):370–389
Mitchell VW (1992) Understanding consumers' behavior: can perceived risk theory help? Manage Decis 30(3)
Mollick E (2014) The dynamics of crowdfunding: an exploratory paper. J Bus Ventur 29(1):1–16
Mondal D, Shrivastava A (2016) Angel funds: the new type of alternative investment fund in India. IUP Law Rev 6(4)
Moscalu M, Girardone C, Calabrese R (2020) SMEs' growth under financing constraints and banking markets integration in the Euro area. J Small Bus Manage 58(4):707–746
Okello Candiya Bongomin G, Mpeera Ntayi J, Munene JC, Akol Malinga C (2017) The relationship between access to finance and growth of SMEs in developing economies: financial literacy as a moderator. Rev Int Bus Strat 27(4):520–538
Ono A, Uesugi I (2009) Role of collateral and personal guarantees in relationship lending: evidence from Japan's SME loan market. J Money, Credit, Bank 41(5):935–960
Paul J, Parthasarathy S, Gupta P (2017) Exporting challenges of SMEs: a review and future research agenda. J World Bus 52(3):327–342
Pedchenko N, Strilec V, Kolisnyk GM, Dykha MV, Frolov S (2018) Business angels as an alternative to financial support at the early stages of small businesses' life cycle
Pepur S, Kovač D, Ćurak M (2020) Factors behind trade credit financing of SMEs in Croatia. Zbornik Veleučilišta u Rijeci 8(1):59–76
Potter J, Proto A (2007) Promoting entrepreneurship in Southeast Europe: policies and tools. OECD Pap 6(12):1
Rahman A, Belas J, Kliestik T, Tyll L (2017) Collateral requirements for SME loans: empirical evidence from the Visegrad countries. J Bus Econ Manag 18(4):650–675
Rauwerda K, De Graaf FJ (2021) Heuristics in financial decision-making: the selection of SME financing by advisers in an increasingly diverse market. Manag Decis 59(7):1728–1749
Shao Y, Sun L (2021) Entrepreneurs' social capital and venture capital financing. J Bus Res 136:499–512
Sohl J (2012) The changing nature of the angel market. Handb Res Venture Capital 2:17–41
Stiglitz JE, Weiss A (1981) Credit rationing in markets with imperfect information. Am Econ Rev 71(3):393–410
Thong JY (1999) An integrated model of information systems adoption in small businesses. J Manag Inf Syst 15(4):187–214
Tornatzky LG, Fleischer M, Chakrabarti AK (1990) Processes of technological innovation. Lexington Books
Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Quart 425–478
Vera D, Onji K (2010) Changes in the banking system and small business lending. Small Bus Econ 34(3):293–308
Wan Muhammad WI (2016) Application of TOE framework in examining the factors influencing pre-and post-adoption of CAS in Malaysian SMEs. Int J Inform Technol Bus Manag 49(1):26–37
Wise S (2013) The impact of financial literacy on new venture survival. Int J Bus Manage 8(23):30
Wong A, Holmes S, Schaper MT (2018) How do small business owners actually make their financial decisions? Understanding SME financial behavior using a case-based approach. Small Enterp Res 25(1):36–51
Wu J, Song J, Zeng C (2008) An empirical evidence of small business financing in China. Manag Res News 31(12)
Yeh HP (2020) Factors in the ecosystem of mobile payment affecting its use: from the customers' perspective in Taiwan. J Theor Appl Electron Commer Res 15(1):13–29
Zarrouk H, Sherif M, Galloway L, el Ghak T (2020) Entrepreneurial orientation, access to financial resources, and SMEs' business performance: the case of the United Arab Emirates. J Asian Financ Econ Bus 7(12):465–474
Zhang H, Chen W (2019) Backer motivation in crowdfunding new product ideas: Is it about you or is it about me? J Prod Innov Manag 36(2):241–262
Zhou W (2009) Bank financing in china's private sector: the payoffs of political capital. World Dev 37(4):787–799
Zhu K, Kraemer KL, Dedrick J (2004) Information technology payoff in e-business environments: an international perspective on value creation of e-business in the financial services industry. J Manag Inf Syst 21(1):17–54
Ziegler T, Shneor R, Wenzlaff K, Wang B, Kim J, Paes FFDC, ... Adams N (2021) The global alternative finance market benchmarking report. Available at SSRN 3771509

Into the Secret Garden or a Dark Pool? An Exploration of Whether DeFi Gardens/ Pools Provide a Democratic Alternative to Principal-Agent Investment Products

Mark Le Page

Abstract

This paper rejects the hypothesis that investment gardens, pools or sets present an imminent danger to the public. Gardens are eBay-style websites that allow anybody to create an investment product, and for it to be run on democratic principles rather than a principal-agent relationship. Gardens contain two inherent weaknesses: (1) no barriers to entry for the creators; and (2) retail customers being exposed to a highly niche and confusing area. This paper sources the limited research noting the opportunities for abuse that producer-managers of complex financial products can exploit against the unsuspecting public. This is considered against the complexity of DeFi products. This paper is a specific consideration of the underlying dialectic in the DeFi debate. How dare access to the riches of the finance sector be denied to anyone? Or is there an indisputable obligation for regulators to restrict DeFi access to protect vulnerable investors? Eight Gardens are regressed against the S&P 500, and Bitcoin and Ethereum prices. Additionally, the research compares the Sharpe Ratio of Gardens to 55 traditional investment fund products investing in DeFi. This leads to the paper proposing an extension of adverse selection from an agency-theoretic approach. The propensity in a Garden (and its circumstances) for adverse selection should determine whether retail investors are allowed to access a specific Garden.

Keywords

DeFi · Game theory · Adverse selection · Investment gardens · Finance democratisation

1 Introduction

"What if the people that [*sic*] know the opportunities share them and get rewarded? You can choose to engage, **join the conversation and discuss opportunities**. Better yet, you can lead and suggest investment ideas to the community. In a group or community…

1. Other participants in the same group can **endorse or downvote these ideas** to ensure that only the best ideas rise to the top.
2. The best ideas will activate the pooled capital. The **rewards for good ideas will be shared** between the ideators, curators, and investors." [bold as per article] Recuero (2021)

This statement on Decentralised Finance ("DeFi") gardens, pools, or sets (hitherto referred to as "Gardens"), reflects the primary dialectic, currently being played out between the cryptoasset industry and financial regulators in Operation Choke 2.0 (cited in SITIE conference proceedings by Cutler et al. 2023). The cryptoasset industry argues for greater access to the finance sector so the public is not denied the financial returns available to the elites: regulators see unacceptable risks being handed down to the public.

Thus, DeFi becomes not only for the professional investors but also "for the people", democratising knowledge distribution and therefore, in theory, wealth creation (Dawson and Allen, n/d). Where is this knowledge imparted? On website platforms. How is it fair? The law courts are replaced by open-source code (Robertson and Robinson 2021) (Table 1).

Yet it is possible that regulators have not yet examined or do not even know about DeFi gardens, pools, or sets. Certainly, there is little academic research to alert them.

M. Le Page (✉)
ESCP Business School, Paris, France
e-mail: mark.le_page@edu.escp.eu

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_16

Table 1 Nomenclature

Term	Definition
Absolute return	Either a positive return or a loss that has been "mitigate[d]…in the event of a market correction" (McCann 2014)
Actively managed index	Per Anonymous (2022) an investment manager can depart from the weightings given an index to improve investment performance
Collective investment fund	Defined by the EU as "investment products created with the sole purpose of gathering investors' capital and investing that capital collectively through a portfolio of financial instruments such as stocks, bonds and other securities. Investment funds play a crucial role in facilitating the accumulation of personal savings…for retirement" (European Commission, n.d.)
Cryptoasset	Per Financial Action Task Force: "a digital representation of value that can be digitally traded, or transferred, and can be used for payment or investment purposes" (Financial Action Task Force 2023)
Decentralised Finance ("DeFi")	Disintermediated financial products (per Belaşcu et al. 2022) using blockchain programmability to perform the service
DeFi gardens, pools, or sets	Internet platforms which seek to remove an intermediary in the financial services industry by either utilising mathematics and programs (smart contracts) within, say, Ethereum (Belaşcu et al. 2022) and/or providing a community where knowledge is imparted to the uninitiated. Henceforth called "Gardens" in this paper
Index tracker funds	Will typically try and track the performance of a financial index—mainly by investing in what comprises it (Jenkins, n.d.)
Minimum investment	Per Tretina (2023) most hedge funds have a minimum investment of USD 100,000 but can extend to at least USD 2 million
Professional investor	Using the United Kingdom Financial Conduct Authority's (2023) Conduct of Business Services Rule 3.5.3—elective professional client: a professional investor who has worked at least 1 year in a position that requires them to work the trades and has traded on their own account for at least 10 significant transactions each quarter over the previous 10 quarters
The public	The retail sector. This paper modifies the United Kingdom Financial Conduct Authority's (2023) Conduct of Business Services Rule 3.4.1 description: it is not a professional client
Regulation D of SEC Rule 506	Permits an accredited investor in an unregulated fund to be someone who can invest sufficient to be clearly of sizeable wealth, which is broadly USD 200,000 *per annum* annual income and net worth (excluding primary residence) above USD 1 million (U.S. Securities and Exchange Commission 2017)
Yield farming	The depositing of cryptoassets into a Decentralised Application, with the expectation of an interest return (Anonymous, n.d.)

2 Problem Statement

The objectives of this research are twofold.

Part I: discover whether or not Gardens have a clear investment strategy and how that performance compares—via ordinary least squares regression—to the S&P500, bitcoin, and Ethereum; and

Part II: show that traditional investment fund structures, invested in the DeFi asset class, perform better than Gardens.

3 Literature Review

The introduction referenced that DeFi has been an increasing phenomenon, particularly since 2020 (Arslanian 2022). Recuero, the author of the above quotation and founder of Babylon Finance, claimed there were US$ 40 billion assets under management by February 2021; an enormous growth, if true, from 2020 of US$ 14 billion (Schmidt 2021). That noted, academics such as Belaşcu et al. (2022) more prudently estimate USD 14.4 billion in February 2021.

The recent advent of DeFi (see Kim 2021; Arslanian 2022) leaves only nascent literature from which to analyse Gardens, and this largely from the wider subject matter of DeFi. Belaşcu et al. (2022) categorise DeFi asset management as composable. Schär (2021) describes composability as enabling a DeFi product to be built from individual pieces, like Lego. He describes layers of sophistication of DeFi asset management, with Enzyme and TokenSets given as two examples of asset management Layer 4 aggregators. This research uses these two aggregators because they contain Gardens.

Belaşcu et al. (2022) agree with Schär (2021) that DeFi asset management protocols should algorithmically allocate capital to the highest returning yield, although the former is

Table 2 Adverse selection relationships per Pouryousefi and Frooman (2019)

	Principal can judge information	Principal cannot judge information
Principal can observe information	Symmetric information	*Homoious hemin*
Principal cannot observe information	Blind-spot	Wardship

clearer in insisting that the protocol's risk appetite should also be defined as a choice for investors. Both similarly agree on the risks, particularly vis-à-vis macroeconomics, i.e. procyclicality, contagion, and cyber security. Belaşcu et al. however go further, perceiving risks in asset/liability mismatches and the ensuing instability in product that will happen thereon. Thus, Part II of this research looks not only at return but also at risk—accounted for in the Sharpe ratio.

The challenges to an investor's decision-making are not limited to deriving a risk-adjusted return. Pouryousefi and Frooman (2019) contend for the principal-agent relationship being different from previous theory. They agree that the agent has the greater knowledge, but not that the principal writes the contract. This is the opposite of Recuero's vision, quoted in the introduction, that Gardens would invest based on democratic decisions. All strategies would be voted on with perfect information shared. This research needs to consider whether or not, in Gardens, agents write the strategy rather than the democratic caucus of investors. If agents do, then the principal-agent relationship has not been overcome.

Pouryousefi and Frooman (2019) conclude that the agent's writing of contract leads to adverse selection problems, which may be defined as incorrect decision-making based on an informed agent but uninformed principal. They continue by subdividing the decision-making ability of the principal on a 2×2 matrix (Table 2), distinguishing between the principal's access to observe and their ability to judge, or evaluate, information. Information asymmetry is abolished when the principal can see full information and understand it.

When the principal has good financial and technological literacy, but not full information, he/she can be said to have a blind spot, the implication of which can only be established depending on the specifics of the case. Where he/she has full information but cannot understand, it is described as *homoious hemin*: Seery (1988), citing Lee, provides the allegory of an "ordinary man [who] is often very uncritical in his beliefs".

Where the principal is neither able to judge nor observe, Pouryousefi and Frooman describe a position akin to a ward of state, where the agent has an almost sacred, patristic duty to the principal.

There is no reason why this adverse selection model should not apply to DeFi. Certainly, DeFi is complex, enabling Reurink's (2018) concerns in respect of traditional finance to translate to DeFi: complexity of financial products; retail investors' comparative financial illiteracy; de facto undisclosed investment strategies; and competitive pressures on agents not to discharge, in an unfettered manner, their fiduciary or regulatory duty to assess a principal's suitability for the product. Reurink (2018) further notes the disintermediation in traditional financial services, through the advent of internet sales. This leads to a difficulty in establishing the locus of fiduciary duty, and the person responsible for suitability assessment. This challenge applies to Gardens, which are fully disintermediated.

Thus, Part I of this research seeks to not only establish how clear the Garden's investment strategy is, but how easily performance can be assessed by a retail investor. In this, the research draws on Liang (2001) and Duanmu et al. (2018) and regresses, for each Garden, its value or price with the S&P 500 index (using Ordinary Least Squares with robust standard errors). Additionally, given that the objectives of the Gardens were designed to attract investors into the cryptoasset space, Gardens are regressed against Bitcoin and Ethereum prices, being market leaders in cryptocurrencies and, as Appendix 1 shows, very closely correlated.

Finally, this literature review asks whether Andrade and Newall's (2023) work on cryptoassets as "Gamblified Financial Assets" would hinder the analysis of Part I. They observe,

> Cryptocurrencies can be traded around the clock and produce the allure of big potential lottery-like wins. Frequent cryptocurrency traders often suffer from gambling-related harm, which suggests that many users are taking on substantial risks.

If investors are this irrational, it could raise questions on whether Pouryousefi and Frooman's analysis applies. Ultimately, if Andrade and Newall's analysis were to apply to Gardens, then the argument would be that frequent investors are Pouryousefi and Frooman's wards.

4 Methodology

4.1 Part I: Ordinary Least Squares Regression

Following Schär's (2021) exposition of platforms, the research reviews the following platforms for Gardens of sufficient size and history: Enzyme Finance; TokenSets—Asset Management Simplified; and dHEDGE. Babylon Finance Gardens are discounted due to insufficient history

or comparable size to a collective investment fund. Prices are taken from Enzyme (n.d.); dHEDGE (n.d.); for DeFi Pulse Index, Scalara (2023, November 23); for TokenSets Metaverse Index, BingX (n.d.); and for TokenSets Sushi DAO House, Zerion (n.d.).

Due to this need for a sufficient history, the lowest number of observations in a Garden was 288. Nevertheless, as Table 3 shows, there is only one pool that would be even close to comparable to a small investment fund (DeFi Pulse Index).

Following Liang (2001) and Duanmu et al. (2018), each Garden is regressed with the S&P 500 index, Bitcoin, and Ethereum prices (using Ordinary Least Squares with robust standard errors). In addition to the reasons stated in the literature review, given that the objectives of the Gardens are designed to attract investors into the cryptoasset space, I regress against Bitcoin and Ethereum prices, as they would be acknowledged as market leaders.

4.2 Part II: Sharpe Ratio

Next, the research selects collective investment funds investing in a similar asset class, whether cryptoasset, metaverse or wider DeFi, from the Crypto Fund Research (n.d.) database. The definitions of investment strategy, and therefore asset class, are as per the Crypto Fund Research. This avoids consistency of taxonomy problems highlighted by Zalan and Barbesino (2023), particularly with regards to the metaverse.

The observations of the selected collective investment funds and Gardens for this Part of the exercise were from 30 September 2021 to 31 August 2022. The Sharpe ratio is calculated for each Garden using end of month prices and, for the 60 collective investment funds selected from the Crypto Fund Research database, the Sharpe ratio is recalculated to take account of the precise time period under observation. Excess returns were calculated using the yield on 3-month Treasury Bills as the risk-free rate (Board of Governors of the Federal Reserve System 2023).

The research does not calculate the Treynor ratio as there is insufficient data to arrive at an individual beta for each fund and Garden. The only credible beta found was a simple crypto beta, calculated by Botte (2022). Applying the same beta to the Gardens and collective investment funds would render a ranking equivalent to a non-risk-adjusted performance. In any event, it is not clear how Gardens, operating as over-the-counter trades on an internet platform, fit into the concept of the market as described in the Capital Asset Pricing Model.

The period reviewed saw the beginnings of what became known as the crypto winter, a rapid decline and crisis in confidence in the cryptoasset industry. Although the Sortino Ratio, which focuses on losses alone, might have some merit in terms of identifying the least worst stocks, the overall volatility would commend a move away from a loss-focused tool.

5 Results

5.1 Part I: Ordinary Least Squares Regression

Table 3 shows the correlation coefficients, all regressions of which gave p-value at 0. The figures within Appendix 1 show how each garden has performed against Bitcoin, Ethereum, and the S&P 500.

dHEDGE Stablecoin Yield

This strategy is clear and has worked well, showing price stability. In this respect, it has outperformed Bitcoin and Ethereum, but perhaps this is not surprising given the exposure to Frax. As the author's (Le Page 2022) previous paper proposed, Frax, with its partial algorithmic trading, would be more likely to stay pegged to the US dollar than most stablecoins. In any event, there is certainly minimal cross-elasticity between this garden and Bitcoin and Ethereum. Compared to S&P 500, its performance has held up.

dHEDGE Top Index

The dramatic crypto collapse in May/June is muted within the dHEDGE Top Index performance, and this is borne out by the coefficients shown in Table 2. The muting would suggest that the clear strategy has, to a degree, worked.

dHEDGE Yield Farming and Trend

The vague, and allowance for many possibilities, investment strategy has at least produced, albeit for a short term, a broadly stable return, largely in line with S&P 500. Investors would find it difficult to argue against this performance.

Enzyme Civa

There was no clear investment strategy given. The performance mirrors and strongly correlates with Bitcoin's and Ethereum's devaluation. The performance since its inception in November 2021 has been poor.

Enzyme DeFiable Mid Small Caps

The language of absolute return and hedge fund defines the strategy. The results define failure, with no mitigation in performance apparent.

TokenSets DeFi Pulse Index

As shown in Table 3 and Appendix 1, Fig. 7, the losses in the cryptoasset sector, exemplified by Bitcoin and Ethereum, are not mitigated. As an index tracker, this is not surprising.

TokenSets Metaverse Index

Table 3 and Appendix 1, Fig. 8 demonstrate that the Metaverse Index has had an even more volatile time—positively and negatively—than cryptoassets. It would, despite the opacity of the strategy, be hard to argue against this performance given the year under review and how even the FANG+ has been hit hard in 2022.[1]

TokenSets Sushi DAO House

Note (b) to Table 3 presents the investment objective as minimising risk in a market downturn. The Garden can claim to have limited the risk through the sub-one positive correlation to Bitcoin and Ethereum in the OLS regression. Therefore, it might be argued it has done what the strategy had indicated.

5.2 Data Collection Part II: Comparison of Gardens with Cryptoasset Hedge Funds

Table 4 shows the top-ranked 21 Gardens and collective investment funds by Sharpe Ratio. Below number 21 the Sharpe ratio is negative, suggesting that no individual should invest as the mean return is less than the mean risk-free rate of return.

Four out of the top five ranked are gardens.

6 Conclusion/Discussion

The above result has placed considerable doubt over the hypothesis that Gardens present an imminent danger of financial loss to retail investors. Part II of the research suggests that four Gardens outperform all but one of the examined sixty traditional investment funds invested in the DeFi asset class. Nevertheless, the *caveat* is that the other four Gardens selected in the sample had a negative Sharpe ratio, presenting a mean return below the mean risk-free rate. Gardens should still be approached with caution by the public.

This caution might be reinforced by the absence of evidence that principals are determining the investment strategy. If anything, there is substantial evidence that agents continue to write the strategy as an advert to investors. Therefore, we are left with a careful reconsideration of the propositions of Pouryousefi and Frooman (2019), in particular whether the investment strategy is clear.

Of the eight Gardens, six would be considered clear strategies/objectives by hedge fund or index tracking standards, if Reurink's (2018) analysis, discussed in the literature review, is fair.

One investment strategy (dHEDGE Yield Farming and Trend) *prima facie* seems opaque:

> You can put any amount you like into this pool, and I will try to manage the risk. This means, the returns will be lower than in high-risk pools, because we'll trade only with 10–50% max.
>
> *I might update this description if I decide to adjust the strategy a bit.*

The italics are this author's emphasis. If ever there were a more obvious agent-written contract…

However, the creator of dHEDGE Yield Farming and Trend may be neurotically over-conscientious, rather than deliberately obtuse, seeking to disclose to investors possibilities in different states of the world. This idea, of a conscientious manager, may indeed be supported by the broadly stable return the pool has provided for investors, as shown in the results section: certainly, the low-risk approach disclosed appears to have paid off.

It is this that exposes a limitation to Pouryousefi and Frooman's (2019) work. Where the performance is understandable, despite its meandering investment strategy, then surely most investors would be able to judge whether the investment pool is a good investment. If, to mimic one of the United Kingdom's legal tests for market abuse within the UK Financial Services and Markets Act (2000), it could be argued that an independent regular user would be able to see the reasonability of the performance in the light of the objective, then it stands to reason that an investor could both judge and observe it. And, as stated in Sect. 5.1.3, the performance seems reasonable in the context. This paper proposes that this concept, outlined in longhand in this paragraph, could be added to Pouryousefi and Frooman's Table 2 and named *post facto* contextual information symmetry.

Now take the other fund with an unclear objective, Enzyme Civa. This would appear to fall within Pouryousefi and Frooman's (2019) definition of wardship. It discloses its current holdings, but there is no strategy/objective given. Its performance over the period of review is abject. This may suggest that, in the absence of a clear strategy to which the agent can be held accountable, failure might have been inevitable. This is a strong statement, but not without merit. If it is not without merit, it would seem reasonable to argue

[1] For chart see ICE. (n.d.), accessed 27 September 2022. FANG+ was defined at the time as Facebook, Apple, Amazon, Netflix, Google plus Microsoft, Alibaba, Baidu, NVIDIA and Tesla.

Table 3 Summary of Investment Pools Correlation with S&P 500, Bitcoin Price and Ethereum Price

Platform	Name	Assets Under Management in August 2022 USD	Number of investors (Number of observations)	Start date	Strategy	Coefficient S&P 500 (*p*-value)	Coefficient Bitcoin price (*p*-value)	Coefficient Ethereum price (*p*-value)
dHEDGE	dHEDGE stable-coin yield	1,960,000	1153 (353)	1 August 2021	Algorithmically farm the highest yield stable pool. Exposure to FRAX, USDC, DAI, USDT	−0.193 (0)	−0.0469 (0)	−0.037 (0)
dHEDGE	Top Index	802,100	247 (353)	1 April 2021	Actively managed index, tracking best dHEDGE traders	3.27 (0)	0.689 (0)	0.578 (0)
dHEDGE	Yield farming and trend	254,700	4 (314)	2 October 2021	See *Note* (a)	0.630 (0)	0.133 (0)	0.107 (0)
Enzyme	Civa	1,136,011	147 (288)	Not available	Discloses token holdings and currently no leverage, no clear strategy given	3.837 (0)	0.937 (0)	0.7243 (0)
Enzyme	DeFiable mid small cap	684,876	57 (353)	May 2021	Long/short multi-strategy	7.229 (0)	1.490 (0)	1.257 (0)
TokenSets	DeFi pulse index	32,463,347	15,665 (353)	9 September 2020	Tracks DeFi pulse index	7.08 (0)	1.45 (0)	1.222 (0)
TokenSets	Metaverse index	5,783,690	10,348 (353)	2 April 2021	Tracks Metaverse index: entertainment, sports, business going to metaverse	8.96 (0)	1.74 (0)	1.502 (0)
TokenSets	Sushi DAO house	4,254,762	4 (351)	27 May 2021	See *Note* (b)	3.96 (0)	0.8302 (0)	0.705 (0)

Note (a) dHEDGE Yield Farming and trend: "The strategy is: Keep the largest part of the portfolio in low-risk LP yield farming at all times…Buy some interesting coins when there is a clear uptrend in the complete space (meaning in BTC), and let the profits run until the trend turns around. Trends will mostly run for weeks or months, and I won't swing-trade much in between. You can put any amount you like into this pool, and I will try to manage the risk. This means, the returns will be lower than in high-risk pools, because we'll trade only with 10–50% max. I might update this description if I decide to adjust the strategy a bit"

(b) "…actively Managed Portfolio that is governed by the Sushi DAO. The investment objective…is to preserve capital and minimize risk during a market downturn in order to support the long-term growth and development of the Sushi Ecosystem. This allows a more dynamic allocation of assets with better risk handling and analysis without loss of decentralization. sushiHOUSE is the one of Yam.Finance's 'DAO House' risk products designed to serve as a DAO cornerstone for achievement of future growth and ongoing success".

Table 4 Top 21 Crypto hedge funds and pools by sharpe ratio

Rank	Name	Mean excess monthly returns over monthly risk-free rate	Standard deviation of return	Sharpe ratio
1	TokenSets Sushi DAO House	0.5147688	0.282067	1.82498768
2	DH dHedge Stablecoin Yield	132.4926	92.75612	1.428397393
3	TokenSets DeFi Pulse Index	1.284234	1.510618	0.850138155
4	Non-Directional Fund	2.613272	3.104797	0.841688523
5	TokenSets Metaverse Index	1.0771734	1.509506	0.713593321
6	Icoinic Delta Neutral	1.863992	2.635546	0.707250794
7	Digital Opportunities Class	0.5459007	0.7767117	0.702835685
8	Ada Capital Fund	0.4950673	0.805705	0.614452312
9	Elysium Global Arbitrage Fund	3.420067	6.418919	0.532810431
10	Delta Neutral Fund	0.7400673	1.653554	0.447561616
11	Altana Specialty Finance	10.72507	34.67657	0.309288664
12	Eltican Neutron SP	1.883401	6.723664	0.280115276
13	Genesis Alpha Fund	13.15173	49.46581	0.265875157
14	Axon Global Fund LP	2.928401	14.48374	0.202185416
15	Full Strategy Fund	0.461224	2.56002	0.180164217
16	Bohr Arbitrage Crypto Fund	1.245067	8.855363	0.140600335
17	Lavaliere Capital, LP	8.277567	90.80263	0.091159992
18	The Etherbridge Fund	3.9534	51.26149	0.077122222
19	Alt Tab Capital	1.498401	21.07689	0.07109213
20	AD Long Short Fund	0.0325673	1.184372	0.027497526
21	Crypto Asset Fund, LLC (Class X)	0.0059007	2.930372	0.002013635

this is also *post facto* contextual information symmetry: an unclear strategy, so steer clear investors.

Combining the arguments in the above two paragraphs, *post facto* contextual information symmetry may be defined as *an investment decision made by the principal on the basis of information that might not be clear from the investment objective/strategy, but the performance of the Garden is sufficiently coherent with, or antagonistic to, the principal's interpretation of the objective/strategy; or has performed either well or badly irrespective of the clarity of the strategy.*

Figure 1 demonstrates the extension of Pouryousefi and Frooman's (2019) adverse selection table into a flowchart. First we consider whether there is performance data, and how coherent—or antagonistic—it is to the investment strategy/objective. If it can be assessed, then we have *post facto* contextual information symmetry. If it cannot, then Pouryousefi and Frooman's table should be consulted in terms of clarity of contract, i.e. investment strategy.

Illustrating this extended model's application against the other Gardens: the two indices were procyclical as expected, but with nuances. One, dHEDGE Top Index, did track Bitcoin and Ethereum downwards, but the data shows some performance mitigation. Again, an investor could consider this performance mitigation in the light of *post facto* contextual information symmetry. The other, TokenSets DeFi Pulse Index, performed worse than Bitcoin and Ethereum with higher positive correlation coefficients. Again an investor could assess this via *post facto* contextual information symmetry as a poor investment opportunity.

Meanwhile, TokenSets Metaverse Index did far worse than Bitcoin and Ethereum. Nevertheless, this also could fall within *post facto* contextual information symmetry, given that it is well known that FANG+ had a terrible year in 2022.

TokenSets Sushi DAO House remains somewhat of a mystery in terms of clarity of strategy. However, an investor could assess it under *post facto* contextual information symmetry, very positively, given the high Sharpe ratio. An investor could see this ratio and rationally decide to invest.

There is one fundamental *caveat* to the findings of this research. The data used on Gardens is from website platforms and, whilst one might expect the platform provider to have ensured the accuracy of the performance data provided by each Garden, there is no reassurance this is the case. Certainly, there is no suggestion of independent audit. All the data could have been manipulated—not necessarily to the point of incredulity, but to a point of mitigated losses or inflated gains. Platforms need to provide academic researchers—and investors—with independently verifiable and audited data for there to be greater confidence in any research findings.

In the absence of such checks and balances, it is difficult for an investor to see how the agent has discharged its fiduciary duty to the Garden. And it is difficult to judge how the platform has allowed a manager on to its systems in the first place. It looks as if anyone can in the name of free market accessibility, cryptoanarchism evolved to DeFianarchism, and no less irrational.

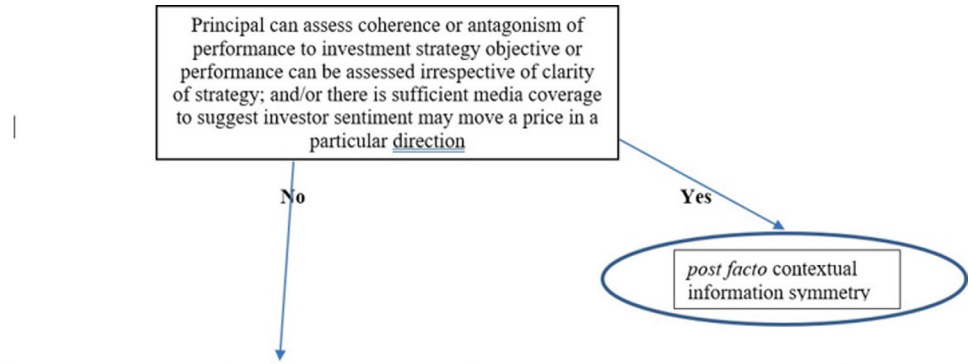


Fig. 1 Revised model for adverse selection

	Principal can judge other performance information	Principal cannot judge other performance information
Principal can observe other performance information	Symmetric information	*homoious hemin*
Principal cannot observe other-performance information	Blind-spot	Wardship

The research is also limited by being bookended within a short time of observation. This is due to the novelty of Gardens, and the decision to avoid the crypto winter at its harshest, with its extreme price drops. Whilst the Sharpe ratio assumes a normal distribution, a period of only 1 year, from 30 September 2021 to 31 August 2022, means all Gardens and collective investment funds are unlikely to be distributed normally. Attempting to solve this via logging the numerator and denominator of the Sharpe ratio by the natural logarithm could not be done due to some observations being negative.

The methodology section indicates that the Sortino and Treynor ratios were not used. Further research establishing whether such Gardens could be pulled into overall market data, thereby enabling betas to be worked out with conviction, would provide a more historically sound analysis on investments and support Vergara-Fernández et al. (2023). For the purposes of this paper, it was not critical given that retail investors are unlikely to be able to diversify investments to the extent of the market, as required by the Capital Asset Pricing Model. Nevertheless, further research could consider the decision-making processes of professional, sophisticated and any investor who is able to diversify to the market portfolio.

For further examination of the limitations on retail investors' ability to make a correct investment decision into a Garden, research should consider Belaşcu et al.'s (2022) concern over risks in asset/liability mismatches and the ensuing instability in product that would happen thereon. Further work could follow up on Reurink's (2018) and Whitty's (2020) concerns of investor susceptibility to fraud. These two risk-focused pieces of research would be obvious next steps to build on this paper.

And what of Ramon Recuero's caring, sharing investment community? Recuero (2022) announced Babylon Finance's decision to close. The reasons blamed a hack into an investment. Whether or not the world was ready for Babylon, and certainly Recuero was no Xerxes conquering all, there has been an unknown quantum of loss of property. Good old-fashioned lack of secure custody put paid to that experiment and confirmed the cyber security risks outlined by Belaşcu et al. (2022). As time passes, further valuable research could be performed on investors' ability to understand the specific cyber security risks and to what extent the cost to *them* of fast shutdowns of platforms, like Babylon Finance, is known and therefore priced into their decision to buy/hold/sell Gardens. Such work would need to consider whether platforms for Gardens need to also consider burial grounds, i.e. regulation insisting on living wills.

Acknowledgements I would like to gratefully acknowledge the considerable input that Professor Pablo Winant, as supervisor, has had in the work leading to this exploratory paper, from which further research is being done for a final paper ("Paper II" for my Global Executive Ph.D). My thanks also goes to Amber Thomas for reading and offering edits and comments over several versions of this paper.

Appendix 1

Comparison of Selected Gardens with Bitcoin, Ethereum and S&P 500

See Figs. 2, 3, 4, 5, 6, 7, 8 and 9.

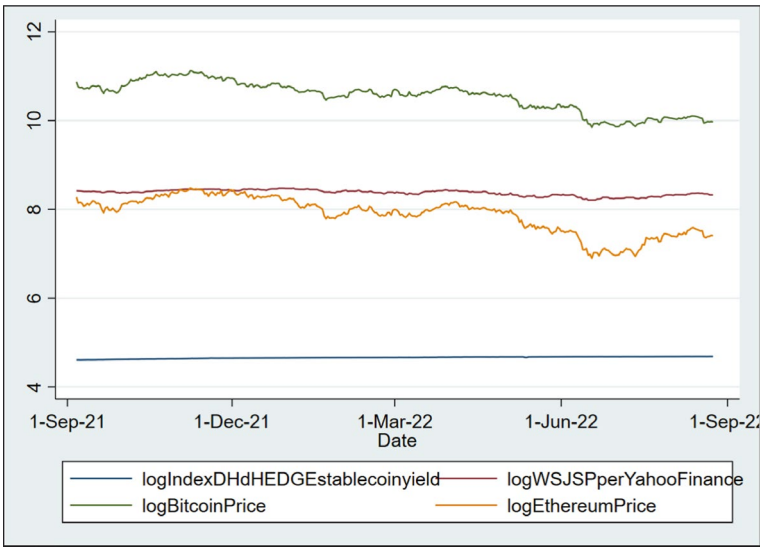


Fig. 2 dHEDGE Stablecoin yield

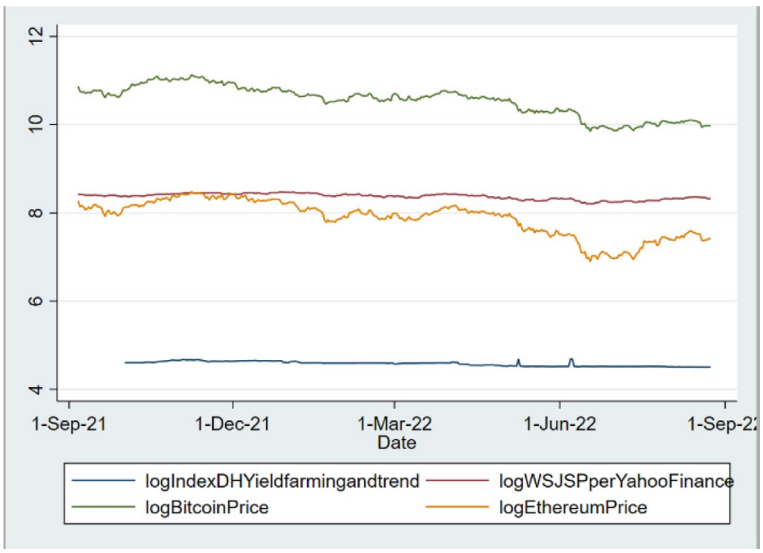


Fig. 3 dHEDGE top index

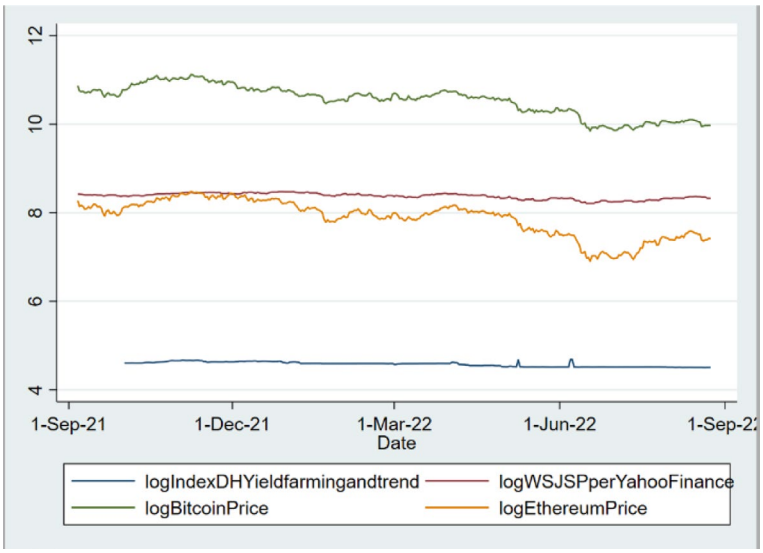


Fig. 4 dHEDGE yield farming and trend

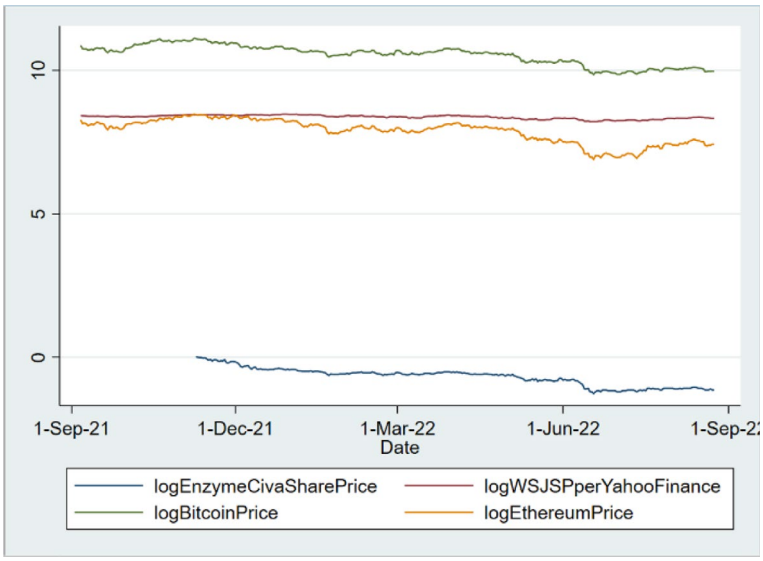


Fig. 5 Enzyme Civa

Fig. 6 Enzyme DeFiable mid small caps

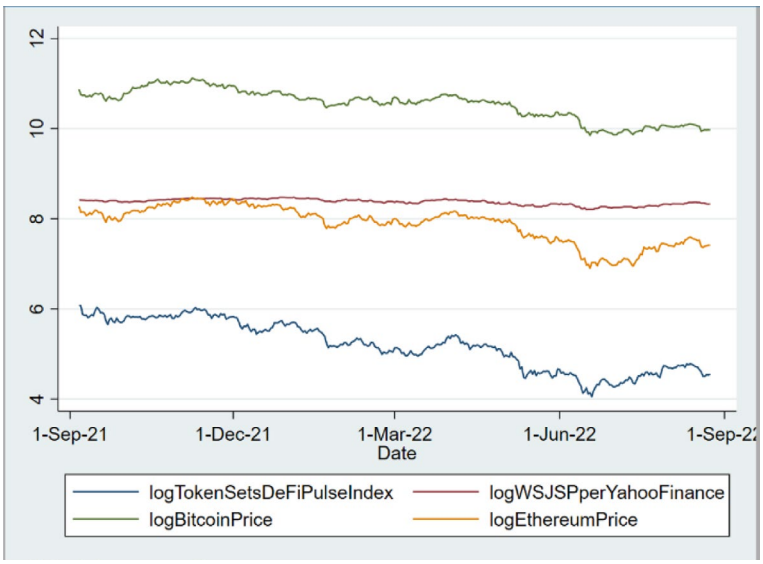


Fig. 7 TokenSets DeFi pulse index

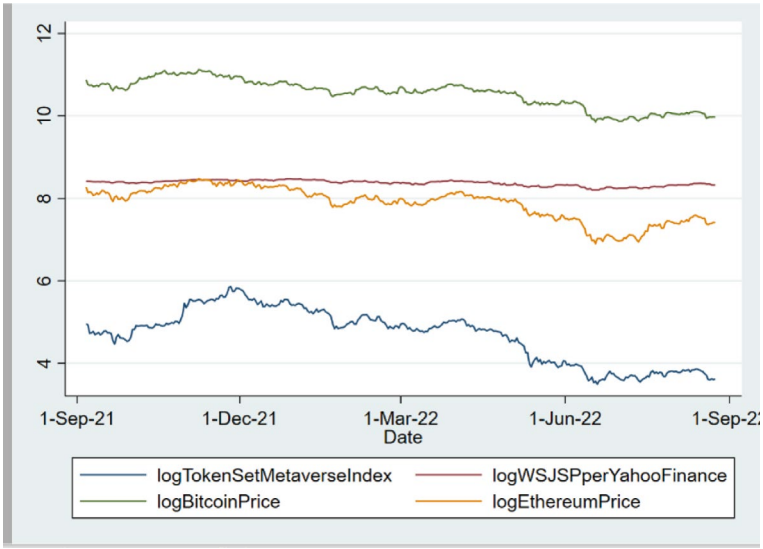


Fig. 8 TokenSets Multiverse index

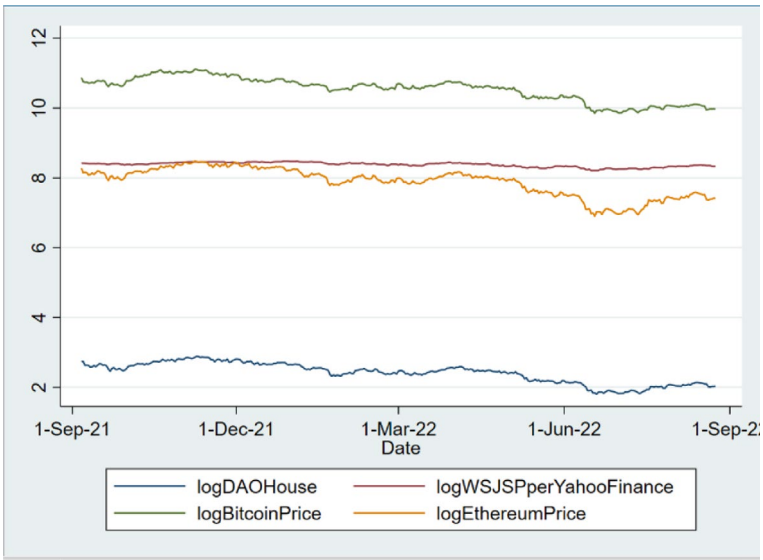


Fig. 9 TokenSets Sushi DAO house

References

Andrade M, Newall PW (2023) Cryptocurrencies as gamblified financial assets and cryptocasinos: novel risks for a public health approach to gambling. Risks 11(3):49

Anonymous (n.d.) What is yield farming?: get started with Bitcoin. com. Bitcoin.com. https://www.bitcoin.com/get-started/what-is-yield-farming/

Anonymous (2022) What are actively managed ETFs? Fidelity What are actively managed ETFs?, June 3. https://www.fidelity.com/learning-center/investment-products/etf/types-of-etfs-actively-managed

Arslanian H (2022) Decentralised Finance (DeFi). In: The book of Crypto, Chapter 1. Palgrave Macmillan, Cham, pp 291–313. https://doi.org/10.1007/978-3-030-97951-5_16

Belaşcu L, Horobeț A, Mnohoghitnei I (2022) Bitcoin is so last decade—How Decentralized Finance (DeFi) could shape the digital economy. Eur J Interdisc Stud, Bucharest Econ Acad (1), March 2022. https://doi.org/10.24818/ejis.2022.01

BingX (n.d.) Metaverse index price. https://bingx.com/en-us/prices/metaverse-index/

Board of Governors of the Federal Reserve System (2023) 3-month treasury bill secondary market rate, discount basis. FRED. https://fred.stlouisfed.org/series/DTB3

Botte A (2022) Crypto beta: the "Layer 1" of a digital asset-specific risk model (pun intended): portfolio for the future. CAIA, March 22. https://caia.org/blog/2022/03/22/crypto-beta-layer-1-digital-asset-specific-risk-model-pun-intended

Crypto Fund Research (n.d.) The #1 list of Crypto funds. https://cryptofundresearch.com/

Cutler JR, Kaplan L, Williamson J (2023) Session 5: Banking, capital markets, and the crypto revolution—a look back and projection of the future of Fintech. Seattle University School of Law Digital Commons. https://digitalcommons.law.seattleu.edu/sitie_symposium/sitie2023/introduction/7/

Dawson K, Allan B (n.d.) Defi and the decentralisation illusion. KPMG. https://kpmg.com/xx/en/home/insights/2022/07/defi-and-the-decentralisation-illusion.html

dHEDGE (n.d.) Managed and automated vaults accessible to everyone. https://dhedge.org/

Duanmu J, Malakhov A, McCumber WR (2018) Beta active hedge fund management. J Financ Quant Anal 53(6):2525–2558

Enzyme (n.d.). https://app.enzyme.finance/discover

European Commisssion (n.d.) Investment funds. Finance. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/financial-markets/investment-funds_en

Financial Action Task Force (2023) Financial action task force recommendations, November 2023. https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html

Financial Conduct Authority (2023) Conduct of Business Sourcebook, Release 31 (UK), November 2023. https://www.handbook.fca.org.uk/handbook/COBS.pdf

Financial Services and Markets Act (2000), s. 118. https://www.legislation.gov.uk/ukpga/2000/8/section/118/2001-12-01#section-118-10

Hayes A. quoted in Robertson B, Robinson E (2021) What's at stake in the U.S. case against a Crypto rebel. Bloomberg.com, N.PAG

ICE (n.d.) NYSE fang+ index futures. https://www.ice.com/fangplus

Jenkins S (n.d.) Index funds. Index Funds. https://www.hl.co.uk/funds/index-tracker-funds

Kim S (2021) New crypto-secured lending system with a two-way collateral function. Ledger, vol 6

Le Page M (2022) Ill-defined. An exploration of whether stablecoins are a significant landmark in the journey to fully-fledged independent currencies, not published

Liang B (2001) Hedge fund performance: 1990–1999. Financ Anal J. CFA Institute

McCann B (2014) Tactical portfolios: strategies and tactics for investing in hedge funds and liquid alternatives. Wiley (Wiley Finance Series), Hoboken, New Jersey, p 1, pp 104–106. https://search-ebscohost-com.revproxy.escpeurope.eu/login.aspx?direct=true&AuthType=uid,ip,url,cookie&db=nlebk&AN=712376&site=ehost-live&scope=site

Pouryousefi S, Frooman J (2019) The consumer scam: an agency-theoretic approach. J Bus Ethics 154(1):1–12

Recuero R (2021) Introducing Babylon finance. Medium, May 8. https://medium.com/babylon-finance/introducing-babylon-finance-8199fa89f918

Recuero, R. (2022) Babylon Finance is shutting down. Medium, November 15. https://medium.com/babylon-finance/babylon-finance-is-shutting-down-b58abf1bc251

Reurink A (2018) Financial fraud: a literature review. J Econ Surv 32(5):1292–1325, December 2018

Scalara (2023) Scalara-Defi Pulse index: Stats, charts and guides: Scalara, November 23. https://www.scalara.xyz/, https://www.scalara.xyz/indices/dpi

Schär F (2021) Decentralized finance: on blockchain and smart contract-based financial markets, Federal Reserve Bank of St Louis Review. Second Quarter 103(2):153–174. https://doi.org/10.20955/r.103.153-74

Schmidt C (2021) 2020: Year in Review. Defi Pulse Blog, July 14. https://www.defipulse.com/blog/2020-year-in-review

Seery JE (1988) Politics as ironic community: on the themes of descent and return in Plato's Republic. Polit Theory 16(2):229–256, citing Lee D. tran. (1980). The Republic, 317n. Penguin Books, New York, New York

TokenSets (n.d.) Asset management simplified. https://www.tokensets.com/#/

Tretina K (2023) How to invest in hedge funds. Forbes, January 12. https://www.forbes.com/advisor/investing/how-to-invest-in-hedge-funds/#:~:text=Minimum%20initial%20investment%20amounts%20for%20hedge%20funds%20range,the%20year.%20Hedge%20funds%20also%20carry%20hefty%20fees

U.S. Securities and Exchange Commission (2017) Eliminating the prohibition against general solicitation and general advertising in Rule 506 and Rule 144A Offerings, July 14. https://www.sec.gov/info/smallbus/secg/general-solicitation-small-entity-compliance-guide#:~:text=Rule%20506%20%28c%29%20of%20Regulation%20D%20Section%20201,to%20verify%20that%20the%20purchasers%20are%20accredited%20investors

Vergara-Fernández M, Heilmann C, Szymanowska M (2023) Describing model relations: the case of the capital asset pricing model (CAPM) family in financial economics. Stud Hist Philos Sci 97:91–100

Whitty MT (2020) Is There a scam for everyone? Psychologically profiling cyberscam victims. Eur J Crim Policy Res 26:399–409. https://doi.org/10.1007/s10610-020-09458-z

Zalan T, Barbesino P (2023) Making the metaverse real. Digit Bus 3(2). https://doi.org/10.1016/j.digbus.2023.100059

Zerion (n.d.) Sushi DAO house price. https://app.zerion.io/tokens/sushiHOUSE-0x7b18913d945242a9c313573e6c99064cd940c6af

The Nature Smart Future—In Search for the Next Gen Innovation

Anne Stenros

Abstract

The 50-year history of innovation, 1980–2030, has seen radical changes from technological to human-driven innovation and will move still further towards human-planetary well-being as the future goal. In our current world of multiple crises, we have only one way out: Nature Smart Design, based on systems thinking and creative thinking, and supported by human-like AI, geodesign, circular design, biophilic design, and a regenerative approach. Through these multiple perspectives and collective wisdom, it is possible to create all things artificial: cities, technologies, transportation, urban food production, culture, and societies that are more resilient, sustainable, equal, innovative, and creative than before. The future Innovation Atelier is a new type of innovation hub with an exciting co-location and diverse groups of creatives. These multi-sector innovation hubs span a range of business models, ownership structures, and physical spaces. Their goal is to create an innovative working culture and environment where businesses of all kinds can learn from each other, make connections, develop new skills, and become inspired to reach the next level. Many of these hubs occupy iconic buildings, including museums, warehouses, train stations, and navy yards, giving new life to micro-localities that have lost some of their vibrant life and attraction. These future Innovation Ateliers are also a great way to explore new ways to express creativity and creative thinking at large.

Keywords

Creative thinking · Circular design · Nature smart · Planetary wellbeing · AI-generated design · Innovation hub

A. Stenros (✉)
Helsinki, Finland
e-mail: anne.stenros@kolumbus.fi

1 Introduction: The Short History of Innovation 1980–2030

> Creativity is thinking up new things. Innovation is doing new things.
>
> —Theodore Levitt

When talking about innovation, it is important to define what we mean with the term. Originally, the term innovation (as we know it today) was first coined in the 1980s to describe new, emerging technologies. The first generation, Innovation 1.0, was the era of techno-optimism that built the foundation for technologies we use today, such as personal computers, compact discs, the space shuttle, and the artificial heart. The second generation, Innovation 2.0, was an era marked by all things digital: the World Wide Web, Text Message SMS, and Google, among other things.

In the turn of the millennium, design started to play a bigger role within leading technologies such as the iPhone and the iPad, as user experience (UX) became the mantra of the first decade of the millennium. The next phase was open innovation and platform innovation, and the breakthrough of Artificial Intelligence AI as well as robotics and self-driving cars. In the next generation, Innovation 3.0, we are talking about systems innovation and quantum computing, but also circular economy and biomimicry. In the coming decade, the focus will be on social and sustainable innovations and nature smart solutions rather than technological innovation alone.* * *

The term *innovation* became a leading mantra in the beginning of the millennium among all the businesses. Governments created innovation agendas and policies, and cities built innovation hubs. It was the heyday of technological innovation, the peak of the development that started a few decades earlier in the 1980s. Originally, the term innovation (as we know it today) was first coined in the 1980's to describe new, emerging technologies. The first generation, *Innovation 1.0*, was an era of techno-optimism that built the foundation for all the technologies we use today, such

A. Olanrewaju and S. Bruno (eds.), *Advancements in Architectural, Engineering, and Construction Research and Practice*, Advances in Science, Technology & Innovation, https://doi.org/10.1007/978-3-031-59329-1_17

as personal computers, compact discs, the space shuttle, and the artificial heart, among others. Industries built R&D teams and businesses started to invest on research and development centers. There was both a change in mindset and a change in behavior in product and technology development when in the 1990s the digital development went mainstream with the World Wide Web and SMS, for example. The second generation, *Innovation 2.0*, was marked by a change of culture and a change of framework. It was the golden era of design thinking, user-centered design, user experience and usability, as well as reinventing tech: the birth of the iPhone and the iPod. The concept of open innovation, followed by innovation hubs and platforms, gathered diverse groups of experts looking for systems solutions, such as self-driving cars, artificial intelligence and virtual reality (Fig. 1).

The next generation, *Innovation 3.0*, is underway. It started as a shift of the worldview ignited by the climate change, soon to be known as the climate crisis. Sustainable design, circular economy, and systems thinking are approaches that support the systems innovation development. Today, we are heading to a shift of paradigm. To survive as a planet, we need to focus on nature smart innovations, supporting overall planetary health and well-being. Next-generation innovation development is built upon the power of creative thinking and regenerative design supported by creative collectives within innovation ecosystems. *Future innovation bridges tech smart knowledge with nature smart wisdom, contributing to the greater good of humanity and the planet.*

2 Towards Next Gen Innovation—Creativity on Demand

> Every child is an artist. The problem is how to remain an artist once we grow up.
>
> —Pablo Picasso

Some 15 years ago, I gave my first speech on innovation in front of a technology industry audience. I had just started my role as the first Design Director at KONE, a leading global elevator and escalator company. At that time, the mantra of the decade was technological innovations, and I praised their importance. However, at the end of my talk, I stated that when companies are craving for innovations, they don't realize that they need creative people to do the job. By sharing the famous picture of Albert Einstein, with his mouth open, showing his tongue, and hair messy, I boldly said: "But when you meet creative people, you don't like them". The truth was—and still is—that creatives are often too messy, too curious, too creative, or too crazy to fit in conventional organizations. My audience of technology CEOs was completely silent at the end of my talk. Ever since, I have tried to speak in support of the creatives, creative thinkers, and creative leaders. They come in many shapes and disciplines. Their curiosity, courage, creativity and craziness make them unique in solving the complex, even wicked problems that we are facing today.

When the World Economic Forum recently published their annual Future of Jobs 2023 survey, they pointed out the importance of creative thinking as the second most

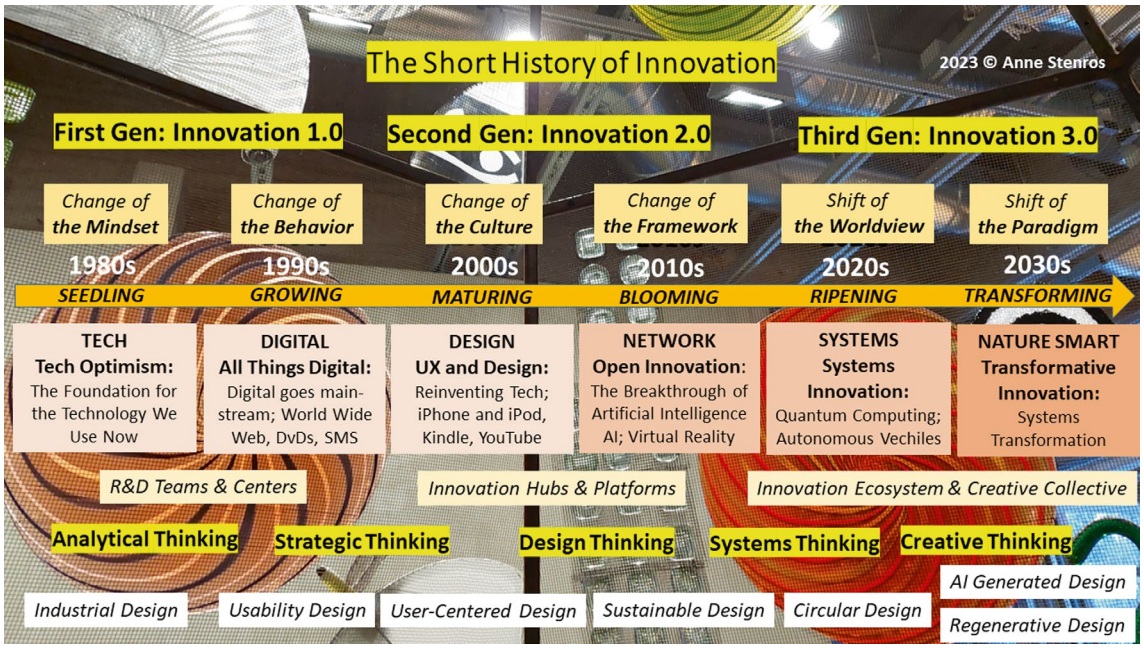


Fig. 1 The short history of innovation 1980–2030. © Anne Stenros, 2023

important skill after analytical thinking. Said in the report: "Analytical thinking and *creative thinking* remain the most important skills for workers in 2023. Analytical thinking is considered a core skill by more companies than any other skill and constitutes, on average, 9% of the core skills reported by companies. *Creative thinking*, another cognitive skill, ranks second, ahead of three self-efficacy skills—resilience, flexibility, and agility; motivation and self-awareness; and curiosity and lifelong learning—in recognition of the importance of workers' ability to adapt to disrupted workplaces" (WEF, 2023). On the list of skills on the rise, *Creative Thinking* is number one, followed by analytical thinking. This clearly shows the utmost importance of creativity, curiosity, imagination and thinking out of the box, next to technological advancement. In the same report it is said: "Technology adoption will remain a key driver of business transformation in the next five years. Over 85% of organizations surveyed identify increased adoption of new and frontier technologies and broadening digital access as the trends most likely to drive transformation in their organization. Broader application of Environmental, Social and Governance (ESG) standards within their organizations will also have a significant impact" (WEF, 2023).—*Summa summarum: In the coming years, creativity, technological skills as well as environmental and social drivers will lead the change within the working culture.*

In 2020, NASA conducted its famous study on human creativity and the impact of education. Based on their research, the results were alarming. The scores showed that 98% of 5-year-old children fell into the "genius category of imagination". However, the number dropped to 12% for 15-year-olds, and only 2% of adults were in the genius category anymore (NASA 2023). Our education system, which is based on *convergent thinking* does not support *divergent thinking*, which is necessary for our capacity to imagine, stay curious, and generate creative ideas. Convergent thinking is the *solution-driven approach* focusing on finding concrete and familiar solutions to problems. Divergent thinking is the creative, *problem-driven approach* generating original ideas and new possibilities.

Our innovation capacity is dependent on creative thinking and divergent thinking. How can we sustain and nurture collaborative creative culture in society? The answer is that we must unleash and nurture our potential as creatives, creators, catalysts, and changemakers throughout our entire educational system, starting from the very early years of the lifelong learning path and ending in the unlearning-relearning phase of adult life. *Only by constantly questioning, by being open-minded and curious, is possible to be both resilient and creative enough for future challenges. Creativity is our innate capacity to cope with the future.*

In the coming decade, to solve the complex network of crises, it is necessary not only to focus on the new frontier technologies but also their adaptation and applications in society through creative thinking and creative problem-solving. Creative thinking should be one of the principal skills of learnt at schools and universities in the future. It has been shown that children who have creative hobbies, music, dance, poetry, etc., perform better at any discipline during their studies. Creativity enhances their learning capacity. "Creativity is making new connections, new synapses," says Ivy Ross, who is vice president of hardware design at Google and co-author together with Susan Magsamen of the bestseller *Your Brain on Art: How the Arts Transform Us* (2023). Says Magsamen: "Children that are playing music, their brain structure actually changes and their cerebral cortex actually gets larger" (Hamilton 2023). Creativity, in general, whether it be exercising music, dance, painting, or storytelling, increases the brain's plasticity and its ability to adapt in response to new experiences. It has been noticed in studies that children who engage in the arts are better learners in general. The arts also can teach the brain skills other than in the classroom, which widens the scope of learning and thinking beyond the basic skills and knowledge. (Hamilton 2023). *On the other hand, studies show that creative mindset and creative activities not only support complex problem solving, but also overall personal well-being and happiness.*

The future of innovation is much more a shift in approach and paradigm than technological change. It means that we must reconsider and redirect our educational systems towards more creative thinking in addition to analytical thinking. *Only by unleashing the hidden powers of imagination, curiosity, and creativity, can we conquer the challenges of the future that lies ahead. A culture based on both creativity and technology as well as humanity and sustainability is the way forward for planetary wellbeing.*

There is one more step to take on the future innovation path: we will need more *creative leadership* in the future to give way and room to creativity and creative thinking. The creatives, creative thinkers, and creative leaders come in many shapes, colors, and disciplines. Their curiosity, courage, creativity and even craziness make them unique in solving complex and wicked problems that we face today. *Only through creative leadership, can we support and nurture creative thinking throughout the organization and its entire culture. Making creativity an imperative will result in a more creative organizational culture, and eventually, more innovative outcomes.* Therefore, creative leadership should be on demand at every level of an organization, from teams to the very top. A creative mind is a curious mind—and vice versa—therefore, it is necessary to support lifelong and life-wide learning, micro-learning, unlearning, and relearning as part of the future skill set of creative leaders. By doing so, we are supporting creativity itself, our human lifeline.

3 The Future Innovation Atelier—A Creative Collective

> When one teaches, two learn.
>
> —Robert Heinlein

The future *Innovation Atelier* is a new way to collaborate, co-create, and co-produce in more diverse and creative ways. The Atelier draws in those who master their discipline as well as talented learners and newcomers. It is the future model of a creative collective for learning, sharing, problem-solving, and innovation. Its educational mission is to teach students to create novel and meaningful solutions in the framework of human and planetary health and wellbeing. This is supported by the learnings from *Creative Leadership*, techno-smart knowledge, and nature smart wisdom.

According to TATE: *"Atelier is a French word that translates literally as studio or workshop and is often used to denote a group of artists, designers or architects working collectively"*. Traditionally, the artist's studio also used to be a place for teaching young artists; but, this function was gradually replaced by the rise of the academy. However, in many fields of arts, the master and apprentice tradition is still valid as a teaching method, especially when learning craft skills. One of the most famous large-scale practicing ateliers was in Germany, when in 1919, Walter Gropius founded the Bauhaus as an attempt to unify the arts with the technology of the mechanical age (TATE 2023).

The original idea of an atelier was to combine several disciplines and professionals with diverse backgrounds under the same roof, led by renowned masters in their field of arts, architecture, and design. The studio became a natural model of learning within an innovative environment. Since there were professionals of different age and with different experience, the overall group was diverse enough to learn from each other. Usually, the atelier consisted of the five critical elements to become a success: (1) *skills* and competence, (2) *knowledge* and wisdom, (3) *arts* and creativity, (4) *technology* and materials, and finally (5) *emotion/passion* and motivation.

Skills refer to learning through making and testing. *Knowledge* refers to the power of learning: sharing information, experiences, and exploration. *Arts* refer to the creative approach, driven by curiosity and questioning. *Technology* refers to knowledge of how to implement and experiment. *Emotion* refers to the purpose and the meaning of doing. The combination of skills, knowledge, arts, technology, and emotion proved a winning formula for creatives to grow and develop their full potential as professionals. Following this educational tradition since the Medieval Ages, the idea of the future *Innovation Atelier* could come to life.

Today, we are talking about *innovation ecosystems* as drivers in disruptive innovation and radical problem-solving. McKinsey recently published their Innovation Ecosystem Playbook (2023) with six essentials for innovation hubs, claiming that "innovation hubs open new avenues for healthier, more diverse, and more connected communities". They also state that "innovation ecosystems are more likely to thrive when local leaders and developers play to a region's existing skill base and institutional strengths" (McKinsey 2023).

Six essentials of the Innovation Ecosystem:

1. *Set the aspiration and a bold vision* > **Mission & Purpose**: Set the shared vision together and focus on meaningful problems.
2. *Focus on specific sectors, partners, and anchor tenants* > **Partners & Locals**: Get locals involved and have an open dialogue with all stakeholders.
3. *Catalyze a critical mass of VC capital and start-ups through a strong innovation backbone* > **Resources & Implementation**: Create programs for supporting implementation and facilitating business creation.
4. *Develop an ecosystem talent and workforce strategy* > **People & Micro Communities**: Get diverse groups of doers and makers together, from experienced professionals to young students.
5. *Design high-quality real estate, infrastructure, and livability* > **Space & Place**: Create a functional and inspirational campus that supports an innovative and future-orientated approach by enabling personal growth.
6. *Cultivate a vibrant, diverse community and a sense of place* > **Creative Culture & Creative Community**: Foster the sense of place by supporting the community building and co-creation through different activities and diverse events.

When combining these essentials of the innovation ecosystem with the critical elements of a creative atelier, we'll get the outline of the future *Innovation Atelier*. It is based on local anchor partners as well as international partners in a specific field. Its professional pool and students are a diverse group of talents from different fields and cultures with diverse professional experiences. There are lifelong learners and master students, even bachelor-level students participating in the shared projects of the atelier. *It is a vibrant, livable and lovable space and place where skills, knowledge, arts, science, and technology are combined through personal passion and motivation for creating new and novel ideas and solutions for the betterment of the world.* The principles of circular economy, biodiversity and sustainability are interwoven into the practices and projects of the atelier. Both teachers and students are exposed to *circular design*, *geodesign*, *biophilic design*, *regenerative design*, and *rewilding* as an approach. The members embrace courage, curiosity, and questioning the status

quo. *Systems thinking* and *creative thinking* as approaches are combined with *design thinking* and *futures thinking* as methods. The best practices of the Innovation Atelier are not only based on collaboration, co-working or co-creation but also on co-production as the highest level of the creative culture and creative collective. The scale of the learnings varies from industrial design to placemaking. *The main purpose is to teach and learn the lifelong journey of Creative Leadership, both on the personal, self-leadership level and on an organizational, championing level. That means, in the end, mastering your creative capacity, competence and potential to the fullest.*

As it is based on co-production, the Atelier sets together, with all its members, the common goal: the mission and purpose for all the activities, learning, teaching, and practicing. The innovation projects undertaken by the atelier are always mission-driven with a greater purpose. Personal passion and emotions are the driving force for motivation and high-level achievements among the participants. *The winning formula of the Innovation Atelier is in nurturing the personal creative leadership and growth supported by other creatives, both teachers and peer-students.* The Innovation Atelier is a *Creative Collective* focusing on new ways of thinking and doing. It is combining traditional *convergent thinking* with *divergent thinking* and *analytical thinking* with *creative thinking*. It is a modern learning-by-doing platform for multi-generational and multi-disciplinary experts and practitioners, as well as students and trainees. The master of the atelier is a *Thought Leader* in their field, whose role is to show the way to future innovative development (Fig. 2).

Summa summarum: The Future Innovation Atelier is a community where the transformation is happening in a cross-sectoral and cross-disciplinary culture, in collaboration with science, arts, economy, and technology, driven by emotion and passion and implemented with a multiple skillset. In the Atelier, doing and learning is not linear, but based on an open-end systems approach, where competence, skills, and knowledge are developing through an evolution of curiosity-driven creative culture and learning by making.

Creative hubs, like the Innovation Ateliers, with "exposure to creativity of all kinds stimulates creativity of all kinds". Learning from what others are doing in an open environment speeds up imagination and will lead to new ideas and solutions faster than in conventional, closed systems. (Greenwald 2018). It is also a great way to explore new ways to express creativity and creative thinking.

Training different types of thinking demands different approaches from touching materials (doing) to thinking out-of-the-box (mental models) and everything in between. The culture of the atelier is inspiring and encourages its members to practice any kind of art that is meaningful to them. In their book *Your Brain on Art* (2023), Susan Magsamen and Ivy Ross describe the concept of an *aesthetic mindset*, which "is simply the ways in which you are aware of the arts and aesthetics around you, and how you bring them into your life with purpose". They continue by saying that the people who have an aesthetic mindset, have four key attributes in common: (Magsamen and Ross 2023, xiii).

1. A high level of *curiosity*
2. A love of playful, *open-ended exploration*
3. Keen sensory *awareness*
4. Drive to engage in creative activities as *a maker* and/or beholder.

Based on experience, the creatives are endlessly curious, and open to exploration; they are self-aware and think intuitively, and they love to test and make things. Their learning is based on trial and error, prototyping, constant testing and improving. They are life-long learners by nature. When Francisco Goya made his famous self-portrait at the age of 80, he signed it by writing: *Aún aprendo*—I am still learning. *The true spirit of an artistic and/or creative endeavor is in the passion of learning, unlearning and relearning*

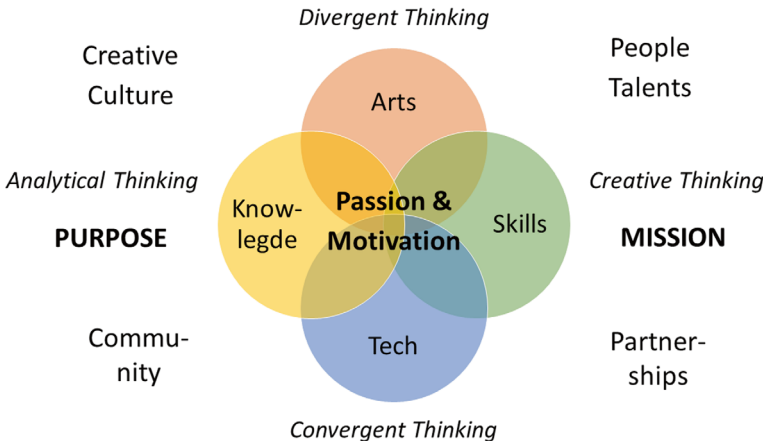


Fig. 2 *Creative Collective:* Elements of the Future Innovation Atelier. © Anne Stenros 2023

throughout life. Creativity is a lifetime learning journey all the time offering new scenes, shores, and peaks, all the way from childhood to the evening of life—creativity IS the spirit of life.

A truly creative ecosystem is rich as is biodiversity, which "is comprised of several levels, starting with genes, then individual species, then communities of creatures and finally entire ecosystems, such as forests or coral reefs, where life interplays with the physical environment" (Carrington 2018). Translated into creativity, this means that there is a *creative culture* that is nurturing the *individual creatives*, who create together microlocal *creative communities* and are networked with similar communities further away, creating an entire *creative ecosystem*. The ecosystem represents also the shared *creative knowledge* learned by creatives during a few thousand years of human history. Said by Magsamen and Ross: "Supporting this core human imperative to live in a community is our unique ability to creatively share our thoughts, ideas, and emotions. The success of our species comes down to this: Art creates culture. Culture creates community. And community creates humanity" (Magsamen and Ross 2023, 202).

Creativity starts with curiosity. Therefore, nurturing curiosity from the early childhood is necessary for a creative mindset. A recent study found out that adventurous play, climbing trees, riding bikes, jumping from high surfaces, or playing out of adult sight, helped children build resilience and supports their holistic wellbeing (Weale 2022).—I remember when we were kids and used to play in the woods, climbing the rocks and trees, testing our skills and courage all summer long. It also created a strong bond with nature, plants, insects, birds, flora and fauna. There was also another perspective: playing in the woods supported our curiosity to go places, to see things, to explore. *Curiosity* goes hand in hand with *courage*, and eventually with *creativity*. Later, I learned that there was yet another element of a creative mindset, namely a personal passion, a *call* that is the driving force for creatives throughout their life and that gives direction to their endless search. The creative ecosystem supports a *creative mindset* on all levels (Fig. 3).

Creativity is, by definition, the ability to imagine and come up with original ideas and solutions.

Creative thinking, on the other hand, allows us to connect information in new and meaningful ways, and therefore, it is the origin of innovation and invention. Creative thinking is one of humankind's most valuable skills (Magsamen and Ross 2023, 185).—Based on research exploring creativity, we can see how creative thinking comes from the interaction of our brain's default and executive control networks, and "that these connections allow us to spontaneously generate ideas and critically evaluate them, respectively". The very same networks we use to recall the past allow us also to imagine the future and think creatively (Magsamen and Ross 2023, 188).

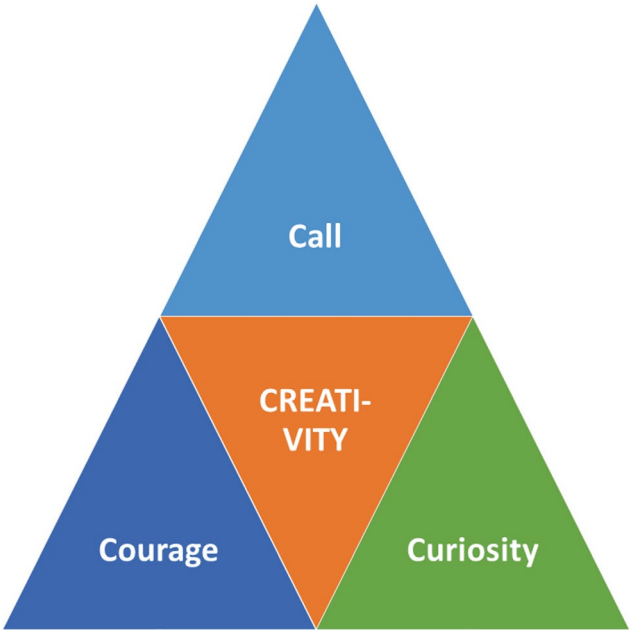


Fig. 3 The creative mindset: creativity is supported by courage and curiosity and driven by a call. AS © 2023

Curiosity, as research has shown, is a building block of positive emotions that lead to happiness. It expands our empathy and strengthens our relationships; curiosity is an emotional state that can be strengthened with practice. (Magsamen and Ross 2023, 173). Curiosity is also linked to open-mindedness and a capacity of lifelong learning and self-renewal. Curiosity is a driving force for *exploration*, for expanding one's understanding of the world around.

Courage is the confidence and faith, and the energy to be able to go to places of uncertainty; to explore and discover, which is in the heart of creative activity and creativity itself. Lorne Buchman, the president of ArtCenter College of Design in Pasadena, supports arts-based studio education, where the connection between the brain and the hand supports learning through making. The hands-on-learning is rooted in discovery and playful engagement with materials and ideas. This artful approach is the authentic expression of any and all individuality. The arts and aesthetics are the building blocks of a full and a vibrant life (Magsamen and Ross 2023, 158–161).

Creative call is our passion, energy, and determination to express our emotions and search for new and novel ideas. It is our North Star leading us to the next and future opportunities. Our creative call is also linked to hope—since the arts have always offered the highest form of hope for civilizations. *The call for creativity is our lifeline; it creates hope for the future to come.*

Creative community is a creative circle of creative-minded people, getting together to share their knowledge and experiences, feelings and emotions, learnings in life. It is a powerful learning community!

4 Towards a Nature Smart Future—The Future Innovation Leader

> The ideal scientist thinks like a poet, works like a bookkeeper, and writes like a journalist.
>
> —Edward O. Wilson

In his latest book, *Design for a Better World* (2023), the renowned design thought leader Don Norman suggests moving from Human-Centered Design to *Humanity-Centered Design*. Norman who coined the idea of user experience and usability in the late 80 s, is now showing the way forward, towards a more sustainable and equitable society. Said by Norman: "These principles [of Human-Centered Design] are important, but they ignore the problems of sustainability, inequity, and bias. In addition, the emphasis is usually on the immediate issues, not on long-term impact. In other words, they describe the things in the past… and not the way we need to do things in the future". Instead, Norman proposes five principles for future Humanity-Centered Design as an approach: Norman (2023, 183).

1. *Identify and solve the root problems* and not the symptoms.
2. *Focus on the entire ecosystem* of people, living things, and the physical environment.
3. *Consider everything as a system* of interconnected parts.
4. *Don't rush to a solution*; instead, embrace the iterative nature of the design process.
5. *Design with the community* and support designs by community.

There is one more principle to add into the list, related to creative and innovative leadership:

6. *Building a strong creative culture* through the creative and innovative leadership

In his book *Thriving* (2022), Wayne Visser writes about innovative leadership as an essential element of a thriving future and the need for breakthrough innovation. Said by Visser: "Leaders with a creative style enjoy playing the role of designer, architect, innovator, game changer, and transformer of systems". Innovation leadership is also missionary leadership, often driven by a cause or a personal call. It is regenerating yourself, as Visser points out: "Personal thriving is about looking after yourself and ensuring that you have the perspective and passion, the inspiration and vision, the energy and support, and the imagination and focus that you need to continue the work of changing the world for the better" (Visser 2022, 289–294).—Into this list of essentials, we can add the dimension of failing and learning, never giving up, as in the Japanese proverb: "Fall down seven times, stand up eight".

The future *creative leadership* is based on *change leadership* and *transformation leadership*. It is *mission-driven* for a better world and *humanity-centered* by approach. It is also about *innovation leadership* by imaging future solutions and building a more sustainable future for all. Without creative leadership, the new type of innovation leaders needed for transformational change, such as the green transition, do not have the right skillset and culture to work with.—"We need to think creatively, to imagine better solutions to complex problems. Then we need to work systematically, to gather the evidence and build a compelling case for action. And finally, we need to be effective storytellers when we communicate about the benefits of changing the status quo" (Visser 2022 295–296).

The future *Innovation Leaders* have the capacity for nature smart thinking, systems thinking, creative thinking, and analytical thinking. They can approach the grand challenges and complex issues with open mind and from several perspectives. They have emotional intelligence for humanity-centered approach and creative spirit to think out of the box. They are imaginative storytellers and can share their compelling call. They are open to regenerating themselves: they are lifelong learners on their way to creative leadership. They are aiming high, and they are mission-driven by a personal purpose. They are leading innovation through their entire character, mind, body, heart, and spirit. They are characterized by wisdom, friendship, love, and bravery. To be a master in innovation and creativity is a lifetime journey of becoming the best version of yourself.

Leadership is learning yourself, knowing yourself, and finding yourself. It is the Japanese three stages of learning any mastery, *shu-ha-ri*: the mastery of learning the rules, then finding your own way to apply them, and finally, abandoning the rules to become a true version of yourself. Innovation leadership also is about learning the rules and abandoning the rules for creating new ones. There is no straight route to the unknown future. We must create our own maps to navigate the future. With great tailwind we will advance fast, but we will also encounter headwinds and storms that slow us down. Leadership is a lonely journey, *hitoritabi* in Japanese, to travel alone. The better a map one can draw of the future, the faster and farther can one travel in leadership.

Changing the perspective is an essential skill for a creative and creative leadership. When focusing on something that is not that evident in the system, it is possible to find new routes, solutions, and openings to the old and existing challenges. *Innovation is all about seeing the world in*

a different way by showing a light and naming something that has been hidden behind the ordinary and well-known. It is about showing something that has been there but has not been seen. *Innovation is a new concept, a new pattern, or a new framework that will eventually rewrite our understanding of the reality and the world around us.*

Innovation needs tools and methods to navigate the unknown areas. Foresight, future design and making of future are ways to explore what is the next. Most creatives are constantly scanning their surroundings, acquiring new knowledge on emerging issues and trends to have a better understanding of the plausible, probable, and possible futures. Many creatives are *living* the future; artists are sensitive enough to recognize the up-and-coming issues within their society. Their intuition is often leading to the right path of sizing up the emerging, hidden challenges and their outcomes. They can also be brutally honest with what they foresee and say it aloud. Innovators, like artists, have imagination to combine small cues and create the bigger picture out of them. In the medieval times, cartographers created '*mappa mundi*' charts, maps of the world. Little by little, these world atlases became more and more accurate, thanks to explorers who traveled to far-away locations, islands, and continents. In the same way, innovators and creative thinkers are creating their own *mappa mundi* to get the big picture of the future to come. The more they work with that, the clearer the picture becomes. Their curiosity is the driving force to learning new things and acquiring new knowledge. For them, reading the future is an essential part of innovation leadership and life in general.

5 Towards Nature Smart innovation—"The Magic Well"

> Nature holds the key to our aesthetic, intellectual, cognitive and even spiritual satisfaction.
>
> —E. O. Wilson

Beyond creativity, there is Mother Nature. In his beautiful book *The Origins of Creativity* (2017), American biologist Edward O. Wilson writes: "For almost all of the 100,000 years that humanity has existed, nature was our home. In our hearts, in our deepest fears and desires, we are still adapted to it. Ten thousand years after the invention of farms, villages, and empires, our spirit still dwell in the ecological motherland of the natural world" (Wilson 2017, 129). Wilson also reminds that we cannot live for long outside this self-sustaining environment. Artificial surroundings are not our original homebase. That is evident when people enjoy walking in the woods, and its positive impact on our health and wellbeing. Said by Wilson: "There's the value of exercise, of course, but there is something else at work deep in our psyche. In our hearts, we are still in some fashion or other hunters and gatherers… This is a point I want to make: the experience of nature, to those who have learned to absorb it, is magic well" (Wilson 2017, 133–134).

This 'magic well'—nature itself—is also a great inspiration for creatives, creative artists, and innovators. By learning the rules of nature, we can also bend and break the rules in a good sense to become real, authentic creators.—"The best artists and performers seek original ways to express themselves in image, sound, and story. Originality and style are everything, measured by the degree to which the innovations attract imitation" (Wilson 2017, 39). Wilson points out that the drive for innovation can be viewed as an analog of genetic evolution: "Cultural evolution adapts our species to the inevitable and constantly changing conditions of the environment. Its innovations are the equivalent of mutations in the genome". Wilson reminds that there are successful mutations and unsuccessful ones. In the same way, only few creative and innovative ideas lead to cultural success and many fail to have an impact. Storytelling is one way to keep cultural innovations alive, even hundreds and thousands of years. Narratives are our way to pass the knowledge and wisdom forward to future generations.

Innovation, originating from the magic well of nature, can be called *Nature Smart Innovation* in the true sense. It is based on regenerative thinking supporting planetary health, it is humanity-centric focusing on the entire ecosystem, and it is original in a creative sense by showing the way back towards a nature smart world. *Future innovation ecosystems are built upon creative collectives, led by curious creatives, who are driven by creative thinking supported by humanity-centric design and nature smart thinking that focuses on regenerative design and regenerative thinking.*

In his insightful book, *The Age of Resilience* (2022), Jeremy Rifkin describes the future resilient revolution and its infrastructure, as well as the elements of resilient economy. Rifkin writes that in human or ecological communities, science clearly shows how in nature, society, and universe, "when agencies interact, they never return to where they were because the interactions themselves change the dynamic regardless of how slight they may be" (Rifkin 2022, 154). In this way, Rifkin points out that resilience never means reestablishing the status quo: "Resilience should never be thought of as a "state of being" in the world, but rather a way of acting on the world". Resilience is openness to constant change.

According to Rifkin, in the coming era, buildings will be retrofitted for energy savings and climate resilience through smart solutions and data. "Buildings in the Third Industrial Revolution will no longer be passive, walled-off private spaces but, rather, potentially actively engaged nodal entities sharing their renewable energies, energy savings,

energy storage, electric mobility, and wide range of economic and social activity with one another at the discretion of their occupants. Self-reliant, smart buildings are a critical component of the emerging resilient society" (Rifkin 2022, 171). In the future, architecture and building technologies will be the focus of urban innovation. The faster we understand the critical role of building and built environment, the speedier the development towards more sustainable local communities and cities. According to Rifkin, communities are also beginning to share their locally generated solar and wind electricity with neighboring communities. This will eventually grow into a worldwide network of sharing green electricity (Rifkin 2022, 178).

Twenty years ago, Finland was known for its technology giant Nokia, a leading mobile phone manufacturer and innovator in the world. Today, only twenty years later, the CEO of Fingrid Jukka Ruusunen says in an interview that the next breakthrough movement in Finland could be wind energy. Fingrid is one of Finland's biggest transmission system operators. Ruusunen also estimates that wind power will overtake nuclear power in 2026–2027 in Finland. Today, already 90% of all energy production is emission-free in Finland (Raeste 2023). This example clearly shows, how fast a new technology sector can replace the old ones on thriving in the future. It is said that innovation needs the right momentum. Today, it is the time for nature smart innovations. A resilient society and economy both require support through nature smart innovations. Together, they foster local communities and make the global ecosystem thrive, eventually rewilding the planet.

Today, we cannot discuss future innovation without talking about Artificial Intelligence (AI) as part of future development and its solutions. From a creative point of view, we should look at the difference between AI as the Researcher-Assistant and AI as the Maker-Creator. One way to see the difference is asking the question: Would you like to live in an artificial forest, or rather in a natural forest? One of the wisest insights about nature and its beauty is the famous *Ode to the Flower*, a monologue in praise of natural beauty by Professor Richard Feynman, Nobel Prize laureate in quantum physics. In his short statement, he describes how his artist friend said: "I as an artist can see how beautiful this [flower] is but you as a scientist take this all apart and it becomes a dull thing". Feynman replies that he didn't agree with this viewpoint. On the contrary, said Feynman: "I can appreciate the beauty of a flower. At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes".—The more we learn about the deep level of the natural world, the more we are awed. To study nature in this almost invisible level can teach us to understand and value even more the mystery of our planet and its amazing organisms and their diverse ecosystems.

In the same way, artificial intelligence AI, as a tool for research and deeper understanding of the natural world, can add on, not subtract from the beauty of the planet or the natural world. But, if we use AI only as a creator/maker and give the leading role and power to the technology itself, we will end up with the dilemma that architect Cedric Price provocatively stated already in 1966: "Technology is the answer, but what was the question?"—Instead, if we say the health and wellbeing of nature are the question, then we can say that technology could be the answer for future innovations.

Professor Emerita Helga Nowotny, the former President of the European Research Council ERC, has said that to cope with uncertanties, we should have a collective effort in producing a shared vision of the future—or even taking a stronger position as collective making of future. When the creative innovative ateliers use the advanced co-creating and co-design tools to support their efforts, explorations, and experiments, they are supporting the collective making of future. This can lead to *Nature Smart Innovations* when arts and science, creativity and knowledge, nature and technology are seen as complementary approaches rather than competing ones.

> The search for knowledge is in our genes. It was put there by our distant ancestors who spread across the world, and it's never going to be quenched.
>
> —Edward O. Wilson

References

Greenwald M (2018) A new wave of innovation hubs sweeping the world. Forbes, April 2. https://www.forbes.com/sites/michellegreenwald/2018/04/02/a-new-wave-of-innovation-hubs-sweeping-the-world/?sh=227f12981265

Hamilton J (2023) Building a better brain through music, dance and poetry. National Public Radio, NPR, April 3. https://www.npr.org/sections/health-shots/2023/04/03/1167494088/your-brain-on-art-music-dance-poetry?utm_campaign=storyshare&utm_source=twitter.com&utm_medium=social

Magsamen S, Ross I (2023) Your brain on art; How the arts transform us. Canongate

McKinsey (2023) Building innovation ecosystems: accelerating tech hub growth. McKinsey, February 28. https://www.mckinsey.com/industries/public-sector/our-insights/building-innovation-ecosystems-accelerating-tech-hub-growth

NASA (2023) Nasa's study on children: how traditional schooling reduces creative spark. YourStory, November 16. https://yourstory.com/2023/10/nasa-study-creative-genius-educational-impact

Norman D (2023) Design for a better world; Meaningful, sustainable, humanity centered. The MIT Press

Raeste J-P (2023) Fingridin Ruusunen ei enää epäile tuuli- ja aurinkovoimaa: "Koko maan teollisuus pultataan siihen". Helsingin Sanomat, HS, June 5. https://www.hs.fi/talous/art-2000009611362.html

Rifkin J (2022) The age of resilience; Reimagining existence on a rewilding earth. St. Martin's Press
TATE (2023) Atelier. Tate, June 2023. https://www.tate.org.uk/art/art-terms/a/atelier
Visser W (2022) THRIVING; The breakthrough movement to regenerate nature, society, and the economy. Fast Company Press
Weale S (2022) Adventurous play boosts children's mental health, study finds. The Guardian, May 20. https://www.theguardian.com/society/2022/may/20/adventurous-play-boosts-childrens-mental-health-study-finds
WEF—World Economic Forum (2023) Future of jobs report. Weforum. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf
Wilson EO (2017) The origins of creativity. W. W. Norton & Company