DOCTORAL THESIS

Data-Driven Urban Modelling: The Case of Explainable Detection of Urban Heat Island

Nasim Eslamirad

Data-Driven Urban Modelling: The Case of Explainable Detection of Urban Heat Island

NASIM ESLAMIRAD

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Engineering
Department of Civil Engineering and Architecture
This dissertation was accepted for the defence of the degree 13/10/2023

Supervisor: Dr. Francesco De Luca
                            School of Engineering
                            Tallinn University of Technology
                            Tallinn, Estonia

Co-supervisor: Prof. Kimmo Sakari Lylykangas
                            School of Engineering
                            Tallinn University of Technology
                            Tallinn, Estonia

Opponents: Prof. Amira Mouakher
                            Laboratoire Espace-Dev
                            Université de Perpignan Via Domitia
                            Perpignan, France

                            Dr. Berk Ekici
                            Department of Architecture
                            Izmir Institute of Technology
                            Izmir, Turkey

Defence of the thesis: 15/12/2023, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.
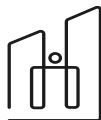
Nasim Eslamirad

Nasim Eslamirad

Signature

Andmepõhine linnamodelleerimine: soojussaarte seletatav tuvastamine

NASIM ESLAMIRAD

TAL
TECH
KIRJASTUS

Contents

List of publications

The thesis is built upon a series of five key publications authored by the researcher. In Paper I, a geoprocessing study forms the foundation, resulting in the creation of a comprehensive dataset encompassing Urban Heat Islands as well as various building and neighbor-related features. Papers II and III delve into the intricacies of machine learning processes, likely outlining the methodologies and algorithms used for the analysis data. Papers IV and V take a deeper dive into the realms of urban heat wave dynamics and the assessment of thermal comfort within urban areas. This structured approach allows for a systematic progression from data preparation and ML analysis to the exploration of urban heat-related phenomena and thermal comfort assessments, creating a cohesive body of work within the thesis.

Paper 1. Eslamirad, N., De Luca, F., Sakari Lylykangas, K., Ben Yahia, S. & Rasoulinezhad, M. (2023). Geoprocess of geospatial urban data in Tallinn, Estonia, *Data in Brief, 48*, 109172. ISSN 2352-3409, https://doi.org/10.1016/j.dib.2023.109172

Paper 2. Eslamirad, N., De Luca, F., Sakari Lylykangas, K. & Ben Yahia, S. (2023). Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment. *Frontiers of Architectural Research, 12*, 541-555. ISSN 2095-2635, https://doi.org/10.1016/j.foar.2022.12.001

Paper 3. Eslamirad, N., De Luca, F., Ben Yahia, S. & Sakari Lylykangas, K. (2023). From near real-time urban data to an Explainable city-scale model to help reduce the Urban Heat Island (UHI) effect. *Journal of Physics: Conference Series*. CISBAT 2023. In print.

Paper 4. Eslamirad, N., Sepúlveda, A., De Luca, F., Sakari Lylykangas & Ben Yahia, S. (2023). Outdoor thermal comfort optimization in a cold climate to mitigate the level of urban heat island in an urban area. *Energies, 16*, 4546. https://doi.org/10.3390/en16124546

Paper 5. Eslamirad, N., Sepúlveda, A., De Luca, F. & Sakari Lylykangas, K. (2022). Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus. *Energies, 15*, 1577. https://doi.org/10.3390/en15041577

Author's contribution to the publications

The contribution to the papers in this thesis is as follows:

In Paper, I, the author of this thesis, served as the first and corresponding author. She developed the study's original idea, conceptualization, methodology, and framework. Additionally, she conducted data collection and programming tasks and wrote the entire manuscript under the guidance and supervision of the co-authors.

In Paper II, the author of this thesis also served as the first and corresponding author. She defined the main idea and objectives of the Paper and developed the study's methodology. The author also conducted conceptualization, programming, modelling, simulation, development of the machine learning model, and visualization tasks. Additionally, she wrote the original draft of the Paper and the entire manuscript under the guidance and supervision of the co-authors.

Paper III has the author of this thesis as the first and corresponding author. She was responsible for developing the original idea and methodology of the study, as well as creating the machine learning model and completing all programming-related tasks. Additionally, she analysed the study's results and outcomes and wrote the entire Paper under the supervision of the co-authors.

The author of this thesis served as the first author and corresponding author in Paper IV. She developed the original idea and conceptualization of the Paper. The co-authors contributed to the case study selection and methodology. She defined the study scenarios and conducted modelling, simulation, and outdoor thermal comfort assessment tasks. She collected and analysed the results, synthesized the outcomes, and wrote the entire manuscript with the second author's contribution under the supervision of other authors.

In Paper V, the author of this thesis assumed the roles of both the first and corresponding author. She played a central role in originating and conceptualizing the study's idea, and also in defining the methodology. Furthermore, she collaborated closely with the second author to shape the research methodology, which involved a combination of qualitative and quantitative approaches. Additionally, she took the lead in composing the study, overseeing all aspects related to the quantitative approach, including data collection and analysis, and contributed significantly to the overall development of the study's findings. Under the guidance of the co-authors, she also bore the responsibility of crafting the initial draft and the final manuscript.

1 Introduction

Urbanization has been a critical driver of development (Bertinelli & Black, 2004), and currently, the world is experiencing a significant wave of urbanization (Erell, Pearlmutter, & Williamson, 2012) Projections suggest that by 2030, almost five billion people, equivalent to 61 percent of the global population, will reside in urban areas (Erell, Pearlmutter, & Williamson, 2012). The world's urban population is expected to increase by an additional 2.5 billion people by 2050 (Akbari, et al., 2016). This growth will likely be distributed among metropolitan areas of all sizes, including small market towns and administrative centers (Cohen B. , 2004). In Europe, nearly 73 percent of the population lives in cities, which is expected to reach 82 percent (Akbari, et al., 2016). The rapid growth of urbanization has led to several interconnected problems, such as the degradation of the thermal environment, land uptake, and climate change, which require integrated solutions (Akbari, et al., 2016). Access to a healthy and comfortable shelter is critical to physical, psychological, and social health. Therefore, it is crucial to implement sustainable urban development strategies that aim to enhance the quality of urban planning and design, reduce the negative impact of cities on the environment, and mitigate the effects of rapid urbanization and climate change on cities.

Urban design is essential for creating a sustainable built environment and improving the quality of life for city dwellers by controlling factors that affect the urban microclimate (Erell, Pearlmutter, & Williamson, 2012). Conversely, poor urban design can exacerbate the impacts of climate change in cities (Akbari, et al., 2016). One of the worldwide phenomena affected by urbanization is the urban heat island effect, which significantly impacts the estimation of global warming (Nakata-Osaki, Lucas Souza, & Souto Rodrigues, 2018).

The urban heat island effect refers to the difference in air temperature between urban and rural areas (Ul Moazzam, Hoi Doh, & Gul Lee, 2022), and it can have severe implications for the economic and social systems of cities (Akbari, et al., 2016), as well as the quality of life and health of their citizens (Aflaki, et al., 2017) (Rizwan, Dennis, & Chunho, 2008) (Vásquez-Álvarez, Flores-Vázquez, Cobos-Torres, & Cobos-Mora, 2022). At the same time, the urban heat island effect directly affects a severe impact on the economic and social system of cities (Akbari, et al., 2016), the quality of life of citizens, health problems for urban dwellers, as well as health, well-being, human comfort, and the local atmosphere (Aflaki, et al., 2017) (Rizwan, Dennis, & Chunho, 2008) (Vásquez-Álvarez, Flores-Vázquez, Cobos-Torres, & Cobos-Mora, 2022). Therefore, there is increasing awareness among scientists, planning authorities, and governmental bodies about the effects of urban design and planning on the intensity of the urban heat island effect during the summer (Aleksandrowicz, Vuckovic, & Kristina, 2017). Many policymakers and governmental sectors are implementing practical solutions to cool cities and improve them to be more sustainable, resilient, and liveable environments (Eslamirad, De Luca, & Sakari Lylykangas, The role of building morphology on pedestrian level comfort in Northern climate, 2021) (Aflaki, et al., 2017).

Offering sustainable solutions to urban microclimatic issues is crucial and challenging, requiring the adoption of a holistic approach that considers urban development, features of buildings and their contents, and characteristics of built environments.

This academic thesis aims to create a comprehensive design framework known as the CLEAR model. This innovative approach integrates data-driven and model-implemented

workflows to gain insights into the causes of urban heat island and heat waves in urban areas.

The ultimate goal is to propose effective solutions that mitigate these phenomena, fostering a better living environment while aligning with sustainability objectives in urban development.

The approach aims to address the challenges faced in various urban areas and promote sustainable urban development strategies to benefit residents and the environment.

The CLEAR model is an innovative and valuable tool to integrate urban data, leverage the model and mitigate the urban heat island effect in Tallinn.

The word of 'CLEAR' refers to:

'**C**': **Capture** data

'**L**': **Learn** from data

'**E**': **Extract** models

'**A**': **Apply** the model

'**R**': **Repeat** the model

The CLEAR model tackles the limitations of existing methods by considering the built environment as a complex adaptive system. It follows a series of iterative steps:

- Gathering and geoprocessing urban geospatial data, urban heat island effect data, and Tallinn, Estonia meteorological data.
- Employing machine learning techniques to process the collected data and create predictive and explainable machine learning models. These models estimate the urban heat island level in different parts of the city and highlight the significance of urban features in the occurrence of the urban heat island effect.
- Developing machine learning techniques to leverage urban data for identifying and implementing effective strategies that mitigate microclimatic urban issues.
- Creating a transparent machine learning model that predicts the urban heat island effect and generates a prioritized list of urban features, which represent both thermal and non-thermal aspects of urban environments. This model is intended to be a valuable tool for designers, researchers, and urban planners, as it enables them to assess the likelihood of urban heat island occurrences by considering various factors such as building characteristics, environmental conditions, and weather data. It succinctly conveys the purpose and potential utility of the machine learning model within the context of urban planning and environmental research.
- Proposing mitigation strategies based on machine learning models' outcomes that integrate urban, meteorological, and urban heat island data involves conducting numerical modelling simulations in urban areas. The machine learning models are analysed in a way that aligns with human perception and thermal comfort, facilitating the development of strategies to mitigate the urban heat island effect and contribute to sustainable urban development.

The accomplishments of the thesis research underscore the importance of the CLEAR model as a groundbreaking and valuable tool in addressing the urban heat island effect in Tallinn. The methodology's versatility allows it to be applied to other locations with varying climatic conditions and urban data, making it a valuable asset for tackling urban heat island challenges in diverse urban settings.

The study has created a comprehensive dataset that includes detailed information on building features and their neighboring areas, encompassing data related to specific areas of the city that have been exposed to urban heat island or have the potential to experience it.

Furthermore, the research has developed various models, such as predictive and explainable machine learning based models, to understand the relationship between urban features, weather data, urban heat island phenomena, and heat waves in the city. These models have provided valuable insights for improving city planning and replanning processes by proposing effective mitigation strategies and solutions aligned with sustainable built environment goals. The findings highlight the potential impact of the CLEAR model in addressing the challenges associated with urban heat island effects and heat waves.

By utilizing the comprehensive dataset and the developed models, urban planners, policymakers, and stakeholders can make informed decisions and implement strategies that promote sustainable urban development. This approach improves the quality of life for residents, enhances the comfortability of urban areas, and contributes to creating more resilient cities.

Architects and urban designers can utilize data-driven insights into significant features impacting the urban heat island effect to make informed decisions about urban design and development. The research has also deepened the understanding of the unique characteristics and challenges specific to Tallinn's urban areas, facilitating the development of sustainable urban development strategies based on evaluating the value and impact of different features on the urban heat island effect.

Overall, the thesis research has significantly advanced understanding the relationship between urban features, weather data, and urban heat island phenomena, providing practical tools and insights that can be applied to real-world scenarios.

The outcomes of this research have the potential to benefit cities and communities worldwide by addressing urban heat island effects, promoting sustainable urban development, and creating more comfortable and livable urban environments.

Research questions and objectives

Research questions

Q1: Which methodologies can be utilized to capture, calculate, and process urban data from diverse sources, such as meteorological and climatic data, urban heat island data, and information about buildings and their surroundings? Furthermore, how can this data be effectively integrated to gain a comprehensive understanding of the urban heat island effect?

Q2: How can machine learning facilitate learning and extract correlations between features that cause the urban heat island effect?

Q3: How can predictive and explainable machine learning models be used to comprehend the impact of critical urban features on the urban heat island effect?

Q4: How can the findings and achievements of machine learning and simulation approaches can be applied to develop mitigation strategies to decrease the urban heat island effect in the urban area and improve sustainable urban development guidelines?

Research objectives

O1: To propose a method for capturing urban spatial data that can provide a comprehensive understanding of the urban environment, considering location and context. The method aims to incorporate essential features related to buildings, their surroundings, and meteorological and climatic data, specifically focusing on the urban heat island effect (Paper I).

O2: To build a Predictive machine learning model that utilizes the captured data and learning algorithm to predict the target (Papers II).

O3: To build a transparent machine learning model that helps comprehend urban features' importance, roles, and values in the urban heat island effect (Paper III).

O4: To provide valuable insights into the design and configuration of urban environments to decrease the urban heat island effect and improve the comfort level in the urban area. This is achieved by generating strategies and solutions to mitigate the urban heat island effect using the captured data (Papers IV and V).

The methods, approaches, and tools of the thesis

The study employed several methods, approaches, and tools to achieve its objectives. The Python 3 programming language was utilized through the Jupyter notebook interface. The packages that were used are: Python libraries such as NumPy and Pandas for data processing, Matplotlib and Seaborn for visualization, and Geopandas for geodata processing and spatial data manipulation were used.

The study utilized Scikit-learn and transparent machine learning models to build the Predictive and explainable models.

In addition, the integration of geodata was accomplished using the QGIS Tool, while geometric modelling, simulation, and numerical analysis were carried out within the ENVI-met environment. Excel was employed for dataset processing and visualization through charts, and figures and diagrams were designed using PowerPoint.

Thesis outcomes

The thesis includes a comprehensive overview of the approval of associated works and the results of the various papers.

Paper I presents a proposed workflow for collecting geospatial data on buildings, their context, and the urban heat island phenomenon in Tallinn, Estonia. The study employs an ascending hierarchical grid system to capture heterogeneous dynamic urban data in a homogeneous static ground. The QGIS Tool is used to perform the hierarchical grid system and capture location-based urban data, which is then referenced to underlying grids for defining urban indices related to both static and dynamic data. The Geopandas package in the Python environment adds additional features to the built environment and buildings. Meteorological data from the summer of 2014-2019 is incorporated into the dataset, assigning urban heat island effect values to each element accordingly.

In Paper II, a predictive and data-generative model is developed based on the ML approach to assess the comfort level of pedestrians on the sidewalks surrounding the buildings in Tallinn. The machine learning approach uses urban data, building characteristics, weather data, and thermal and wind comfort. The results of simulation in the ENVI-met environment for thermal and wind comfort are used together to define the dataset used

in the machine learning process. The machine learning pipeline based on the Random Forest classifier model is performed in the Python environment, and the model generates new trustworthy results about outdoor comfort at the pedestrian level in Tallinn, Estonia.

Paper III combines machine learning approaches to building predictive and explainable ML models. The study aims to develop a framework to employ machine learning based models using geo-processed urban data to build a dataset that is used to build an ML-based transparent model that explains the importance of features impacting the urban heat island effect in urban areas. The outcomes indicate that the transparent machine learning model results show that the feature importance in the classified model is trustable.

In Paper IV, a multi-objective optimization-based workflow is proposed to improve the quality of the built environment in the urban area of Tallinn. The study uses captured geodata and the urban heat island effect of Tallinn to identify critical residential buildings that experience a high urban heat island effect. The simulations in an ENVI-met environment in different orientations of the urban canyon are performed to optimize the surface temperature and outdoor thermal comfort in the urban area of the case studies. The study results can be used to propose mitigation solutions to enhance thermal comfort, create suitable conditions for thermal comfort levels, and leverage the urban heat island level in cities.

Paper V introduces an innovative research framework for enhancing thermal comfort in urban areas, with a particular focus on the relationship between microclimatic conditions, human thermal sensation, and preferences. It utilizes simulations within an ENVI-met environment to conduct quantitative assessments of outdoor thermal comfort and gather subjective thermal sensation ratings. The study's findings enable the identification of areas with varying levels of thermal comfort, thereby contributing to urban sustainability goals.

The theoretical and practical novelty of the thesis

The present work explores the novel application of data-driven approaches and techniques in capturing and processing urban data, meteorological data, and urban heat island data in the context of Tallinn, Estonia. The study utilizes geoprocessing, programming, modelling, simulation, and analysis techniques to develop a comprehensive data-driven approach with the aim of promoting sustainable built environments.

The novelty of the study lies in multiple aspects. Firstly, the geoprocessed dataset, encompassing building samples, location information, building and urban characteristics, urban heat island data, and climate data, provides a valuable resource that can be applied in future studies and research endeavours. This dataset provides detailed insights into the urban environment and holds significance across multiple research fields, encompassing environmental studies, urban sustainability, and studies related to the built environment.

Secondly, the developed machine learning models offer a unique approach to proposing solutions and mitigation strategies for reducing the occurrence of heat waves and urban heat island phenomena in urban areas. These models enable architects, urban planners, policymakers, and local authorities to monitor urban data, investigate relationships between data, and enhance the quality of future development projects. The machine learning models also aid in identifying patterns and relationships between urban features and microclimate conditions, informing strategies and policies for mitigating the urban heat island effect on human health and the environment.

Furthermore, in conjunction with machine learning based models, the CLEAR model provides a comprehensive framework for evaluating the effectiveness of different strategies for mitigating the urban heat island effect. By simulating microclimate conditions, considering the impact of urban features, and analyzing the results of machine learning models, the CLEAR model informs decision-making processes to create more resilient and sustainable urban environments.

Overall, this study's methodology and framework have significant implications for comprehending the urban heat island effect and heat waves in cities located in colder regions and for developing effective strategies to mitigate their impact on urban environments and human health. The combination of data-driven approaches, geoprocessing techniques, and machine learning models contributes to advancing the understanding and management of the urban heat island effect, ultimately fostering sustainable and livable cities.

Limitations of the work

While this thesis presents substantial contributions, it is imperative to acknowledge the limitations encountered throughout the research process. Despite yielding promising results, the study does recognize certain constraints that warrant consideration.

A notable limitation pertains to the compilation of urban data, which encountered hurdles in terms of data source availability and completeness. Data quality emerged as a concern, impacting the precision and dependability of the findings. Furthermore, certain features necessitated on-site inspections, introducing both time and cost implications. Additionally, specific data sources with a higher level of detail were indispensable to ensure the approach's efficacy. The study's efficacy was profoundly influenced by the distinctive attributes inherent in the dataset.

Challenges in data acquisition arise from the necessity for more specific and detailed data sources, as well as concerns surrounding data accuracy and quality. Another limitation relates to the data-driven approach and utilization of specific machine learning models. While the selected models were suitable for the study, there exists a need to delve deeper into exploring alternative models to enhance prediction accuracy and reliability. It is noteworthy that the accuracy of machine learning models directly hinges on the quality and reliability of the data utilized for training and testing. Overfitting poses another challenge when models become overly intricate and closely tailored to training data, thereby leading to inadequate generalization to new data. Likewise, striking a balance between interpretability and predictive performance holds significance, as excessive simplification may compromise accuracy. The success of the approach hinges on the distinctive attributes of the dataset, necessitating the exploration of alternative models to effectively address these limitations.

Consequently, while the CLEAR model serves as a valuable methodology with the capability of being updated with new data and helping researchers and designers better understand urban features to mitigate phenomena like urban heatwaves and the urban heat island effect, it necessitates supplementation with other approaches, such as policy interventions, to encompass a more comprehensive and holistic strategy for mitigating climate-related concerns in cities and advancing sustainable urban development.

In spite of these limitations, the findings and methodologies presented in this study contribute significantly to a broader comprehension of the urban heat island effect and offer insights into potential strategies for mitigating its impact.

Abbreviations

AT	Air Temperature
CFD	Computational Fluid Dynamic
CRS	Coordinate Reference System
CSV	Comma-separated values
DL	Deep Learning
DT	Decision Tree
E-W	East-West
GIS	Geographic Information System
GLMs	Generalized Linear Models
HSR	Hierarchical Spatial Reasoning
KNN	K-Nearest Neighbours
LIME	Local Interpretable Model-agnostic Explanations
LIR	Linear Regression
LR	Logistic Regression
LST	Land Surface Temperature
MEMI	Munich Energy-balance Model for Individuals
ML	Machine Learning
NE-SW	North East-South West
NN	Neural Network
N-S	North-South
NW-SE	North West-South East
OTCA	Outdoor Thermal Comfort Autonomy
PDP	Partial Dependence Plot
PET	Physiological Equivalent Temperature
RF	Random Forest Classifier
RH	Relative Humidity
SE-NW	South East- North West
SHAP	SHapley Additive ExplanatIon
SL	Supervised Learning
SML	Supervised Machine Learning
S-N	South- North
ST	Surface Temperature
STD	Standard Deviation
SUHI	Surface Urban Heat Island
SVC	Support Vector Classifier
SVF	Sky View Factor
SVM	Support Vector Machines
SW-NE	South West- North East
UHI	Urban Heat Island

USL	Unsupervised Learning
WD	Wind Direction
Wm	Weighted mean
Wm-PET	Weighted mean of PET
WS	Wind Speed
Ws	Weighted score
Ws-PET	Weighted score of PET

2 Background

2.1 Urban heat island and surface temperatures in urban areas

Urban areas tend to experience higher temperatures than non-urban areas (Rosenzweig). Even urban areas can be 2–5°C hotter than their surrounding rural areas (Aflaki, et al., 2017) (Onishi, Cao, Ito, Shi, & Imura). Even it is projected that they will continue to experience warming throughout the 21st century due to global climate change and urban development (Rosenzweig). Rapid urban growth and the increase in impervious surfaces in dense urban areas can have significant impacts on the ST of cities (Erell, Pearlmutter, & Williamson, 2012) can have significant impacts on the ST of metropolitan cities (Ul Moazzam, Hoi Doh, & Gul Lee, 2022).

In 2012, Brian Stone declared that urban areas are experiencing a more significant increase in temperature compared to the overall planet, indicating the presence of global warming in cities (Ching, 2013). Urban planners are showing growing concern regarding the influence of climate on urban planning (Chirag & Ramachandraiah, 2010). They continuously work towards enhancing the well-being of urban residents by establishing a comfortable and enjoyable atmosphere (Luo, He, & Ni, 2017).

High heat emissions can influence the surface energy balance and emit toxic gases that can cause environmental problems, such as air and water pollution, and impact the Surface Urban Heat Island (SUHI) (Ul Moazzam, Hoi Doh, & Gul Lee, 2022). Remote sensing techniques are widely used to monitor Urban Heat Island (UHI) and understand its features and urban climatology, and Surface Temperature (ST) helps define the SUHI (Leiqiu & Brunsell, 2013). The SUHI effect can be observed due to the loss of natural landscapes to built-up land (urban areas), making the urban area warmer than the rural area. Thermal imagery has shown that urban centers are hotter than natural landscapes (Ul Moazzam, Hoi Doh, & Gul Lee, 2022).

The occurrence and intensity of UHI can be classified into two primary categories, as Rajagopalan, Lim, and Jamei claimed: meteorological factors and urban design factors. Meteorological factors, including Wind Speed (WS), Wind Direction (WD), humidity, and cloud cover, can impact the transfer of heat energy between surfaces and the atmosphere, thereby affecting the intensity and spatial distribution of heat islands. Conversely, urban design factors pertain to the physical features of the built environment, including building density, aspect ratio, sky view factor (SVF), and construction materials (Eslamirad, De Luca, & Sakari Lylykangas, The role of building morphology on pedestrian level comfort in Northern climate, 2021).

Heat waves and the UHI effect are significant climate risks affecting cities, and various studies have been conducted to research and classify different methods to reduce their impact (Sagris & Sepp, 2017). The UHI effect and heat waves are critical climate risks for cities, and many studies have explored and categorized various strategies to mitigate their impact (Akbari, et al., 2016). Therefore, it is crucial to explore sustainable UHI mitigation solutions (Aflaki, et al., 2017). Mitigation aims to modify the urban microclimate through physical environmental changes (Aleksandrowicz, Vuckovic, & Kristina, 2017).

The concept of "mitigation" and its application to UHI can be attributed to the research of Hashem Akbari, Arthur Rosenfeld, and Haider Taha at the Lawrence Berkeley National Laboratory (Aleksandrowicz, Vuckovic, & Kristina, 2017). Akbari identified various mitigation strategies to reduce UHI and improve thermal comfort for people living in cities (Akbari, et al., 2016). The implementation of many mitigation measures on a large

scale can significantly impact the urban microclimate as a whole, and many studies have examined mitigation strategies. In 2009, Giguère provided an extensive inventory of UHI mitigation strategies, categorizing them into four groups: vegetation and cooling, sustainable urban infrastructure, sustainable water management, and reduction of anthropogenic heat. While the list is comprehensive, scientific evidence supporting the effectiveness of some measures in reducing UHI intensity is lacking. Additionally, there is partial overlap between categories, particularly concerning trees and vegetation (Aleksandrowicz, Vuckovic, & Kristina, 2017).

Some UHI mitigation strategies proposed by Giguère's comprehensive list lack definite scientific proof for effectively reducing UHI intensity. Additionally, there is some overlap between Giguère's categories, particularly regarding trees and vegetation (Aleksandrowicz, Vuckovic, & Kristina, 2017). In Toronto, a study evaluated various UHI mitigation strategies in urban neighbourhoods using numerical simulations with ENVI-met software. The study found that the duration of direct sunlight and mean radiant temperature, influenced by the urban form, are crucial factors in determining urban thermal comfort (Wang, Berardi, & Akbari, 2016).

Another study analysed the impact of various urban design factors on the distribution of ambient and ST in open spaces, using the Sydney metropolitan area as a case study. The results showed a strong correlation between the gradient of temperature decrease along the precinct axis and the average aspect ratio of the precincts, with and without mitigation. The study suggests that urban design interventions that modify the aspect ratio of buildings and streets may effectively mitigate UHI effects and improve thermal comfort in open spaces (Kolokotsa, et al., 2022). Furthermore, a study by Xu et al. investigated the potential of using the spatial equity of green areas in cities and LST to mitigate UHI effects. The study found that increasing the amount of urban green spaces can be beneficial in reducing urban average temperatures and mitigating UHI effects (Xu, et al., 2022).

In another study, the authors explored UHI mitigation in Mandaue, Philippines, by increasing vegetation, adding open spaces, employing green roofs, or a combination. The changes in AT, ST, and thermal comfort of the study areas were considered to understand the impact of changes in green areas and green roofs on reducing the UHI effect (Cortes, Jesfel Rejuso, Ace Santos, & Blanco, 2022). In addition, Farhadi et al. conducted research on daytime UHI mitigation strategies in Tehran, which is a city that experiences urban warming. They found that lower ST strongly correlates with the UHI effect and thermal comfort (Farhadi, Faizi, & Sanaieian, 2019). In a similar vein, Arnfield discovered that street orientation is more useful for estimating the amount of solar energy that walls absorb (Arnfield, 1990).

Van Esch et al. discussed the effects of urban elements such as street width and orientation, building parameters like roof shape and envelope on solar access to the urban canopy, and the feasibility of passive solar heating strategies in residential buildings (Esch, Looman, & Bruin-Hordijk, 2012). In the Netherlands, the evaluation of UHI effects in many small to large cities and villages has shown that most Dutch cities experience significant UHI. The 95th percentile of the UHI is well-correlated with population density (Steeneveld, Koopmans, Heusinkveld, Hove, & Holtslag, 2011). Additionally, the design of streets, the urban canyons' orientation, and the trees' presence significantly influence ST and outdoor thermal comfort (Yahia & Johansson, 2012), and thus the UHI effect. Research on the UHI phenomenon often focuses on the canopy layer and examines it at micro- and local scales, such as single-street canyons and neighbourhoods (Renganathan

& Emmanuel, 2018). The design configuration of urban areas, including optimized building and street geometry and orientation, plays a crucial role in influencing solar radiation and airflow within urban canyons (Shishegar, 2013).

By employing various approaches, it becomes evident that the discernible influence of urbanization and other land-use alterations on the long-term trajectory of global temperature is apparent (Erell, Pearlmutter, & Williamson, 2012).

Ambient temperatures, including air and ST, are vital indicators for assessing the UHI effect (Aflaki, et al., 2017). Solar radiation directly affects pedestrians' thermal comfort in urban street canyons (Shishegar, 2013). Therefore, integrating these strategies into urban development plans can create more sustainable, resilient, and liveable cities (Eslamirad, De Luca, & Sakari Lylykangas, The role of building morphology on pedestrian level comfort in Northern climate, 2021).

2.2 Urban spatial data

Urban planning relies heavily on urban data collection and spatial data analysis. To address the challenges posed by complex urban systems, advanced interdisciplinary analysis methods such as urban informatics or urban data science are necessary (Kovacs-Györi, et al., 2020). One powerful urban studies and planning tool is Geographic Information System (GIS), which can capture, store, analyze, and display spatial data and capture numerical and spatial interactions between geographic objects (Nakata-Osaki, Lucas Souza, & Souto Rodrigues, 2018).

With GIS, users can create, edit, and analyze spatial data by linking geographic features to attribute data. GIS can also perform complex spatial analysis, such as overlaying multiple data layers to identify relationships or patterns, modelling scenarios based on input parameters, and creating maps and other visualizations to communicate results (Nakata-Osaki, Lucas Souza, & Souto Rodrigues, 2018).

However, spatial data analysis presents a challenge due to the non-stationarity of relationships and processes across space, referred to as spatial non-stationarity (Brunsdon, Fotheringham, & Charlton, 1999). Local analysis emerged in the late 1960s (Fotheringham & Brunsdon, 2010) to tackle this issue, and urban analysis, which draws on multidisciplinary knowledge and skills to solve urban issues, has been defined (Páez, 2005). Additionally, urban areas have different spatial boundaries created by various departments, leading to inconsistencies. The urban basic grid concept has been introduced as a solution (Peng, 2009), but the boundaries have been created uncoordinatedly, resulting in information fragmentation. A proposed solution involves using Hierarchical Spatial Reasoning (HSR) and a GIS-based algorithm to reorganize the spatial environment and automate the delineation of boundaries (Eagleson, Escobar, & Williamson, 2002).

HSR theory is a way of breaking down complex problems into smaller ones using spatial structuring and reasoning. While HSR has been used in spatial information theory, it has yet to be widely implemented in GIS environments (Eagleson, Escobar, & Williamson, 2002). In 1997, Car introduced spatial information theory's hierarchical structuring of space and reasoning as fundamental components for defining HSR (Car, 1997). In addition, in a study conducted by Mingjun, the concept of a primary urban grid was proposed, defined as the basic urban grid, and an algorithm based on HSR was designed to help delineate the primary urban grid, taking into consideration various factors such as land use, geometry compactness, and major roads (Peng, 2009).

The proposed solution to the issues of scale and aggregation in spatially aggregated data studies involves identifying a set of zones that optimize an objective function related

to model performance rather than trying to model the effects of scale and aggregation (Openshaw, 1977). This approach has implications for spatial data analysis. As urban design projects become increasingly complex and massive, there is a growing need for efficient analysis tools and rational design methods (Xiao Wang and Yacheng Song and Peng Tang, 2020).

Another study utilized a framework of digital description for block form, focusing on block morphological complexity. This framework was put to the test in an urban design setting. The study incorporated hierarchical structure and access structure theory to evaluate the spatial form of blocks under investigation (Wang, Song, & Tang, 2020).

The other study discussed the issue of incompatible boundary systems that hinder effective data integration in GIS. It highlights that the uncoordinated manner in which individual organizations generate boundaries has limited the potential of technologies for analyzing geospatial information. The authors propose an algorithm that uses the hierarchical spatial system to structure administrative boundaries and facilitate accurate data analysis while meeting agency requirements (Eagleson, Escobar, & Williamson, 2002).

2.3   Machine learning approaches and application in urban studies

Integrating spatial science with AI and related technologies such as Machine Learning (ML) and Deep Learning (DL) has been made possible through digitalization (Balogun, Tella, Baloo, & Adebisi). Such tools and principles have become increasingly important in sustainable development (Balogun, Tella, Baloo, & Adebisi), and although AI and ML are transforming scientific disciplines, their full potential has not yet been realized (Milojevic-Dupont & Creutzig, 2021). With the urgency of climate change, there is recognition among researchers and practitioners of the need for systematic research to upscale and implement place-specific climate solutions while respecting local variation and context (Milojevic-Dupont & Creutzig, 2021).

In smart cities, ML and big data analytics can facilitate assessment by analyzing large amounts of data from air quality sensors, weather stations, and other sources to identify patterns and predict air quality levels in different parts of the city (Balogun, Tella, Baloo, & Adebisi). High granularity analyses and improvements to climate solutions are possible using the wealth of data provided by significant data sources like satellite and aerial imagery, GIS data, social media data, and surveys (Milojevic-Dupont & Creutzig, 2021). However, generating mitigation-relevant digital models of cities requires spatialized climate data, which is a crucial challenge (Milojevic-Dupont & Creutzig, 2021).

Combining ML and DL techniques with geospatial data has opened the possibility of extracting valuable insights and knowledge from large and complex datasets. This approach can help identify effective solutions tailored to specific locations and contexts, discover hidden patterns and insights that traditional methods may miss, and optimize processes and outcomes (Balogun, Tella, Baloo, & Adebisi) (Milojevic-Dupont & Creutzig, 2021) (Belle & Papantonis, 2021).

In a recent study, the authors reviewed the use of DL and meta-analysis in various studies, focusing on data sources, preparation methods, training details, and performance comparisons. The study found that DL outperforms traditional methods in terms of accuracy and can address challenges previously faced. The authors also suggest future research directions in this field (Neupane, Horanont, & Aryal, 2021).

Advancements in AI research, increased computational power, and data availability have paved the way for a new era of science focused on data-intensive methodologies (Papadakis, et al., 2022). However, despite the successful application of ML approaches,

they are still primarily black boxes (Ribeiro, S, & Guestrin, 2016). In order to improve transparency and trust in ML models, it is essential to understand and interpret the model's inner workings. This is where the idea of explainable ML models comes in (Ribeiro, S, & Guestrin, 2016). Transparency in ML refers to the ability to understand and interpret the inner workings of a model (Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, 2016), and it is often used in contrast to the concept of opacity or black box-ness (Lipton Z. C., 2016). Explanations are essential for ensuring that the predictions made by models are accurate and trustworthy. Black box ML models can be challenging to understand and explain, so the need for explainable ML models has arisen (Ribeiro, S, & Guestrin, 2016).

ML refers to various families and traditions of techniques that can handle high-level tasks such as classification, regression, or probability density estimation (Milojevic-Dupont & Creutzig, 2021). There are different families within ML, such as Supervised Learning (SL), Unsupervised Learning (USL), and reinforcement learning (learning occurs through interactions with an environment), each with their own sub-families like kernel methods or tree-based methods for SL (Milojevic-Dupont, Nikola, and Felix Creutzig, 2021).

The advantages of using ML-based approaches include their ability to handle simple to complex methods. For example, linear classifiers are sufficient for simple tasks like classifying built-up areas on satellite imagery. At the same time, generative adversarial networks can achieve complex tasks like mimicking the style of an image, such as generating a realistic image of a climate-induced flooded area from a picture taken with typical weather (Milojevic-Dupont & Creutzig, 2021).

ML tasks are embedded in the larger life cycle of a data science project, which includes data collection, problem formulation, and model maintenance when a model is used with new streams of data over time. Many resources are available on generic aspects of the data science life cycle, focusing on interpretable approaches. Practical tips on model training are also available (Milojevic-Dupont & Creutzig, 2021). While the largest neural networks (NN) require training billions of parameters and consume extreme amounts of energy, most ML methods can run within seconds to hours on personal computers. The energy impact of training models can be explicitly considered to contain the deep NN energy footprint (Milojevic-Dupont & Creutzig, 2021).

AI and ML have great potential to transform scientific disciplines (Milojevic-Dupont & Creutzig, 2021). With the increasing urgency of climate change, there is growing recognition among researchers and practitioners for systematic and well-grounded research to implement place-specific climate solutions while considering local variation and context (Milojevic-Dupont & Creutzig, 2021). In the field of urban data science, ML techniques, such as unsupervised ML algorithms and semi-supervised ML systems, are being applied to extract insights from large and complex datasets (Kovacs-Györi, et al., 2020). These techniques are significant in smart cities' air quality monitoring and assessment. ML algorithms can analyze data from various sources to identify patterns and predict air quality levels (Balogun, Tella, Baloo, & Adebisi).

Integrating spatial science with AI and ML has become possible through digitalization, enabling the extraction of valuable insights from large and complex geospatial datasets (Balogun, Tella, Baloo, & Adebisi). In the context of climate solutions, ML methods combined with big data sources offer the potential to improve climate solutions with high granularity and tailored to specific locations (Milojevic-Dupont & Creutzig, 2021).

ML models have become integral in urban studies because they can process and analyze large volumes of data accurately and efficiently (Jing & W Biljecki, 2022). While

supervised ML techniques are commonly used in urban studies, ML is also employed to develop predictive models and estimate unknown values while explaining relationships among phenomena (Jing & W Biljecki, 2022). The application of ML techniques, including linear regression (LIR), classification methods like K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and tree-based models such as Decision Trees (DT), Random Forest (RF), and XGBoost, has demonstrated their effectiveness in urban data-related research fields (Jing & W Biljecki, 2022).

In specific studies, ML models have been applied to predict land use/land cover (LULC) and seasonal Land Surface Temperature (LST). The influence of land use patterns on the UHI effect has also been investigated using ML methods. Random Forest classifier (RF) has been used to figure out how non-linearly morphological factors affect the intensity of UHI, which is helpful for reducing UHI and planning land use when there is not much green space (Lin, Qiu, Tan, & Zhuang, 2023). Furthermore, ML techniques, including RF, have been employed to enhance the land use regression model for obtaining spatial distributions of ozone and the UHI, with the interpretation of feature variable impact using SHapley Additive Explanation (SHAP) and Partial Dependence Plot (PDP) techniques (Han, Zhao, Gao, & Gu, 2022).

In another study, authors (Balogun, Tella, Baloo, & Adebisi) used ML-based temperature spatial downscaling to look at the effect of UHI on the atmosphere. They did this by making a regression model that shows how urban structure affects temperature.

Considering the prediction of the UHI effect, the integration of ML models prioritizing interpretability is gaining attention. The review of existing literature highlights two main approaches: white-box and black-box methods, with the former characterized by easily interpretable working processes and the latter involving less transparent procedures (Lipton Z. C., 2016).

In summary, AI and ML techniques, including black-box and transparent methods, are increasingly being applied to address climate-related challenges in urban environments. ML models enable the analysis of large and complex datasets, offering valuable insights for climate solutions, air quality monitoring, land use prediction, and understanding the UHI effect. Integrating spatial science, AI, and ML opens new possibilities for extracting knowledge and enhancing sustainable development.

2.4 Outdoor thermal comfort optimization in urban areas

The potential for reducing outdoor Air Temperature (AT) in a square in Rome was studied by using a numerical model created using the ENVI met tool to simulate different mitigation scenarios to reduce warming in urban areas. The study found solutions like using grass pavers to provide the most significant advantages that could enhance the thermal conditions of the air and reduce outdoor AT (Gabriele, Ocłoń, & De Lieto Vollaro, 2023).

Fazia Ali-Toudert et al. discuss the role of street design, like aspect ratio and orientation, in developing pedestrian-level comfortability. The study benefits from the three-dimensional numerical model ENVI-met, which simulates microclimatic changes within urban environments with high spatial and temporal resolution in Ghardaia, Algeria. The study analyzed the symmetrical urban canyons with various height-to-width ratios and different solar orientations (i.e., East-West (E–W), North-South (N–S), North East-South West (NE–SW), and North West-South East (NW–SE). In addition, the study assessed outdoor thermal comfort values in the physiologically equivalent temperature

(PET) index. The results show contrasting patterns of thermal comfort between shallow and deep urban streets and the various orientations.

Moreover, the results prove that PET at the street level depends strongly on aspect ratio and street orientation (Ali-Toudert & Mayer, 2006). Another study evaluated the potential for UHI mitigation by greening parking lots and the relationships between LST and LULC in different seasons in Nagoya. The results show that different LULC types play different roles in different seasons and times, and using more green areas slightly reduced the LST for the whole study area in spring or summer (Onishi, Cao, Ito, Shi, & Imura). The other study that used qualitative and quantitative approaches to assess outdoor thermal comfort as a mixed method identified which urban areas needed more improvement during the summer. The results of thermal comfort assessment through the PET index and subjectively perceived thermal sensation using ENVI-met environment to do Computational Fluid Dynamic (CFD) simulation and thermal comfort assessment (Eslamirad, Sepúlveda, De Luca, & Sakari Lylykangas, 2022).

In the study by Giridharan, the author defines urban compactness as a combination of various urban design factors, including the building area-to-volume ratio, aspect ratio (height to width), SVF, distance to the nearest wall, width of the street, built-up area, green areas, albedo, water surface size, roads, open areas, and distance to a heat sink (Giridharan & R, 2018). These factors influence the urban microclimate, and their combination can affect the level of UHI and outdoor thermal comfort in urban areas. In addition, Aleksandrowicz et al. outline the physical features of the urban environment, such as the density of buildings, the area of land used and unoccupied areas, and the type of materials in urban components, which all affect UHI level and strength (Aleksandrowicz, Vuckovic, & Kristina, 2017).

The background studies and literature review show that heat waves and the UHI effect are significant climate risks affecting cities. There are many ways in which urban design can be modified to mitigate the UHI effect in cities, such as increasing green spaces, using reflective or high albedo materials, modifying the built environment, reducing anthropogenic heat, and optimizing building and urban canopy orientation and layout. These changes can significantly reduce ST in urban areas, which can have a negative impact on the UHI effect and the level of thermal comfort felt by those who live and work there. Furthermore, an assessment of building orientation and urban canyon extension, both of which influence solar radiation levels, plays a role in local temperature variations. This, in turn, affects subjective thermal perceptions, particularly in urban areas. This approach aids in mitigating elevated temperatures, enhancing pedestrian-level comfort, regulating thermal conditions within urban canyons, and alleviating the UHI effect on scorching summer days. Moreover, these adaptations yield additional advantages, encompassing improved urban comfort, heightened livability, and the cultivation of sustainable and resilient environments that foster the well-being of residents.

3   Methodology

3.1   The research workflow; CLEAR model

The overall flow of the thesis is closely tied to the CLEAR model, which serves as the foundation for the data-driven Model. This model encompasses four essential steps, as illustrated in Figure 1.

Step 1: Data **Capture** (**C**) – In this step, urban data is collected and aggregated to be used in the ML models for UHI analysis.

Step 2: Learning (**L**) – The captured data is utilized to develop learning algorithms and create ML models, including predictive and explainable models.

Step 3: Rule **Extraction** (**E**) – From the ML models, rules are extracted to develop a model that enables easy prediction of UHI phenomena in urban areas.

Step 4: Application (**A**) – The captured dataset and research findings are applied to propose effective mitigation strategies aimed at reducing heat waves and UHI in the urban environment.

Additionally, the research workflow includes an iterative process (R) where the model can be repeated and applied to other locations with different urban scenarios and climatic conditions. This allows for the adaptability and applicability of the CLEAR model in diverse contexts.
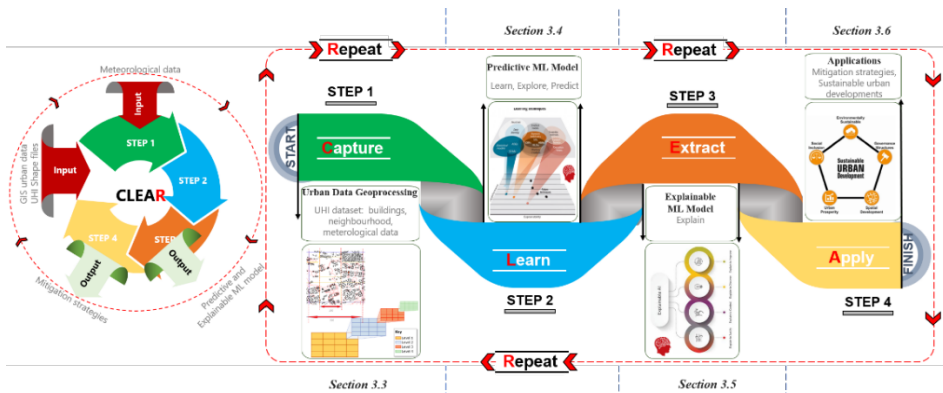

Figure 1. General structure of the thesis and defined steps.

3.2   The Location of the case study

The location of the spatial data is Tallinn, the capital city of Estonia. The latitude and longitude coordinates are 59.436962 and 24.753574 (Weatherdata, n.d.). The geographical information of Tallinn is 445,005 people and 159 km^2 of area. The number of city districts in Tallinn is 8 (GeoJournal, 2005). Moreover, Tallinn is characterized by a humid continental climate with cold winters, according to the Köppen-Geiger classification Dfb (Peel, Finlayson, & McMahon, 2007). (Figure 2)
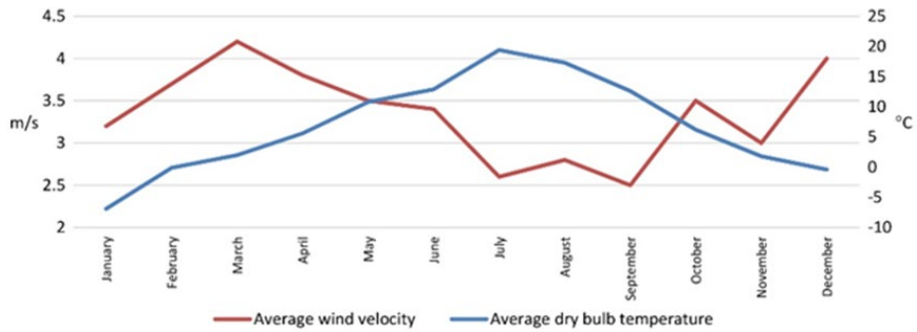
Figure 2. Monthly wind velocity and temperature averages for Tallinn, Estonia.

Furthermore, the weather data is related to the days of UHI value in the summer of 2014–2019. The weather data resource is the Underground website (Weather Underground, n.d.).
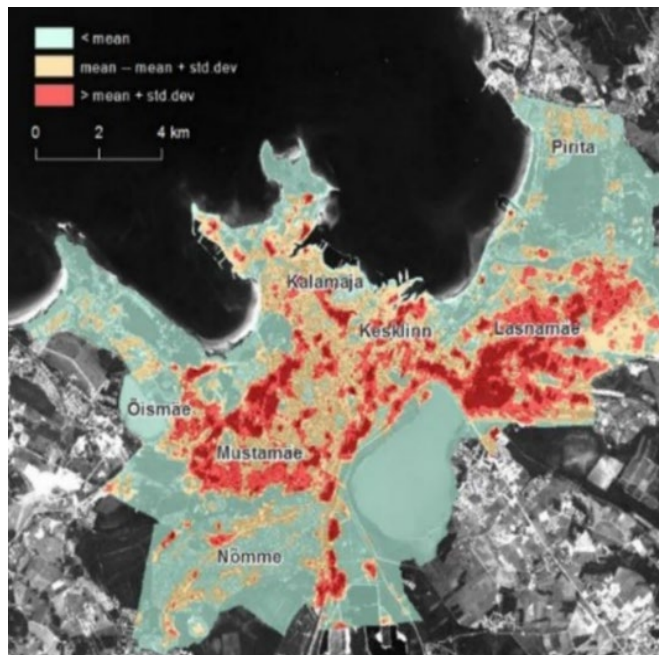

Figure 3. SUHI map of Tallinn, July 25, 2014, 12:30 Estonian summer time (9:30 GMT). The green areas are outside of the SUHI, the yellow areas are the areas over the mean SUHI, and the red areas are "inside" the SUHI (over mean + std.dev) (Sagris & Sepp, 2017).

The UHI data downloaded from the Environment Agency is based on national monitoring data collected by the Environment Agency and analyses carried out on Landsat-8 (USA) satellite data (UHI 2014–2019) (keskkonnateadlik hub, n.d.). For example, Figure 3 shows the heat map of the UHI on July 25, 2014. All UHI data used are Shapefiles in the formats SHP and SHX (Märtens, Pärg, & Sluzenikina, 2020). The data shows the UHI value in Tallinn is categorized into three levels: lower than 30°C (29°C), 30°C and 35°C.

This research concentrates on the residential buildings with the greatest volume (maximum height and area) that went through an intense heatwave and UHI in 2014,

2018, and 2019. Moreover, case studies were chosen by sampling and utilizing the histogram to locate the more critical and more severe residential buildings from the Tallinn UHI dataset (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023).

3.3   Step 1: Capture (the C' of the CLEAR' model); Data capturing

This section of the methodology, described in Paper I, focuses on creating a dataset for the Machine Learning (ML) model using urban data, meteorological data, and UHI data specifically obtained from Tallinn, Estonia. The dataset is a valuable resource that enhances the understanding of urban environments and enables professionals in various fields, including urban planning, architecture, policymaking, and stakeholder engagement, to make well-informed decisions. It is available as supplementary material and provides essential baseline information for future research and planning endeavors.

Additionally, integrating geospatial data into the QGIS tool facilitates the application of the study's findings and analysis to propose effective mitigation strategies. The analysis and modelling processes are conducted using the Python 3 environment in Jupyter Notebook, ensuring comprehensive data analysis and interpretation.

To summarize, the methodology comprises three key steps: data collection, the establishment of an ascending hierarchical grid system for data capture, the application of experimental methods to extract data, and the completion of the dataset.

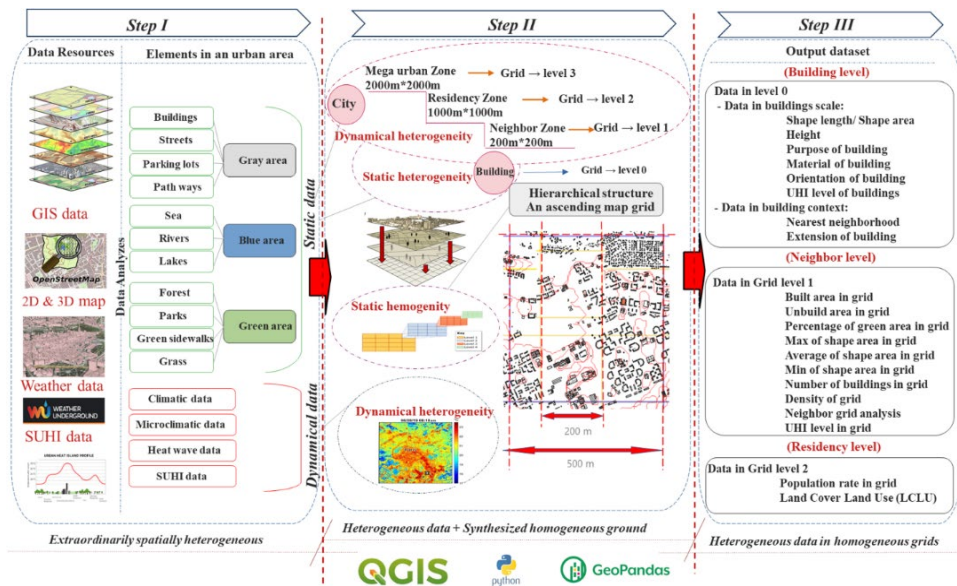Figure 4 visually represents these three main steps in the methodology.

Figure 4. Data capturing framework: structured steps for effective data collection.

The first step involves extracting urban data (Sagris & Sepp, 2017) (General data of Tallinn., n.d.), weather data (Weather Underground, n.d.), and SUHI data from Tallinn, Estonia. The SUHI data was obtained from the Environment Agency, which collected national monitoring data and analyzed Landsat-8 satellite data (SUHI 2014–2019)

(keskkonnateadlik hub, n.d.). Figure 5 displays a heat map of the UHI on July 25, 2014, illustrating the UHI values categorized into three levels: lower than 30°C (29°C), 30°C, and 35°C. The collected data, including UHI Shapefiles and other relevant data, serves as valuable input for subsequent research steps.
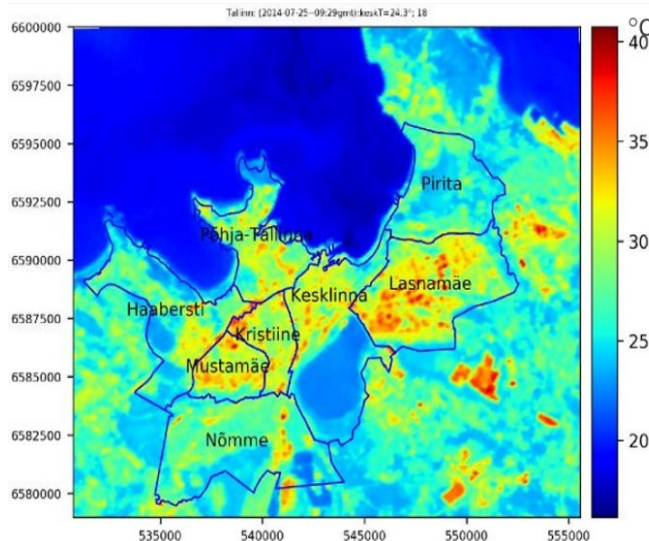

Figure 5. Heat map of UHI, Tallinn, July 25, 2014. Tallinn 25 of 2014 (Märtens, Pärg, & Sluzenikina, 2020).

3.3.1   Substep 1: Data description

The considered features of buildings are described in Tables 1 and 2. The focused features of buildings in Tallinn are general and technical characteristics. The main features of buildings are the purpose of use, material, absolute height, area, length, and number of floors.

Table 1. The description of building data based on Open data information and instructions.

Object ID	The ID of each building is a unique combination of numbers
Material	The main materials of the construction: are stone, wood, metal, composite material, and stone, composite material and metal
Absolute height (m)	The highest point of the highest structure
Height (m)	The largest vertical dimension of the building from the ground or pavement immediately surrounding the buildings to the highest point of the highest structure of the building, without taking into account local smaller depressions and elevations.
Area (m²)	All building areas are in common use by residential and non-residential users.
Length (m)	The length and shape of the building
Number of floors	The horizontal plane in a building, on which it is possible to use the building according to its purpose.

Table 2. Purpose of use of buildings.

The purpose of buildings	Description
Residential building	Buildings for dwelling purposes.
Public building	Buildings for particular public or common use (such as schools, shopping centers, banks, and offices)
Outbuilding	Buildings that are not residential not industrial facilities, and not public
Underground building	Building with no floors above ground
Industrial facilities	Building for manufacturing and production processes, or a warehouse; not public building
Underground storage space	Storage building with no floors above ground
Parking facilities	Underground garage or parking lot for cars

The data available is split into three categories: Urban, Weather and Climate, and UHI. The Estonian government and local authorities own the Tallinn Land Authority Geospatial Data Portal, which is a source of urban data. It is available as spatial data in the form of ESRI shapefiles (.SHP) and Shape Index files (.SHX). This data contains details on buildings, land, green spaces, roads, water bodies, and infrastructure in Tallinn. For example, according to the Data Portal of Tallinn (Maaamet) (Märtens, Pärg, & Sluzenikina, 2020). Tallinn has 67,113 buildings (General data of Tallinn., n.d.). Data formats are Spatial data (ESRI shapefile), Shapefile (.SHP), and Shape Index file (.SHX). The data format indicates information about the buildings, land, green spaces, roads, water bodies, and infrastructure in Tallinn. The number of buildings located in Tallinn is 67, 113 (General data of Tallinn., n.d.)

The data on SUHI was obtained through a study that used Landsat-8 images and LST data, which were suitable for analyzing the UHI effect. The study aimed to evaluate the impact of heatwaves on Estonian cities and measure the extent and magnitude of UHI effects (Märtens, Pärg, & Sluzenikina, 2020) (Sagris & Sepp, 2017).

The threshold for defining a heat wave in Estonia is an AT of over 30°C that lasts for several days, with dangerous temperatures defined as over 27°C. UHI effect data was produced for Tallinn, with an intensification of 5°C according to Landsat-8 images, and the threshold for UHI data is categorized into three ranges: 30–35°C, 35–40°C, and over 40°C. The related weather data was also included in the dataset based on the UHI effect in Tallinn (Sagris & Sepp, 2017).

The related weather data was added to the dataset based on the UHI effect in Tallinn. The geoprocessing methodology involves collecting geospatial data on buildings and UHI phenomena, integrating them into QGIS, and analyzing them in Python. The methodology consists of three steps: (1) considering the available resources of urban data and their schema, (2) categorizing data schema into homogeneous or heterogeneous, static or dynamic, and implementing the hierarchical grid system, and (3) using the homogeneous ground to define urban indices mainly related to the heterogeneous data, accommodating both static and dynamic data in the hierarchical grid system. The aim of creating the hierarchical grid system is to provide a solid basis for collecting data based on their

locations (General data of Tallinn., n.d.) (Building data of Tallinn, n.d.) (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023).

3.3.2 Substep 2: Geoprocessing and dataset development

The raw data (General data of Tallinn., n.d.) is used in the second sub-step to develop a geoprocessed dataset. This dataset incorporates multi-scale urban data, allowing urban planners, researchers, and practitioners to integrate urban data into their activities. The process involves using an ascending hierarchical grid system to collect heterogeneous data within a homogeneous ground. This approach ensures the organization and integration of data at different grid levels, from level 0 to level 3.

The final sub-step focuses on processing, arranging, and collecting the data. Python libraries and packages are employed to facilitate this process. Additionally, the geoprocessing attributes and expressions of the QGIS Tool are utilized, enabling the development of methods for data collection across different grid levels. The systematic arrangement and processing of the data ensure its accessibility and usability for further analysis and research.

By following these steps, the methodology ensures the extraction, geoprocessing, and arrangement of relevant urban, weather, and SUHI data from Tallinn. The resulting dataset provides a comprehensive foundation for integrating urban data into research and decision-making processes.

Figure 6 displays the defined ascending hierarchical grid system used to organize the spatial data in Tallinn. The hierarchical structure allows for the collection and analysis of data based on location, with grids characterized by homogeneous static data in one layer and including all recognized elements within the grid. The grids are divided into different levels, such as Neighborhood Zone, Residential Zone, and Mega Urban Zone, with the building level (level 0) being the smallest scale.
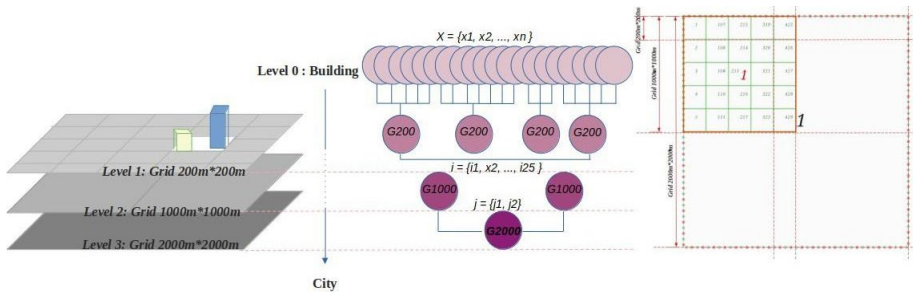

Figure 6. The defined ascending hierarchical grid system.

The available data is divided into two components: static heterogeneity data and dynamic heterogeneity data. Static heterogeneity data includes elements in an urban area that remain unchanged or change little over time, while dynamic heterogeneity data refers to frequently changing data such as climate, microclimate, and UHI data. A hierarchical grid system is created to aid in analyzing extensive spatial data by dividing the areas of the earth into identifiable grid cells. This structure provides a solid basis for collecting data based on their locations, and each grid is characterized by homogeneous static data in one layer and includes all points, geometries, and other recognized elements below the grid.

The aim of creating the hierarchical grid system is:

- First, a spatial index is created for each object on the map, and the objects are referenced to the underlying grids. The grid system aids in analyzing extensive spatial data by effectively dividing the areas of the earth into identifiable grid cells.
- Secondly, use the homogeneous ground to define urban indices mainly related to the heterogeneous data, which means accommodating both static and dynamic data in the solid hierarchical grid system.

This means that the hierarchical structure provides a solid basis for collecting data based on their locations. Thus, each grid is characterized by homogeneous static data in one layer and includes all points, geometries, and other recognized elements below the grid. The total building area located in grid level 1 is considered an example in the definition of built-up area.

This section uses a hierarchical grid system in QGIS Tool to represent polygons covering all parts of the map of Tallinn in three independent layers. The system has three ascending levels of geometric grids, starting from a 200-meter square to a 2,000-meter square, which are respectively called levels 1 to 3. This grid system aids in analyzing the spatial data of the target area on the map.

Neighbourhood Zone, Residential Zone, and Mega Urban Zone, while the base level is the building level (level 0) as the smallest scale (Figure 6).

The set X= {x_1, x_2, ..., x^n} is the set of building objects in the building dataset (level 0). First, this set is aggregated into the defined ascending hierarchical grid system (Peng, 2009). For example, x_i is the building located in the ith grid of level 1, so the ith grid is aggregated by the jth grid of level 2 and the kth grid of level 3. Then the spatial index of each building is appended to the dataset to determine the exact position of the building in the ascending hierarchical grid system. For example, figure 6 shows that each grid unit of level 3, a grid of 2,000 meters by 2,000 meters, contains four squares of level 2 grids of 1,000 meters by 1,000 meters and at the same time includes 25 squares of level 1 at a 200-meter dimension.

In the third step of the methodology, data processing, arranging, and collection are carried out using Python libraries and QGIS geoprocessing attributes and expressions. The hierarchical structure system is used to collect data related to the level of the grid system and the defined indices mentioned in Table 3.

Table 3. Defined indices using a hierarchical structure system.

Index	Grid level	Description
Building data	Grid: 200 m*200m level 1	Building data
built-up area	Grid 200 m*200m level 1	B = Sum of buildings' areas located in grid Level 1
Density	Grid: 200 m*200m level 1	D = Sum of all buildings' areas located in grid level 1/40000
Green area	Grid: 200 m*200m level 1	G = Sum of the area of green spaces, located in grid level 1

3.3.3 Substep 3: The experimental methods to capture data

This section describes the experimental methods used to capture data for the geospatial dataset.

Data acquisition I: object detection

The first step is object detection in the hierarchical grid system, which involves detecting the UHI value of each building. This is done in Grid 2, Level 1 of the hierarchical grid system. The UHI intensity levels in the main levels of SUHI are also mentioned as being lower than 30°C (29°C), 30°C, and 35°C. The dataset is then completed by adding more details about the buildings.

The QGIS Tool deals with polygons when detecting objects like buildings, and the Coordinate Reference System (CRS) is used to track objects located in the same place. The code developed in Python uses this relationship to list buildings in each UHI area with unique identification codes. The libraries used in Python are Pandas and Numpy for data import and processing, and Geopandas for object detection using GeoSeries.intersection attributes. Within the hierarchical grid system's coverage area are these layers. The codes were developed and executed using the Jupyter Notebook interface.

The method needs to use the unique code of buildings to find out the intersected geometries and return the ID of the building and the UHI map for the next step of appending other features. To do this, a function to iterate pairs is required, such as the zip () function in Python, which pairs the first element in each passed iterator and then the second element in each passed iterator.

To develop the geospatial dataset, the for-loop function is utilized to iterate over a sequence of tuples. The for loop runs a set of instructions for each element in the tuple. This loop is repeated twice to build the dataset and the UHI data with the zipped iterated objects. Using Geopandas' intersection attributes, Python's zip function, and finding the intersection between two geometries of the two datasets, the building's ID that intersects with the UHI values is obtained. The collected data includes the UHI value and the date, enabling the identification of buildings with different heat values in the city. The code to determine which building is under which grid is similar to this code.

The codes are given below:

```
for r, s in zip (Buildingdata.geometry, Buildingdata.ID): for v and o in zip (UHI.geometry, UHI.ID):
    Intersection_area = r.buffer(0).intersection(v.buffer(0)) if not intersection_area.is_empty:
        p = gpd.GeoSeries(intersection_area) print(f"{int(o)} has intersection with {s}")
```

Data acquisition II: the nearest neighbour

The distance between buildings is crucial to understanding urban density and city compactness. To determine the shortest Euclidean distance between buildings in a neighbourhood, the spatial index defined in the previous step can be used, which is faster than looping through a large data frame to find the minimum distance.

The code uses the zip () function to get the ID of each building and identify neighbouring buildings using a for-loop in Python. The Geopandas package's distance attributes are also used to find the shortest Euclidean distances between neighbouring buildings at the neighbourhood scale.

The code generates a list of distances between each building and its nearest neighbour. The list starts with 0 as the minimum distance (since the distance of each building to itself is 0), followed by the distance to the nearest neighbour. The code then uses this list to get the distance to the nearest neighbour of each building by calling the

second item in the list (since the first item is the distance to itself). Figure 7 illustrates the shortest Euclidean distance between buildings, which is used to determine the nearest neighbour to each building.

The code presented can find the nearest neighbour of a given geometry by calculating the distance between the outer wall of the building and the other geometries. It does this by iterating through each member of the building dataset and calculating the distance to all other members. If the distance between the centers of the geometries is needed, centroids can be calculated using the centroid attributes of the Geopandas package. The distance function of the Geopandas package is used to calculate the distances, which are usually defined as the smallest Euclidean or straight-line distance between two geometries. The method can be used to find the nearest street, green space, or other location as a geometry, point, or line, even in two different data resources.
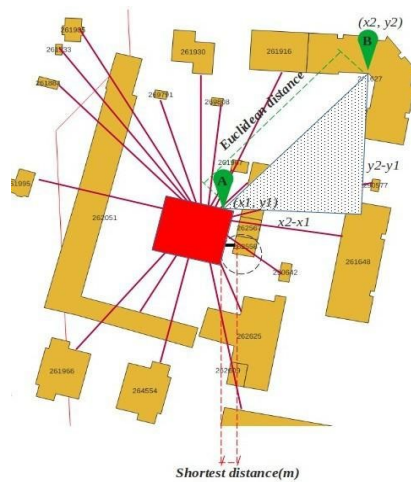

Figure 7. The nearest neighbour to each building.

Through the utilization of distance, alignment of index values within the GeoSeries (geometries) can be achieved, enabling the comparison of elements possessing identical indices via align=True. Alternatively, an index-independent comparison can be conducted, based on the sequential order of matching elements, employing align=False. (geopandas, 2022). (Figure 8)

Figure 8. The GeoSeries distance, index value True or False.

32

The code back ID of two adjacent buildings with the minimum distance in meters between them. The codes for calculating the nearest neighbour are given below:

for i, y in zip (DFB.geometry, DFB.objectid):
distance = DFB.distance(i, align = True) sorted_list=sorted(distance.to_list())
min_sorted_list=sorted_list
print (f"the nearest neighbour of {int(y)} building ID is located in min_sorted_list} meters")

In addition, if the maximum, mean, and average distances are needed, the following attributes of Geopandas in the codes can be applied:

maximum_distance = distance.max() mean_distance = distance.mean()

Data acquisition III: main angle and orientation of buildings

The orientation of the building is a factor of paramount importance that affects the incident solar radiation and the absorbed heat wave. According to Mondal, a building with an E–W orientation has maximum solar gain, and a building with an N–S orientation has minimum solar gain (Mondal, 2018). Since data collection aims to create a predictable and explainable ML UHI model, the sub-scoring of buildings oriented in the NE–SW and NW–SE is crucial because it shows how buildings are exposed to solar radiation and receive heat waves.

The orientation of a building is important for understanding its exposure to solar radiation and heat waves. The conventional axes are used to denote directions, with east being 0 degrees and angles calculated in an anticlockwise direction. The main angle of a building is located on the right side, while its orientation is the direction or angle to which its length is facing. Figure 9 shows how building direction and angle are calculated using QGIS Tools. The east direction (positive x-axis) is assumed to be 0 degrees, as shown in Figure 9. The orientation of a building is the direction or angle to which its length is facing. Since the axis of a building is parallel to its length, or conversely, perpendicular to its orientation, the main angle of the building located on the right side of the following figure is 45°. In contrast, the orientation of the building is 135° in the NE–SW direction. Considering the building on the left side of the figure, the angle is 135°, the orientation is 45°, and the direction is NW-SE.
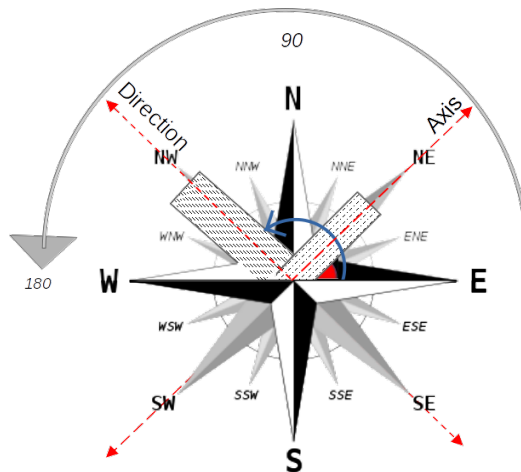

Figure 9. Direction, angle, and orientation of buildings.

The main_angle expression is used in the field calculator window of QGIS Tools to determine the main angle of all buildings in the dataset. It is applied to all geometries by calling $geometry in the expression part.

The final step in preparing the dataset involves appending other general and technical characteristics of buildings at different levels of the hierarchical grid system. Tables 1, 2, and 3 are used to collect this information. The resulting dataset includes information about the characteristics of the buildings, their spatial indices on the hierarchical grid system, the UHI value of each building, defined indexes based on buildings and different zones in the city area, and weather data.

3.4   Step 2, Learn ('L' of 'CLEAR' model); Learning from the captured data

3.4.1   Machine learning process, predictive models

This section focuses on developing ML models, particularly predictive models, to construct a learning algorithm based on ML. The goal is to apply these models in the subsequent stage to extract ML models that are transparent and explainable.

The methodology employed in this research section draws upon the concepts presented in Papers   II and III. The study is situated in urban studies and focus on the implementation paradigm of urban assessment practice. These ML models were built using a geoprocessed dataset, which forms a crucial component of Paper I.

The first step in designing an ML dataset is finding the most relevant data through pre- processing. As a result, the type of features, their classes, and their range create and define a raw tabular data set. Because the ML model algorithm requires not only precision, accuracy, and minimum error to make an accurate prediction but also works better with features with higher correlation rates, the dataset needs to be evaluated in the pre-processing step to find and omit the data that is out of range. Therefore, while the feature engineering process is performing, the raw input data was modified and cleaned of outlier data (that falls outside the range) and features with the lowest correlation. Figure 10 shows the processes of training and testing of a ML model.
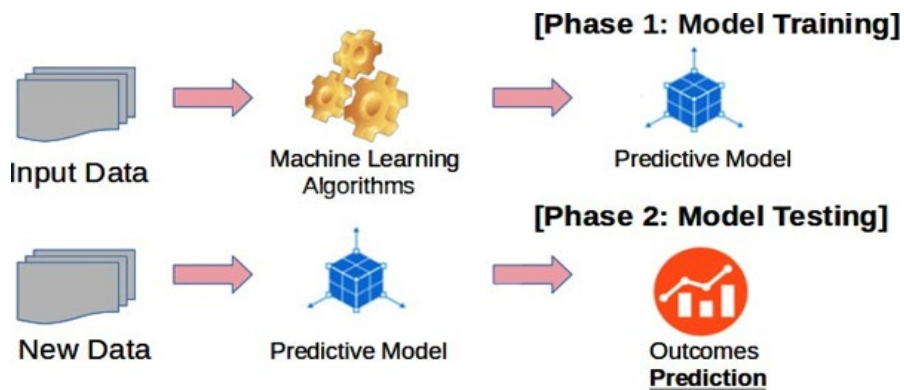

Figure 10. ML training and scoring process.

An extensive multidisciplinary dataset that contains 34,001 building samples (rows) and 30 features (columns) of Tallinn, Estonia (Eslamirad N., 2023) including the characteristics of the buildings and their neighbours in the hierarchical grid system (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023).

The attributes of the dataset, derived from the original data sources, can be summarized as follows:

- Object ID: A unique numerical identifier for each building
- Purpose of use: Categorization of buildings into Residential (for dwelling purposes), Public (for public or common use like schools, shopping centers, etc.), outbuildings (not used for residential or public purposes), Underground buildings (no floors above ground), and Industrial facilities (for manufacturing, production, or warehousing).
- Material: The primary construction material of the building, such as stone, wood, metal, composite material, or a combination of stone, composite material, and metal.
- Absolute height (m): The highest point of the tallest structure in the building.
- Height (m): The vertical dimension of the building from the ground or surrounding pavement to the highest point of the tallest structure, disregarding minor local depressions and elevations.
- Area (m^2): The total area of the building that both residential and non-residential users commonly use.
- Length (m): The measurement of the building's shape in terms of length.
- Number of floors: The count of horizontal planes within the building that can be used according to their intended purpose.

The features of the dataset are Material, Height (m), Absolut Height (m), Number of Floors Above Ground, Shape Length (m), Shape Area (m^2), the spatial indices of buildings on the hierarchical system grid, like, Built up area (G200, level1), Urban Density D1 (G200, level 1), Urban Density D2 (G1000, level 2), Urban Density D1 (G2000, level3), Average Building Area in G200 (m^2), Max Area in G200 (m^2), Number of Buildings in G200, the defined indexes based on buildings and different zones in the city area, The Nearest Neighbour (m), Green Area in G200 (m^2), The Ratio of Green Area / Grid Area (G200), Purpose of Building, Main Angle, Orientation, Height to wide (G200), the weather data, and the UHI value of each building (Eslamirad N., 2023) (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023).

Furthermore, the dataset incorporates meteorological data from the recorded UHI dates in Tallinn. Table 4 provides an overview of the meteorological data within the dataset.

Table 4. The meteorological data in the dataset (Eslamirad N., 2023).

Date	Dry Bulb Temperature (°C)		Dew Point Temperature (°C)		Wind Speed (WS) (km/h)
	Highest	Lowest	Highest	Lowest	
02-06-2018	24	12	12	5	35
05-06-2019	28	14	15	12	15
25-07-2014	29	15	17	12	13

The ML approach relies on the fundamental principle of learning algorithms derived from identified patterns, structures, and correlations among labelled features and targets. The Python environment is utilized for this purpose, enabling the prediction of outdoor comfort data within an urban context.

To initiate the ML process, the first step involves importing the data in comma-separated values (CSV) format into the Python environment. Subsequently, a data frame (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023) is created to facilitate further analysis and modelling (Eslamirad, De Luca, Sakari Lylykangas, & Ben Yahia, Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment, 2023).

The aim is employing ML based models using geoprocessed urban data to clarify how different urban features affect the UHI effect (Eslamirad, De Luca, Sakari Lylykangas, Ben Yahia, & Rasoulinezhad, Geoprocess of geospatial urban data in Tallinn, Estonia, 2023). The framework of the study comprises two sections, as Figure 11 shows.

The first section is related to the urban assessment practice of building explainable ML models based on UHI phenomena to perform geoprocessing on the data (Eslamirad N., 2023). The steps in this section to build the explainable ML model are:
1- Building the RF Classifier model
2- Finding the essential urban features (Permutation feature importance)
3- Finding the marginal contribution and threshold of the essential attributes
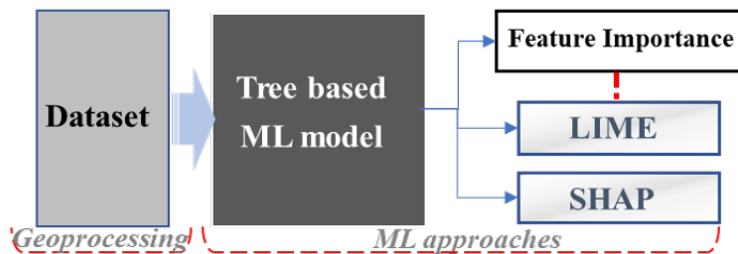

Figure 11. Schematic diagram of the framework of building predictive and explainable ML models.

The geodata is integrated into QGIS Tool, and the data analysis is implemented in Python 3 in the Jupyter notebook interface. The Python libraries used are NumPy and Pandas for data manipulation, Matplotlib and Seaborn for visualization, and Geopandas for geodata manipulation.

The objective of this section is to ascertain the significance of features influencing the UHI effect and determine the pivotal thresholds within the ML model. The initial step involved constructing a tree-based RF Classifier model, followed by the application of methodologies aimed at comprehending feature roles within the dataset. Feature importance analysis using the permutation importance technique was carried out on the RF Classifier model.

For the preliminary stage of feature engineering, the UHI effect was categorized into two classes: 30°C and 35°C. This consolidation facilitated the creation of a binary classification model, distinguishing urban areas with and without the UHI effect, represented by 29°C. Subsequently, the dataset was segregated into independent and dependent variables to represent features and targets. The UHI effect level was designated as the target variable, while the remaining factors served as features.

3.5 Step 3, Extract ('E' of the CLEAR' model); Extracting from models

3.5.1 Explainable machine learning model

This step presents an ML-based transparent model, explainable ML model, which explains the importance of features impacting the UHI effect in urban areas. The explainable model based on the RF model was built to analyze the performance and transparency of the model. Transparency in ML refers to the ability to understand and interpret the inner workings of a model (Ribeiro, S, & Guestrin, 2016). They typically operate by relating the inputs of a model to its outputs without making assumptions about the internal workings of the model. Based on learning theory, ML methods can extract meaningful information and patterns from this data, allowing for the identification of practical solutions tailored to the specific location and context (Milojevic-Dupont, Nikola, and Felix Creutzig, 2021). The importance of informing how the machine makes models and predicts features develops the idea of using the explanation of ML models. Explanations are essential to trust that the predictions made by models are correct. The needs for explainable ML models are because black-box ML models make it hard to understand and explain the behavior of a mode (Ribeiro, S, & Guestrin, 2016).

Building and implementing explainable models
The implemented explainable ML-based models in the study are Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) models. Both models aim to provide insight into how specific predictions were made (Lipton Z. C., 2016).

3.6 Step 4, Apply ('A' of 'CLEAR' model); Application of the research

This section focuses on the practical implementation of the research findings, utilizing a geoprocessed dataset encompassing UHI phenomena and urban data to apply in outdoor thermal comfort and urban livability. Specifically, in this section the correlations between building data, heat waves and ST were examined.

Urban planners increasingly focus on the impact of climate on urban development, with a primary goal of improving residents' well-being by creating more comfortable environments. The quality of outdoor life is a crucial measure for assessing the urban microclimate, and various design solutions are employed to enhance public spaces. Moreover, research into urban-scale thermal comfort highlights the link between urban and landscape planning, emphasizing the importance of considering pedestrians and climate. This emphasis on outdoor thermal comfort has grown significantly since the early 2000s. The thermal comfort assessment used in Paper IV constitutes the quantitative approach of Paper V, which seeks to assess the outdoor thermal comfort and thermal sensation in an urban area.

The objective of this section is to propose effective mitigation strategies that enhance thermal comfort in the studied areas of Tallinn, encompassing both the city and its urban surroundings. Leveraging the insights obtained from the research outcomes and the comprehensive geoprocessed dataset, the section aims to identify suitable measures and interventions that improve thermal comfort in these specific areas.

In summary, this section utilizes the research outcomes, geoprocessed datasets related to UHI phenomena and urban data to find the samples, and the established correlations between building data, heat waves, ST, and outdoor thermal comfort.

The ultimate goal is to propose mitigation strategies that optimize thermal comfort in the case-studied areas of Tallinn. To achieve this, three case studies were conducted, analyzing the area between the target building and the nearest adjacent neighbour across the street. These case studies incorporate a geometric model of a real residential building in Tallinn, the urban canopy, including neighbouring buildings, and the connecting street. By representing the orientation of building mass, the aim is to explore the optimal urban environment orientation that ensures the highest level of outdoor thermal comfort and lowest ST in the analyzed area on a hot summer day.

The building direction often describes the orientation of the canyon axis (e.g., N–S, E–W) or (NW–SE, NE–SW) (Erell, Pearlmutter, & Williamson, 2012). In the definition of scenarios and the simulated models of the study, the orientation of the canyon axis represents the direction of an elongated space, measured (in degrees) as the angle between a line running N-S and a significant axis running the length of a street or other linear area, measured counterclockwise. Figure 12 shows the mentioned orientations on the axes of four main directions.
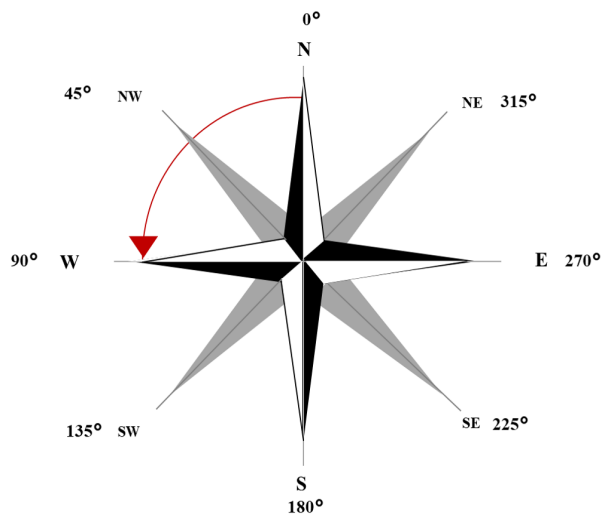

Figure 12. The eight main orientations of the urban environment used in the study.

This study analyzes existing buildings in their actual orientation and other hypothetical orientations. Thus, the orientations of 0°–180° refer to the extension of N–S and S–N, and 270°–90° means the extension of the urban canyon is in the direction of E–W.

Furthermore, the orientation of 45°–225° refers to NW–SE, while the orientation of 135°- 315° goes back to South West- North East (SW-NE), respectively. Figure 13 displays three selected residential case studies in Tallinn.

Case study 1: XY coordination based on Google earth: 59°24′51″N24°44′18″E
Harju county, Tallinn, Kesklinna district, Pärnu mnt 110

Case study 2: XY coordination based on Google earth: 59°26′08″N24°49′15″E
Harju county, Tallinn, Lasnamäe district, Pae tn 68

Case study 3: XY coordination based on Google earth: 59°25′57″N24°46′45″E
Address: Harju county, Tallinn, Kesklinna district, Vesivärava tn 50

Figure 13. Case studies of three residential buildings in Tallinn, Estonia.

Table 5 outlines the specifications of each case study. Moreover, Table 6 outlines the characteristics that were considered when defining different scenarios during modelling and simulations for case studies 1 through 3.

Table 5. Features of the main building in case studies.

Sample	Height (m)	Floors	Length (m)	Total area (m²)
Case study 1/M1	50	17	291.3	2258.4
Case study 2/M2	45.4	14	208.5	2093.3
Case study 3/M3	37.3	9	202.8	2470.6

Table 6. The general features of the simulated models.

Scenarios of simulated case studies in the different extensions of canopy			
Model	Case study	Orientation (°)	Extension
M1	Cs1	347	NE-SW
M2	Cs2	22	N-S
M3	Cs3	325	NE-SW
M1-1	Cs1	0	N-S
M1-2	Cs1	45	NW-SE
M1-3	Cs1	90	W-E
M1-4	Cs1	135	SW-NE
M1-5	Cs1	180	S-N
M1-6	Cs1	225	SE-NW
M1-7	Cs1	270	E-W
M1-8	Cs1	315	NE-SW
M2-1	Cs2	0	N-S
M2-2	Cs2	45	NW-SE
M2-3	Cs2	90	W-E
M2-4	Cs2	135	SW-NE
M2-5	Cs2	180	S-N
M2-6	Cs2	225	SE-NW
M2-7	Cs2	270	E-W
M2-8	Cs2	315	NE-SW
M3-1	Cs3	0	N-S
M3-2	Cs3	45	NW-SE
M3-3	Cs3	90	W-E
M3-4	Cs3	135	SW-NE
M3-5	Cs3	180	S-N
M3-6	Cs3	225	SE-NW
M3-7	Cs3	270	E-W
M3-8	Cs3	315	NE-SW

The methodology is designed with five sequential steps, beginning with Step 0 for data acquisition and concluding with Step 4 for inventory creation and the application of the study (Figure 14).

- level 0: Capturing data
- level 1: Sampling
- level 2: Simulation
- level 3: Assessment
- level 4: Application
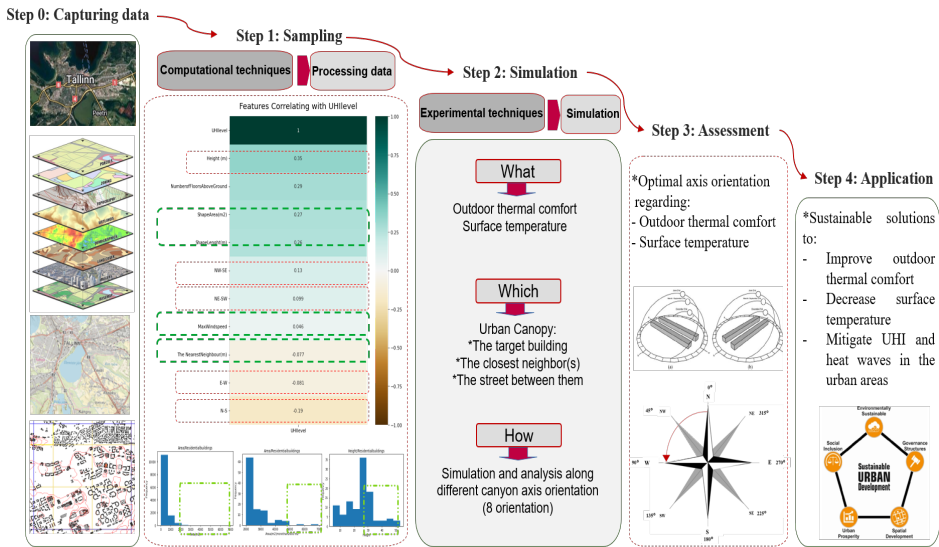
Figure 14. Application of the model: structured steps for urban canyon optimization.

3.6.1 Substep 1: Sampling and finding critical urban canyons

In this section, the study mainly concentrated on residential buildings that are critical cases in the city and experience high intensities of temperatures in the environment, according to the city's UHI data. Thus, the dataset was initially filtered to divide the data according to purpose and directed attention to residential building use. Figure 15 shows the correlation between features that helps to understand which feature has more significance to be concentrated when the heat wave and UHI phenomena in the urban area are concentrated. This is the concept of sampling and finding case studies.
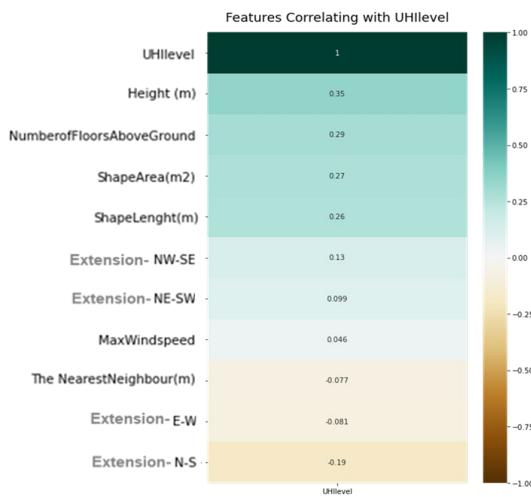
Figure 15. Dependency values between the features of the dataset and the UHI level.

In the initial histogram, the area of the buildings was considered. This data is random and lacks an identifiable pattern. As illustrated in Figure 16.a, buildings larger than 2,000 m² have less frequent footprint areas in the dataset. Because of this, the entire dataset was examined to differentiate between buildings with an area greater than 2,000 m² and identify the tallest buildings, which are not as frequent in this dataset. Thus, as the histogram shows, filtering the dataset to the highest volume buildings helps to find which buildings are the critical cases to study. In addition, Figure 16.b shows samples in the highest area and height of residential buildings in the UHI dataset (more correlated features with the UHI effect) with an area of more than 2000 m² and a height of over 30 m. The graph shows fewer samples with an area higher than 2000 m² and a height higher than 30 m. Consequently, Figure 16.b reveals samples that meet the research question's goal, pointing to some residential buildings in the UHI dataset with an area greater than 2000 m and a height above 30 m. The selected samples with an area of more than 2000 m and a height of over 30 m are shown in red.
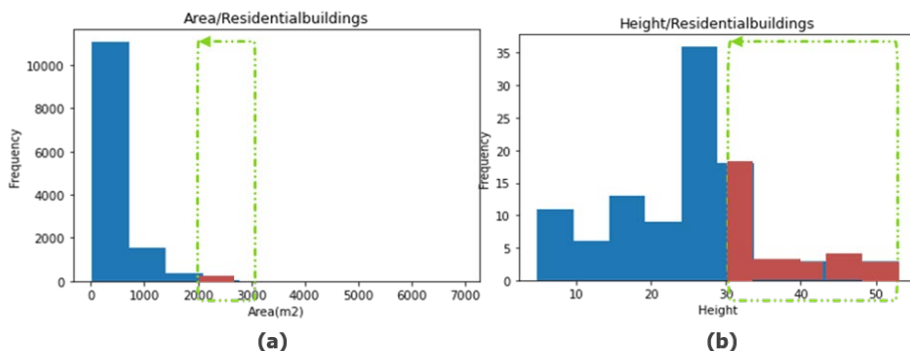

Figure 16. Sampling by using a histogram to choose case studies from urban UHI dataset.

3.6.2   Substep 2: Simulation of buildings at an urban scale

The focus of step 2 is on the geometric modelling and simulation process of case studies in different scenarios to verify outdoor thermal comfort and solar access. Based on the sampling results and the UHI dataset, three urban canyons with high-volume residential buildings in Tallinn, Estonia, were chosen.

To conduct CFD simulation, a three-dimensional computational model of fluid dynamics and energy balance, designed to simulate the microclimate of the study at street level, was employed. The simulation is performed using ENVI-met, a software package specifically developed for urban microclimate modelling.

To assess outdoor thermal comfort, the PET index, which measures the combined effect of AT, RH, WS, and radiation on human thermal comfort, was utilized. PET uses the range of thermal index predicted by human beings and physiological stress on human beings (Cohen, Potchter, & Matzarakis, 2013). The PET index is based on the Munich Energy-balance Model for Individuals (MEMI), which is a two-node model that simulates the thermal balance of the human body in a physiologically relevant way (Deb & Ramachandraiah, 2010).

Through the utilization of this index, it becomes feasible to assess the thermal comfort of the outdoor environment across various scenarios and detect potential concerns associated with heat stress and discomfort. Table 7 shows thermal perception and the ranges of PET in each thermal comfort class. Also, ST, a crucial parameter for

comprehending the potential effects of absorbed solar radiation, was evaluated. This assessment holds the potential to elucidate heightened ST, which in turn can contribute to the UHI effect (Ul Moazzam, Hoi Doh, & Gul Lee, 2022), exacerbate heat waves, and influence the overall thermal performance of buildings and outdoor areas in response to various factors.

Table 7. The range of thermal index predicted thermal perception by human beings and physiological stress on human beings (Cohen, Potchter, & Matzarakis, 2013).

Thermal Perception Grade of physiological stress		
PET(°C)	Thermal Perception	Grade of physiological stress
4	Very cold	Extreme cold stress
8	Cold	Strong cold stress
13	Cool	Moderate Cold stress
18	Slightly cool	Slight cold stress
23	Comfortable	No thermal stress
29	Slightly warm	Slight heat stress
35	Warm	Moderate heat stress
41	Hot	Strong heat stress
	Very hot	Extreme heat stress

4   Results and discussion

4.1   Results of step 1: Data capture; geoprocessed urban data

The dataset used in the thesis was obtained through geoprocessing, programming, and analysis techniques (Eslamirad N., 2023) as a result of the first step of the CLEAR model (Paper I). The dataset served a dual purpose in the subsequent stages. It was not only employed for constructing the ML models but also possesses the versatility to be utilized in other research endeavors and projects associated with the building and urban data of Tallinn, Estonia.

The utilization of an ascending hierarchical grid system is based on the concept of dynamic variation within urban areas. It considers the structure of the data, its features, and its location. To process the data, Python programming packages, and the QGIS Tool were employed for geoprocessing, capture, and analysis.

The dataset presented is extensive and encompasses various disciplines. It consists of 34,001 building samples collected from all eight districts of Tallinn. The dataset includes location, building characteristics, urban characteristics, UHI data, and climate data (Buildingdata, n.d.; keskkonnateadlik hub, n.d.). The proposed methodology to capture urban data establishes a framework to classify the data into homogeneous or heterogeneous, static or dynamic schemes, and then collects the data while considering the homogeneous grid system.

In practical terms, the hierarchical structure guarantees an effective method of gathering data according to location. It achieves this by creating a spatial index for every object on the map and associating them with specific grid cells. This division of the Earth into identifiable grid cells allows for efficient analysis of large sets of spatial data. Each grid in this structure contains consistent and unchanging data in one layer and encompasses all the points, shapes, and other identifiable components within that grid. This hierarchical approach improves the organization and retrieval of spatial data, making it easier to access and manage. Additionally, this solid hierarchical grid system can handle static and dynamic data, allowing for the integration of different types of information (Eslamirad N., 2023).

Implementing the hierarchical grid system in the data collection process has a number of advantages. First, it enables the creation of a spatial index for each object and establishes a connection between the objects and the grid system. Second, it utilizes the homogeneous ground to define an urban index primarily based on the heterogeneous data (Eslamirad N., 2023). The meteorological conditions related to the days that the UHI value was measured in the summer of 2014–2019 were assigned to the dataset. The weather data resource is the Underground website (Eslamirad N., 2023).

4.2   Results of step 2: Learning; predictive models

The outcome of Step 2 adds value to the field of ML by underscoring the advantages of model simplification through numerical feature emphasis. The findings obtained from this study have the potential to facilitate the development of efficient ML models that utilize the intrinsic characteristics of the dataset's features.

This section presents spatial data and UHI data acquisition to build a dataset that is used to build an ML-based transparent model, which explains the importance of features impacting the UHI effect in urban areas.

The dataset was further divided into training and testing subsets. After establishing a model based on the RF model, rigorous assessments of model performance and accuracy were conducted to ensure its efficacy. Notably, precision assumed a critical role given the emphasis on maximizing model performance within this context. The model's accuracy is 90%, while the precision is 82% and 92% in classes 1 and 2, respectively. Impurity-based feature importance breaks the relationship between the feature and the target. Thus, the drop in the model score indicates how much the model depends on the feature. For example, according to the bar chart in Figure 17, the numerical feature related to the 'Built Area in G200', the built area in the minor grid of the defined ascending hierarchical grid system, is the most significant; after that, 'Angle,' 'Height,' and 'Shape Area' of the building are the most important features. The last essential features are 'Purpose of Building' and 'Material,' which are categorical.
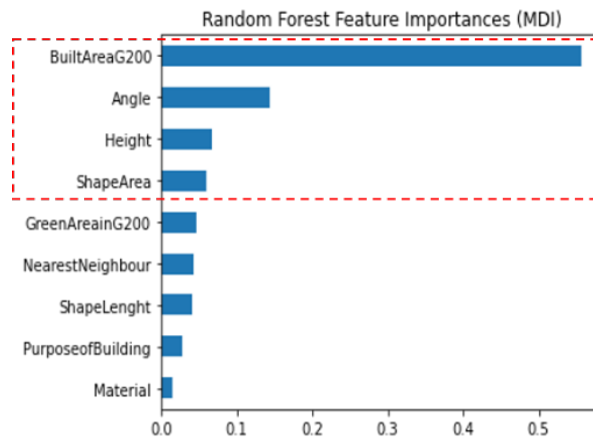
Figure 17. The importance of features in the classifier ML model (Eslamirad, De Luca, Ben Yahia, & Lylykangas, 2023).

According to the impurity-based feature importance, 'BuiltAreaG200', 'Angle,' 'Height,' and 'ShapeArea' of the building are the most important features that impact the labelling of a sample in the class with the UHI effect. Furthermore, the results show that when the value of the feature of 'BuiltAreaG200' increases to more than 10000 m^2, the building is more likely to be classified in UHI class 2, and the UHI effect is more than 30°C. To check the results of feature importance, the initial dataset used for the model creation was considered.

The samples show that just 5.9% of samples with a value of 'BuiltAreaG200' greater than 10000 m^2 are in UHI class 1 (lower than 30°C). Therefore, the transparent ML model results show that the feature importance in the classified model is trustable. Subsequently, the methodology of the study and Explainable model can be applied to develop mitigation strategies to improve the quality of the microclimate conditions in the urban area in other climatic conditions. However, the results rely on site-specific data.

4.3 Results of step 3: Extracting; Results of explainable models

Each attribute holds a certain weight in predicting the occurrence of a class depending on a particular threshold. To explain the decisions behind the ML model, the marginal weights of four top features ranked as the most significant attributed to the impurity-based feature importance were interpreted.

In the next step, two explainable models based on the Random Forest (RF) model were employed to assess the model's performance and transparency. These models typically function by establishing a connection between the model's input and its outputs, without making assumptions about the model's internal mechanisms. The study implemented two explainable ML-based models: LIME, which simplifies explanations, and SHAP models, which provide insights into feature relevance. Both models are geared towards shedding light on the specific reasoning behind the model's predictions (Lipton Z. C., 2016). Each attribute carries a particular weight in predicting the occurrence of a class, contingent upon a specific threshold. To elucidate the rationale behind the ML model's decisions, the marginal weights of the top four features, ranked as the most significant based on impurity-based feature importance in Figure 17, were interpreted.

4.3.1 Substep 1: The explainable model of LIME

According to Figure 18, LIME predicts class 1 with 46% and class 2 with 54% confidence. The first important attribute, 'BuiltAreainG200', indicates the threshold of 10805, meaning above the value, the value increases the chance to be labeled in class 2 with a weight of 0.12, whereas below, it increases the chance of being labeled in class 1 with a weight of 0.12. The following important attribute, 'Angle,' indicates the threshold of 133°, which means when the angle of the building is higher than 133° the UHI class is more likely to be class 2, but a lower angle will lead the sample to be labeled in class 1 (without UHI). The next is 'Height,' while 4 m is the threshold to label a building in class 2, otherwise in class 1. In addition, the feature of 'ShapeArea', which refers to the area of buildings in the dataset, with a threshold of 70.5 m^2, shows that when the area of the building is higher than 70.5 m^2, the sample has a better chance of being labeled in class 2 (UHI level = 30°C); otherwise, it will be in class 1.
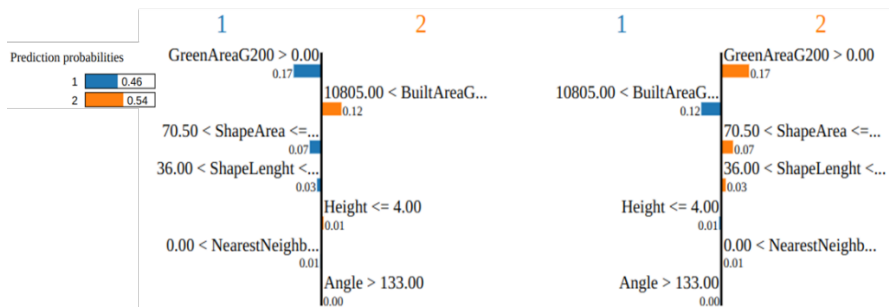

Figure 18. The threshold and value of top features in the LIME Explainable model (Eslamirad, De Luca, Ben Yahia, & Lylykangas, 2023).

4.3.2 Substep 1: The explainable model of SHAP

SHAP is the most popular explainable model, based on coalitional Game Theory (Shapley values), and involves identifying the essential features that contribute to the model's predictions to explain and interpret the coefficients as the feature's importance. In the plotted results of the SHAP model, the red marks push the prediction of the UHI value higher toward the base value (about 30°C), while the blue marks just push the value lower than 30°C (Figure 19).

Figure 19. Performance of the SHAP explainable Model (Eslamirad, De Luca, Ben Yahia, & Lylykangas, 2023).

Moreover, according to the plot of dependency in Figure 20. a, between all the most significant values in the model, the shape value increases with an increase in the value of 'BuiltAreainG200' as the highest feature value in the SHAP model, in the wide range, and then 'ShapeArea' and 'Angle' as the second and third features. Also, Figure 20.b shows dots as a single prediction (row) from the dataset. In addition, the x-axis is the value of the 'BuiltAreainG200,' and the y-axis is the SHAP value for that feature, which represents how much the value of the feature changes the model's output for that sample's prediction. Moreover, the red dots represent the feature of 'ShapeArea' as a second feature that may interact with the plotted feature. Therefore, as the plot indicates, the high value of 'BuiltAreainG200' and 'ShapeArea' maximizes the SHAP value.
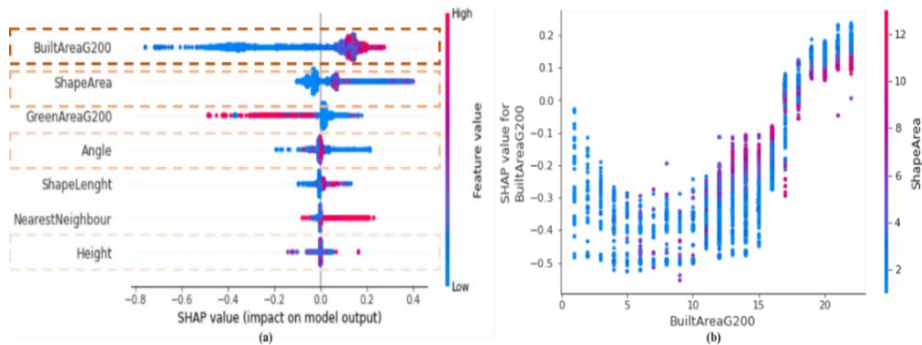

Figure 20. The plot of the dependence (a), the scatter plot of 'BuiltAreainG200' in SHAP value (b) (Eslamirad, De Luca, Ben Yahia, & Lylykangas, 2023).

4.4   Results of step 4: Application of the research findings

4.4.1   Substep 1: Assessment of thermal comfort and surface temperature

Through the CFD simulation and analysis, output data related to the thermal comfort of people at the pedestrian level and the ST of the urban area under different scenarios were acquired.

To evaluate thermal comfort, a comprehensive approach to evaluating the thermal performance of the studied urban area under different scenarios was used. The first thermal comfort analysis is about finding the non-uniform spatial distribution of PET in each particular scenario in the urban canopy between the target building and the nearest neighbour.

The second analysis aims to create a model based on the scoring system to show the uniform or normalized spatial distribution of PET. The normalized PET of the urban canyon in each scenario is the weighted PET value calculated by considering the quality and quantity of PET data. The highest value of the weighted PET leads to finding the optimal degree of orientation in the urban canyon.

The scoring system that was implemented considers the thermal comfort level at each point in the canopy area. The scoring system allows one to calculate the overall level of comfortability at the pedestrian level by combining the scores of all the individual points. In addition, the ST assessment of the studied areas was performed using the results of the CFD simulation. Overall, the orientations lead to the lowest ST highlighted as the optimum building mass extension and taken into account in the final assessment to determine the best urban environment orientation to ensure comfortability.

4.4.2   Substep 2: Application to find the optimal building mass orientation

Output data related to the optimal building mass orientation is helpful to imagine more sustainable solutions in cities. Furthermore, by improving outdoor thermal comfort and ST in urban areas and leveraging UHI data, the study provides valuable insights into the thermal performance of the studied urban area.

The evaluation is particularly beneficial to guide urban planners and architects in proposing mitigation solutions to enhance thermal comfort in cities and create suitable conditions for achieving approved thermal comfort levels with complementary solar access in the city area. With this information, planners and architects can make more informed decisions about the design of new buildings, the placement of green spaces and other urban elements, and the use of shading devices and other technologies to reduce heat gain and improve outdoor thermal comfort.

Overall, the study's findings highlight the importance of considering outdoor thermal comfort and solar access in urban design and planning to create more livable and sustainable cities.

CFD simulation and numerical analysis

The section is related to the CFD simulation and the numerical analysis to assess the outdoor thermal comfort and ST in the studied areas. CFD simulation is used to simulate the microclimate of the studied area at street level, considering the influence of various factors on thermal comfort, such as AT, RH, WS, and radiation. By combining CFD simulation and numerical analysis, a holistic comprehension of the thermal performance of the studied area is achieved across diverse scenarios. This information can then be used to identify potential issues related to heat stress and discomfort and propose mitigation solutions to improve outdoor thermal comfort.

Figure 21 and Table 8 show more detailed information about the areas focused on in the CFD simulation and the area of the urban canyon in which PET results were evaluated.
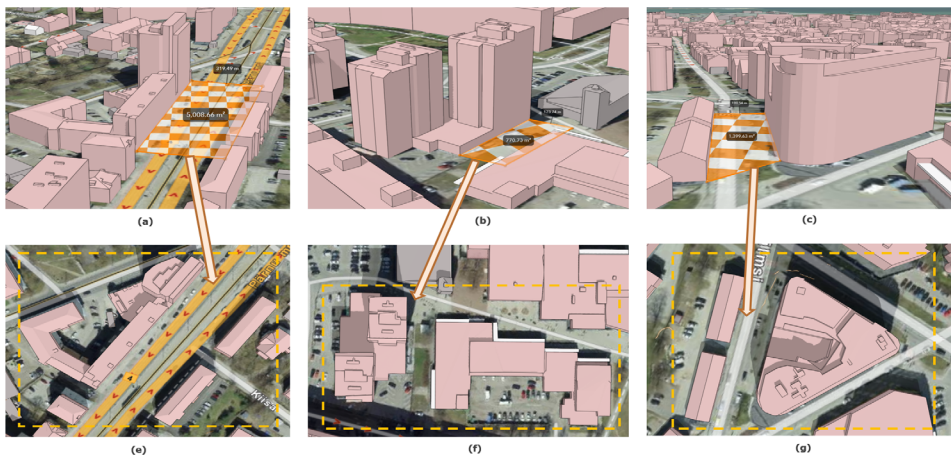
Figure 21. The Canopy areas (the hatched rectangular shapes) and the whole simulated areas of each case study (the dashed rectangular shapes) (a) The Canopy area, cases study 1; (b) The Canopy area, cases study 2; (c) The Canopy area, cases study 3, (e) The whole simulated area, case study 1; (f) The whole simulated area, case study 2, (g) The whole simulated area, case study 3.

Table 8. Detailed information about the CFD simulated areas of each case study.

Case studies	Canopy area (m²)	Simulated area (m²)
Case study 1	20,650	5,000
Case study 2	15,000	770
Case study 3	12,700	1,400

Different scenarios of case studies were simulated based on the specific orientation of the canopy extension, including E–W, N–S, Southeast Northwest (SE–NW), and Northeast Southwest (NE–SW).

ENVI-met requires information about surrounding environmental features and meteorological data to conduct the CFD simulation of the generated geometric models.

The input data for the simulation models are the physical properties of the studied urban areas (buildings, soil, and vegetation) and geographic and meteorological data (Yahia & Johansson, 2012). This study conducted the CFD simulation on July 25, 2014, during a high UHI and heat wave period (Yahia & Johansson, 2012). The simulated period

lasted from 16:00 to 17:00, including the maximum AT during a summer day of 28°C. The outdoor thermal comfort assessment was conducted at 17:00 and evaluated at 1.80 m, the average human height. Simple forcing was used in all scenarios to adjust for the meteorological conditions, creating a 24-hour weather data cycle that defined the meteorological boundary conditions for the ENVI-met simulation.

Table 9 gives information about the weather condition of Tallinn on the supposed date and time that was chosen for the CFD simulation.

Table 9. The input meteorological data during the CFD simulation by ENVI-met.

Date: July 25, 2014. Time: 17:00.	
Temperature (AT) (°C)	Max 28/Min 17
Relative Humidity (RH) (%)	Max 75/Min 45
Wind speed (WS) at inflow border at inflow border (m/s)	2.00
Wind Direction (WD) at inflow (°)	90.00
Roughness length (m)	0.010
Specific humidity in 2500m (g/kg)	8.00

Outdoor thermal comfort assessment

PET, expressed in °C, is based on the human energy balance model MEMI and includes the physiological thermoregulatory processes of human beings to adjust to a climatic situation outdoors. The thermal comfort zone for the PET index was initially defined as 18–2°C (Yahia & Johansson, 2012). In this section, the authors listed all the parameters used in the CFD simulation. Also, during the simulation, the building's indoor temperature was set to a constant value of 20°C. Therefore, the outside microclimate did not influence the building temperature. Overall, using the PET index and CFD simulations is a useful approach to assess the thermal comfort of the studied area. By taking into account the physiological thermoregulatory processes of human beings and using advanced simulation techniques, a more accurate and comprehensive understanding of outdoor thermal comfort in urban areas will be achieved.

Thermal comfort is a subjective concept that depends on personal features and describes a person's state of mind regarding whether they feel comfortable (Faria Neto, Inácio, Wurtz, & Delinchant, 2016). Thus, once the meteorological data and environmental characteristics are added to the input data used in the CFD simulation, thermal comfort in PET indices needs to set the individual personal data that are supposed as the users of the urban areas.

In this section, PET is taken as the outdoor thermal comfort assessment and calculated just for a male pedestrian wearing very light summer clothes in the standing position with the walking speed of 1.2 m/s. For a simple PET assessment process, just male pedestrians wearing unique clothing values with normal body parameters were considered. Table 10 shows other personal parameters used in PET evaluation.

Basic personal parameters	
Age of the person	35
Weight (kg)	75
Height (kg)	1.75
Surface area of the body (sm²)	1.91
Clo	0.10
Metabolic work(W)	164.70

Surface temperature assessment

The assessment of the ST of the urban canopy in each scenario is a valuable approach to understanding the impact of changes in canopy orientation on urban temperature. By analyzing the minimum, maximum, and median values of ST, the optimum orientation of the canopy for maximizing thermal comfort and decreasing temperatures on urban surfaces can be identified. Additionally, the ST is a crucial metric for evaluating the UHI effect because it shows how much heat the surfaces of the urban environment have absorbed. Thus, by reducing ST, it is possible to mitigate the UHI effect and improve thermal comfort for pedestrians.

Through the ST analysis, the impact of canopy orientation on ST and the optimal orientation that reduces ST and maximizes thermal comfort can be determined. This information can then inform urban planning and design strategies that prioritize thermal comfort and sustainability. It is important to consider these orientations as they can cause more heat on urban surfaces and potentially result in lower levels of thermal comfort for pedestrians.

According to Figures 22.a and 22.b, in case studies 1 and 2, scenarios M1-8, M1-6 (24.9 and 28.3°C), and M2.8, M2.6 (24.3 and 24.4°C) have the lowest median ST when oriented at 315°C and 225°C, respectively. This suggests that these orientations can provide the highest thermal comfort for pedestrians in the case studies. Likewise, Figure 22.c shows in case study 3 that the median ST data is observed in M3-5 and M3-4 (22.4 and 23.6°C) with orientations of 180°C and 135°C, respectively. The finding indicates that these orientations can also provide high thermal comfort for pedestrians in this case study. Furthermore, in case study 1, the orientations with the highest median of ST are M1-1, M1-2, and M1-3, with orientations of 0°, 45°, and 90°, respectively. In case study 2, the orientations with the highest median of ST are M2-1, M2-4, M2-2, and M2-3, with orientations of 0°, 135°, 45°, and 90°, respectively.

Finally, in case study 3, the orientations causing the highest median of ST in the analyzed areas are M3-2, M3-1, M3-3, and M3-8 with orientations of 45°, 0°, 90°, and 315°, respectively.

This section presents the results of outdoor thermal comfort, expressed in terms of the PET index and ST in degrees centigrade (°C), to choose the best orientations in each case study that lead to the highest comfort level. The spatial distribution of outdoor thermal comfort in terms of the metric PET was calculated via simulation for all scenarios

in three case studies. The thermal comfort assessment results are explained in two forms: non-uniform and normalized spatial distributions of PET.
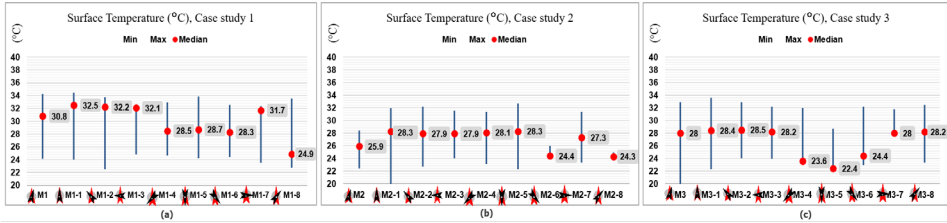

Figure 22. Results of ST (°C) in the urban canopy of different scenarios (a), Case study 1, (b) Case study 2, and (c) Case study 3.

Non-uniform spatial distribution of PET

Figures 23, 24, and 25 show the initial results of PET for case studies in a different scenario, taking into account the input setting, meteorological data, and parameters during the CFD simulation in the defined canyon orientation.

For example, in Figure 23, the initial results of PET assessment in different scenarios of case study 1 are demonstrated. Furthermore, Figures 24 and 25 show the results of the PET assessment of case studies 2 and 3, respectively.

It can be seen in the line graph in Figure 23 that, although the minimum value of PET in all scenarios is almost the same, different scenarios have different values in the average, median, and maximum of PET. Likewise, M1 with the original orientation of 347° NE–SW has the lowest minimum value of PET, while the other values are even higher than others. Moreover, M1-1 and M1-5, with orientations of 0 and 180°, have the highest values of the maximum and median of PET. Likewise, a comparison of all different orientations in the urban canopy of case study 1 indicates M1-2 with an orientation of 45°has the lowest median PET value compared to other scenarios.
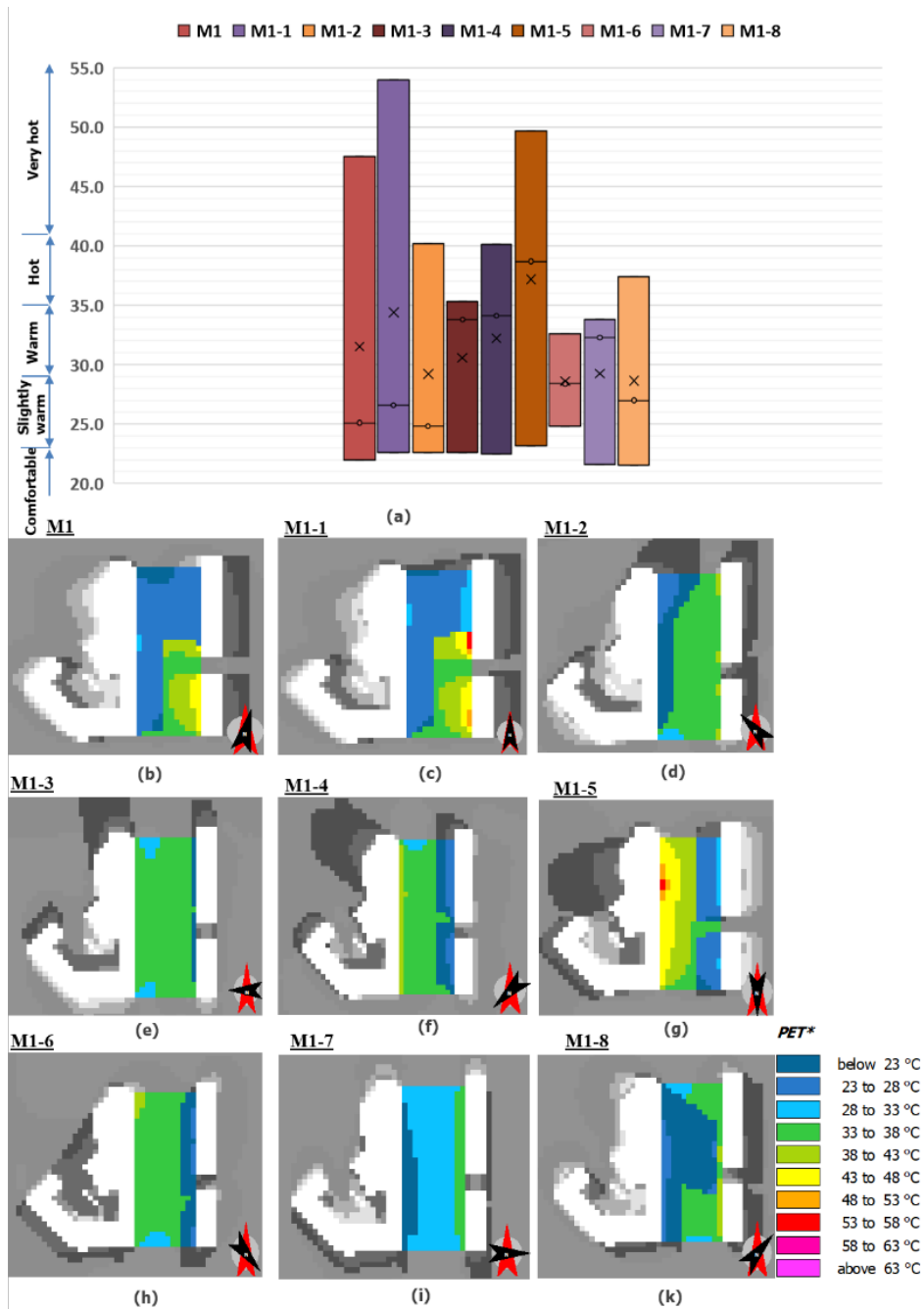
Figure 23. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 1, Scenarios: M1, M1-1, to M1-8. For the configuration shown in (a), The line graph shows the minimum, average, median, and maximum of PET in each scenario, Spatial distribution of PET: (b) M1 (c) M1-1, (d) M1-2, (e) M1-3, (f) M1-4, (g) M1-5, (h) M1-6, (i) M1-7, (k) M1-8.
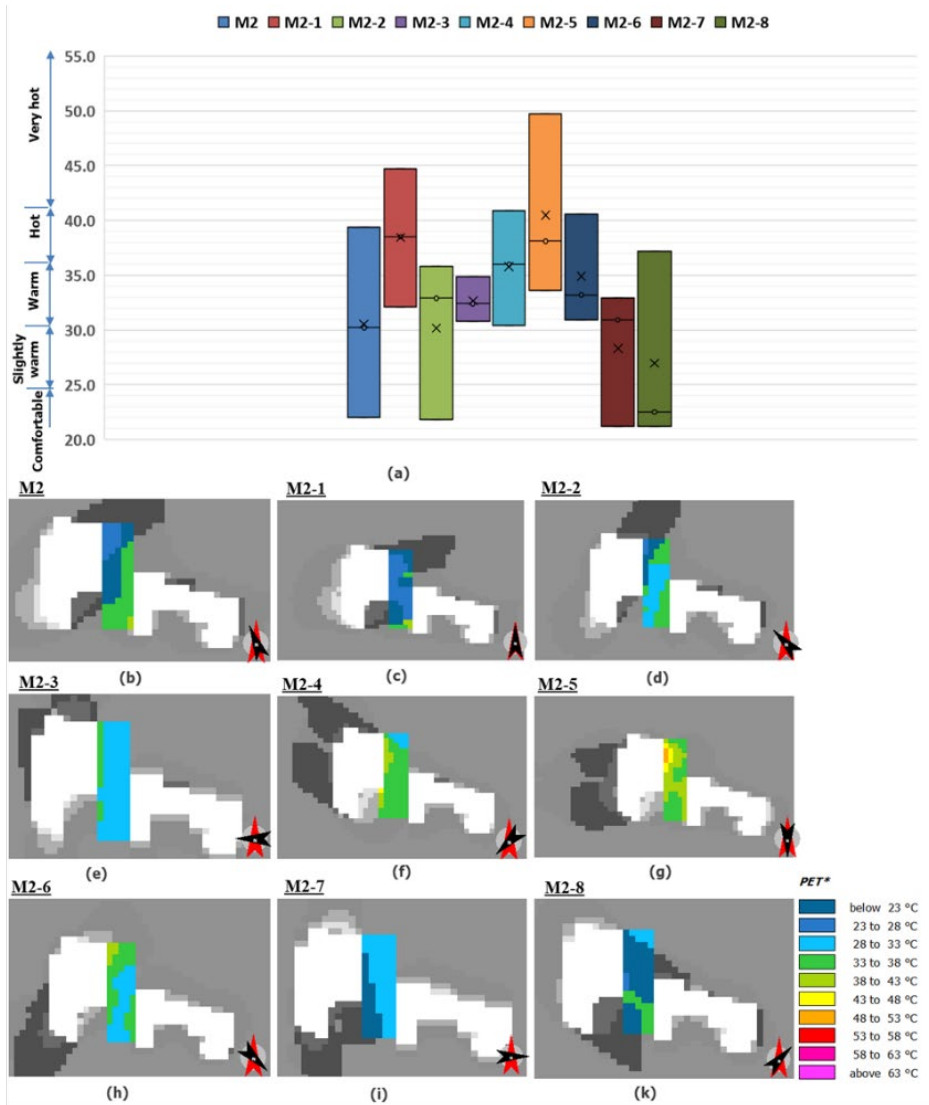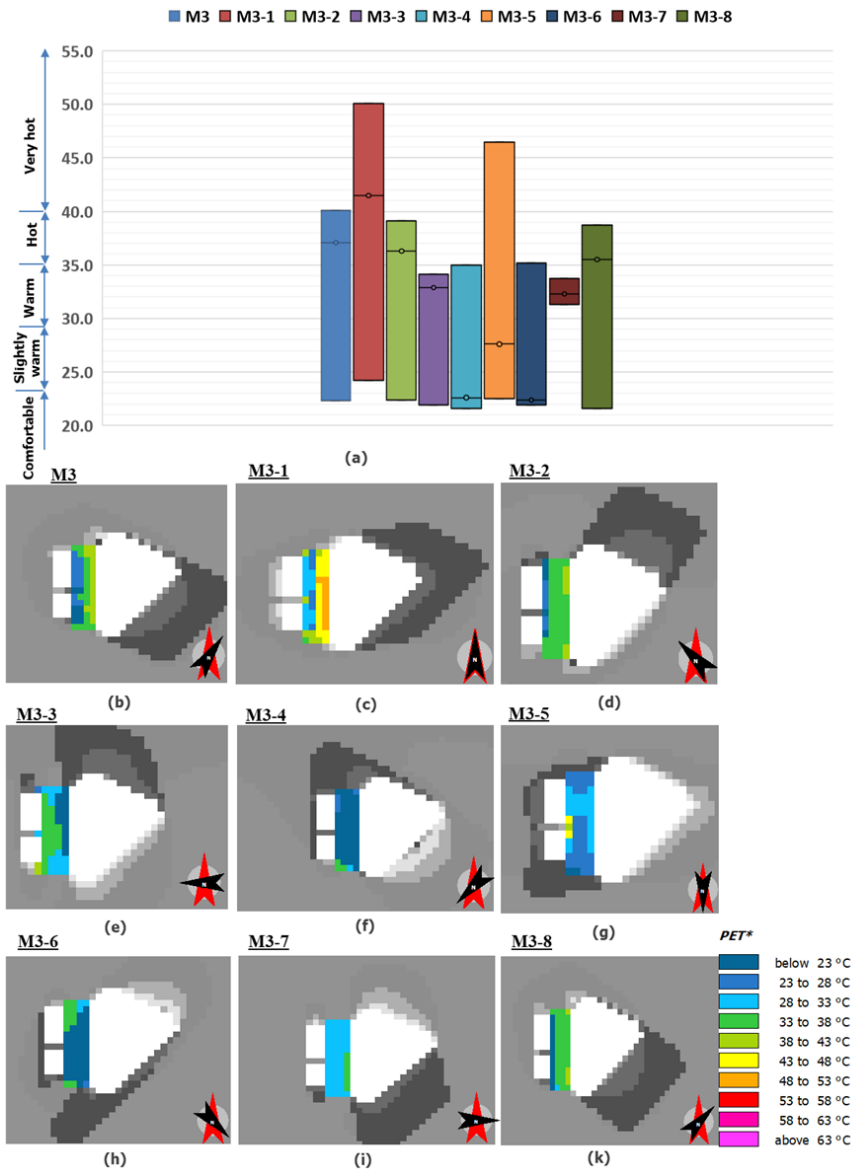
Figure 24. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 2, Scenarios: M2, M2-1 to M2-8. For the configuration shown in (a), The line graph shows the minimum, average, median, and maximum of PET in each scenario. Spatial distribution of PET: (b) M2 (c) M2-1, (d) M2-2, (e) M2-3, (f) M2-4, (g) M2-5, (h) M2-6, (i) M2-7, (k) M2-8.

Figure 25. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 3, Scenarios: M3, M3-1 to M3-8. For the configuration shown in (a), The line graph shows the minimum, average, median, and a maximum of PET in each scenario, Spatial distribution of PET, (b) M3 (c) M3-1, (d) M3-2, (e) M3-3, (f) M3-4, (g) M3-5, (h) M3-6, (i) M3-7, (k) M3-8.

Thus, there are different measures of PET, such as the minimum, maximum, median, and average values, which can vary depending on the scenario. But, based on the initial evaluation of PET of the simulated scenarios, it is impossible to conclude which orientation offers a higher level of thermal comfort. Therefore, the non-uniform spatial distribution of PET can indeed make it difficult to interpret the results of the study and draw definitive conclusions about which scenario offers a better level of thermal comfort at the pedestrian level.

Normalized spatial distribution of PET

According to Joshi et al., measuring subjective experiences or phenomena can be a challenging task, as they are often difficult to quantify using conventional measurement techniques (Joshi, Kale, Chandel, & Pal, 2015). Therefore, for a thorough analysis and to determine the optimal orientation of the urban canopy within various scenarios, it is imperative to take into account both the frequency of PET data at each level and the highest, lowest, and average PET values within each urban canyon.

Therefore, evaluation scales can be presented in various graphical ways with different levels of detail, and no standard gives specifications on the choice of the most suitable configuration; thus, the selection is often a matter of the specifications of the study (Giampaoletti, et al., 2020).

In the study, Nazarian et al. used the continuous Outdoor Thermal Comfort Autonomy (OTCA) scale as a metric to measure outdoor thermal comfort. According to the authors, OTCA considers the percentage of time an outdoor space is within the desired thermal comfort range, including periods where the thermal comfort level is below the threshold. It is an extension of Spatial OTCA, which is defined as the percentage of outdoor space that is within the desired thermal comfort range for at least half of the occupied time (over a year or a prescribed period of use) (Nazarian, Acero, & Norford, 2019).

Here, a weighted scale was designed to consider the level of thermal comfort in the studied areas and rank the PET data based on the frequency of data in each PET class. As a widespread scale used in different areas like psychology, sociology, health care, marketing, attitude, preference, customers' quality perceptions or expectations, and subjective well-being in health care, Likert scales have wide applications in different sciences (Chakrabartty, 2014). In addition, Likert scales are examples of such scales in Psychometrics used widely in social science & educational research (Joshi, Kale, Chandel, & Pal, 2015). Therefore, the Likert scale was designed to weigh the PET classes and rank the importance of data at each thermal comfort level.

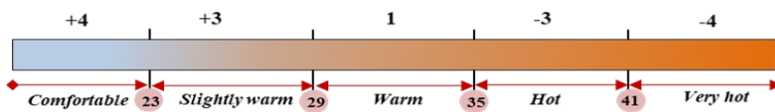Figure 26 shows the Likert scaling system applied in the study.

Figure 26. Five-point Likert scale used in the PET analysis of scenarios.

Each item in the Likert scale usually has an odd number of response categories, up to five or seven levels (Chakrabartty, 2014), and is named the five-point or seven-point Likert scale. Here are five points: The Likert scale was applied by considering the comfortability of the area that the Likert scale should measure and assigning the highest indicator equal to +4 to the PET class of Comfortable, +3 to Slightly Warm, +1 to Warm, and the negative scores to the worst classes of PET that cause a high level of discomfort in the urban area, meaning -3 and -4 to Hot and Very Hot classes.

Statistical methods and exploration of data

In this section, to better interpret the results of the overall thermal comfort level of each scenario, not only considering the arithmetic and mathematical average of PET, but also taking into account the frequency of data in each level of PET is essential. Therefore, it is an excellent approach to consider the frequency of data at each level of PET to better interpret the results of each scenario's overall thermal comfort level. Accordingly, at first, the results of the PET assessment of each scenario were sorted as the experimental data in a matrix and split into five levels of thermal perception.

To describe the process of analyzing the results of a study on thermal comfort levels in different urban scenarios, it should be mentioned that, at first, the results of the PET assessment of each scenario were split into five levels of thermal perception based on a five-point Likert scale. To better understand the distribution of PET data in each thermal comfort level, an experimental matrix and pie charts that show the percentage of PET data distribution of scenarios in different classes of PET (Figures 27, 28, and 29) were created.
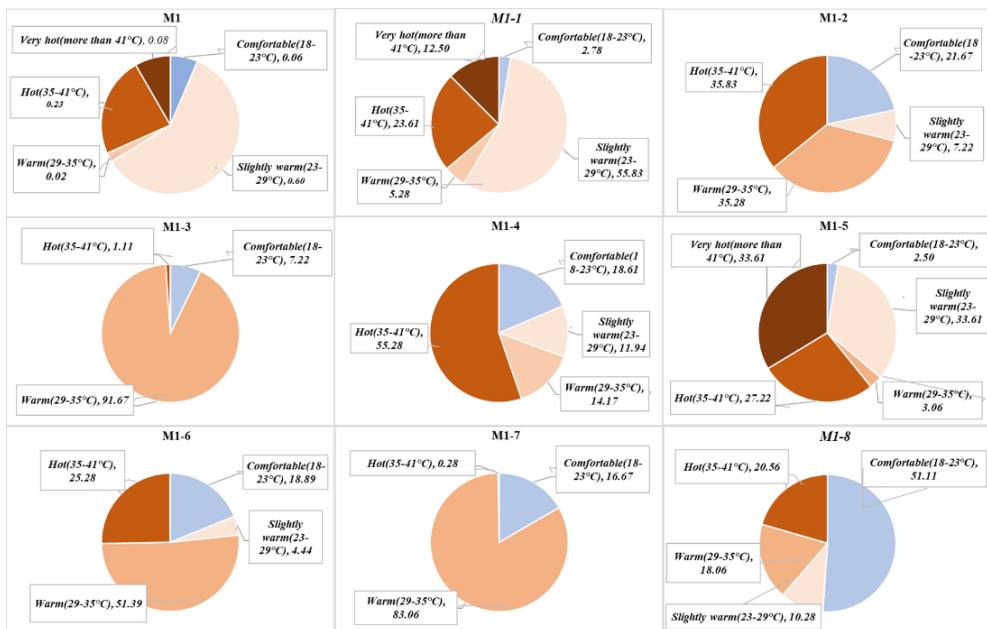
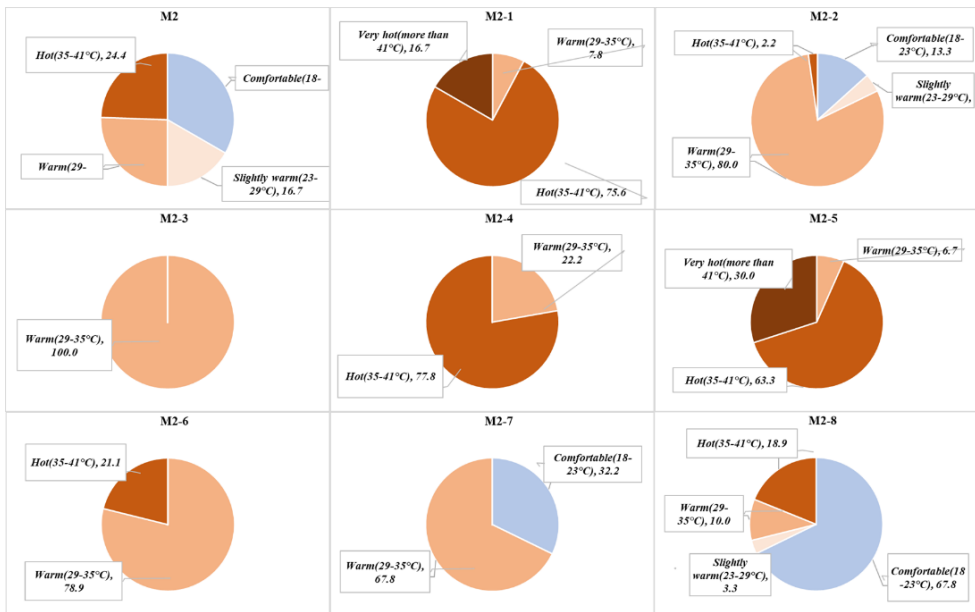Figure 27. The distribution of PET data in each class of PET/ Case study 1.

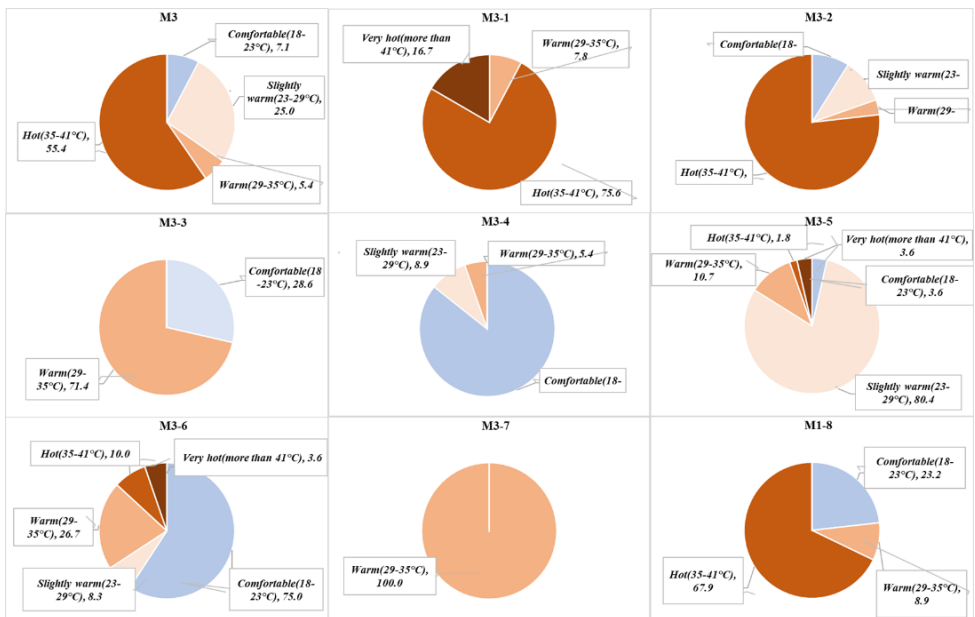*Figure 28. The distribution of PET data in each class of PET/ Case Study 2.*

Figure 29. The distribution of PET data in each class of PET/Case Study 3.

The experimental matrix contains the thermal perception of all scenarios and the arithmetic average (mathematical average) of PET at each thermal comfort level, starting from comfortable to very hot. Moreover, the quantity of PET data in each thermal comfort level is counted to understand how much of each urban area in the urban canopy offers the considered thermal comfort level. To further refine the analysis, the Likert

scaling system was applied to assign a score to each thermal comfort level that reflected its importance. By combining the arithmetic average of PET in each thermal comfort level with the weight or value of each thermal perception level and the quantity of PET data in each level, it is possible to obtain a more accurate and meaningful measure of the overall thermal comfort level of each scenario.

Hence, a statistical approach was established, employing the arithmetic mean of PET within each thermal comfort level. This classification ranged from the comfortable zone with PET below 23°C, signifying the comfort zone, to the very hot level with PET surpassing 41°C.

The next step is calculating the statistical average of each PET level by multiplying the arithmetic average of each PET level by the weights of the Likert scale and the respective count of data in each PET level. Therefore, Equation 1 shows the final formula that shows the Weighted mean (WM) of PET (Wm-PET), that is used in the PET exploration method, and Figure 30 is the schematic diagram of the model to evaluate PET in each scenario.

$$Weighted\ Mean\ (Wm) = \sum_{ni=1}^{n}(xi*wi)/\sum_{ni=1}^{n}wi$$

$$Wm = w_1x_1 + w_2x_2 + \ldots + w_nx_n/w_1 + w_{\ldots} + w_n$$

Where: ∑ denotes the sum

w is the weights, and x is the value of PET

In cases where the sum of weights is 1,

$$Wm = \sum_{ni=1}^{n}(xi*wi)$$
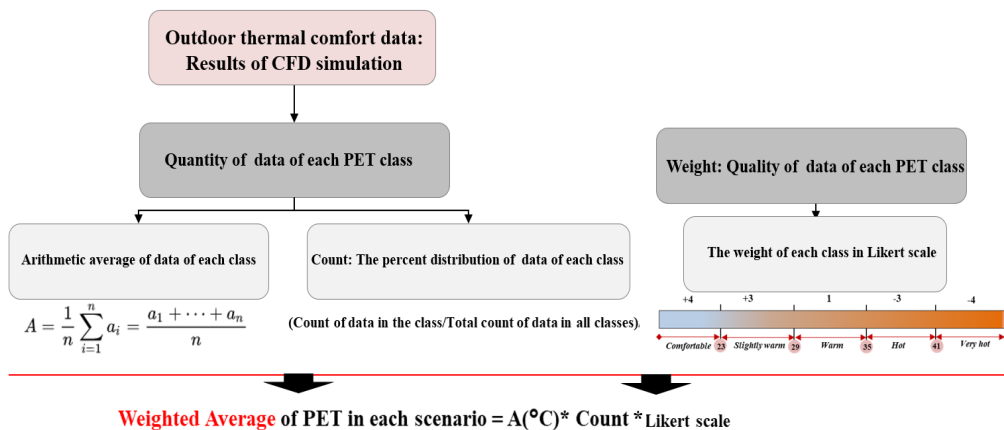
Equation 1. The formula of Weighted mean of PET

Figure 30. The evaluation model for PET value.

Assessment and analysis of urban thermal environment

The next step is calculating the statistical average of each PET level. For example, to calculate the Wm-PET for scenario M1, the arithmetic average of each PET level (22.5, 24.6, 34.5, 38, and 43.4) was multiplied by the weights of each response in the Likert scale (4, 3, 1, -3, and -4), and then each of those values was multiplied by the count of responses (0.064, 0.597, 0.022, 0.233, and 0.083).

For example, for PET level 1 (Comfortable), the Weighted score (Ws) would be:

$$(22.5 * 4 * 0.064) = 5.76$$

The process is repeated for each PET level, and then the weighted scores of PET (Ws-PET) for all PET levels are summed up to get the overall Wm-PET in scenario M1. The calculation for the overall Wm is:

$$Wm\_PET, M1 = (Ws\_PET\_level1) + (Ws\_PET\_level2) + \dots + (Ws\_PET\_level5)$$
$$= (22.5 * 4 * 0.064) + (24.6 * 3 * 0.597) + (34.5 * 1 * 0.022) + (38 * (-3) * 0.233) + (43.4 * (-4) * 0.083)$$
$$= 5.7 + 44.1 + 0.8 - 26.6 - 14.5 = 9.5$$

Therefore, the Wm-PET in scenario M1 is 9.5, which represents the overall level of perceived exertion for this scenario, taking into account both the quantity of data in PET classes and the weight of each class as well as the arithmetic average of PET data in each level. To report the result regarding the optimal orientation for achieving the highest thermal comfort in different scenarios, the summarized weighted average of each scenario is shown in Figure 31.
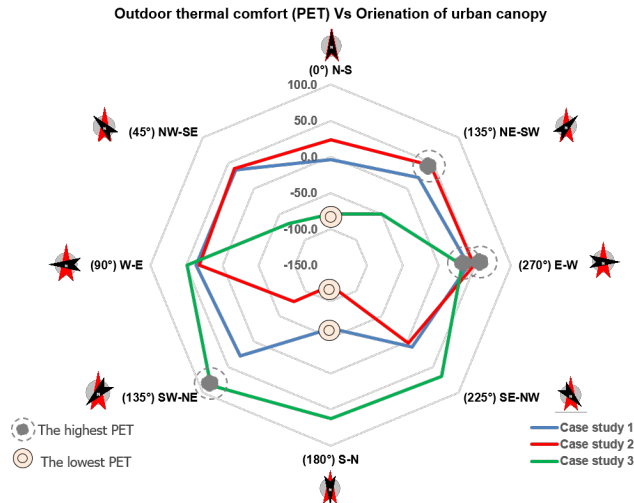

Figure 31. The results of the weighted average PET of each case study in different orientations of the urban canopy.

Based on the information in Figure 31, the highest thermal comfort value in the PET index in case study 1 is observed in scenario M1-7 when the urban canyon is oriented in the E–W direction (270°). For case studies 2 and 3, the highest PET levels were observed in scenarios M2-7 and M3-4, respectively, with an orientation of 270°, an extension E–W, and an orientation of 135°, the extension of South West to North East in urban canyons. Based on the assessed outcomes of outdoor thermal comfort in the urban canopy of the case studies, it seems that the extension of the canopy from N to S has the lowest thermal

comfort value in all three case studies. This could be due to the fact that the sun is at a high angle during the day, which can create hot and uncomfortable conditions in these orientations. Additionally, as Figure 31 shows, the optimum orientation for outdoor thermal comfort can vary depending on the specific features of the case study.

The study's findings suggest that the orientation of urban canyons can significantly impact thermal comfort and ST in urban environments. By considering the orientation of urban canyons, urban planners and designers can optimize the design of urban environments to prioritize the well-being of residents and visitors.

The study results are shown in Figure 32, which summarizes the finding that the West-East and E-W orientations offer the best thermal comfort and lowest ST for all case studies. However, the orientation of urban canyons should also consider the tallest wall's location in the canyon, as this can significantly impact thermal comfort and s ST. Additionally, the study found that orientations of NE-SW and NW-SE offer optimum thermal comfort for case studies 1 and 2 at the pedestrian level but create the lowest thermal comfort for case study 3. Finally, the orientation from Southwest to Northeast offers optimum outdoor thermal comfort for case studies 1 and 3. In contrast, the orientation of south-east-northwest causes good thermal comfort only for case study 3. The optimal orientation for canopies varies depending on the specific characteristics of each case study and its surroundings. In cases where a tall building is located on one side of a canyon or street, the orientation of the canopy can play a significant role in determining the levels of thermal comfort experienced in the shaded areas. For example, in case studies 1 and 2, extending the canopy from NE-SW does not provide ideal shading for the area due to the orientation of the taller building on the left side of the canyon. This could result in high levels of PET in the shaded area, which could lead to discomfort at the pedestrian level. On the other hand, in case study 3, where the taller wall of the canopy is located on the right side, extending the canopy from NE to -SW could provide better shading and improve thermal comfort. Therefore, it is essential to consider the specific morphological features of the building and surrounding area when designing.
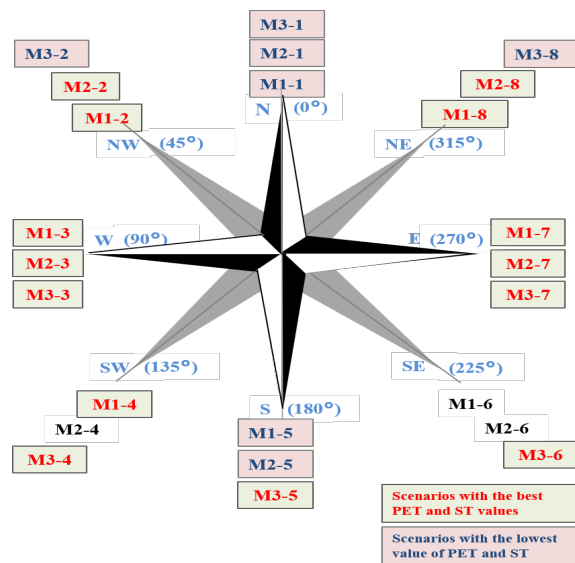

Figure 32. Final results regarding optimum thermal comfort and ST of each scenario in different urban environments.

The results can be beneficial in determining the optimal extension of the canopy layer that maximizes thermal comfort, thus providing more areas of thermal comfort to pedestrians.

Overall, this study's findings can be helpful for urban planners and designers to create more livable and sustainable urban environments that prioritize the well-being of residents and visitors. In addition, by taking into account the specific features of the urban canyon, planners can design and implement targeted strategies for shading, ventilation, and other thermal comfort measures that can further improve outdoor thermal comfort in the area.

This investigation aims to explore the effect of various orientations of the urban canopy on the level of outdoor thermal comfort and ST during hot days, while the case studies taken from the urban UHI data experienced substantial UHI. In this study, numerical analysis and CFD simulations were used to evaluate thermal comfort and ST in urban areas. This investigation aims to find the connection between building mass orientation, pedestrian-level outdoor thermal comfort, and ST in the urban canopy area. By identifying building orientations that lead to maximum ST and thermal comfort in the urban area, urban planners and designers can choose more appropriate insights to optimize thermal comfort and mitigate the UHI effect in cities. Thus, the findings of the study can be applied to create more livable and sustainable urban environments that prioritize the well-being of residents and visitors and mitigate the UHI effect in the urban environment.

The results of the assessment of outdoor thermal comfort in the urban canopy can provide important insights for urban planners and designers to optimize the extension of the canopy layer for maximum thermal comfort. By identifying the ideal orientation for the extension of the canopy layer, planners can create more areas of thermal comfort for pedestrians, which can help improve the residents' quality of life and well-being. Furthermore, urban planning and associated elements extensively affect microclimatic conditions in cities and the lifestyle and well-being of those who inhabit them. Consequently, the study's result can enhance the quality of urban areas by improving thermal comfort and the temperature of surfaces in urban areas. In addition, the findings of this study have the potential to apply to other locations in similar climates and latitudes, especially in the early stages of designing cities or redeveloping and renovating projects in urban areas.

Overall, prioritizing outdoor thermal comfort in the design and planning of urban areas can have significant benefits for both residents and the city as a whole. By designing urban areas that prioritize outdoor thermal comfort on hot summer days, planners can create more livable and sustainable cities that are comfortable for residents and visitors alike. In addition to the social and environmental benefits, there can also be significant economic benefits. By reducing the demand for cooling energy during hot weather conditions, cities can save on energy costs and reduce their carbon footprint. The study shows that the orientation of the urban environment, buildings, and streets in an urban canyon significantly impacts the thermal comfort levels experienced by people in the area. The study also highlights the importance of building mass and the features of the environment, such as the height and layout of canopy walls, in evaluating thermal comfort and ST. By considering these factors when designing and planning urban environments, urban planners and designers can create more comfortable and sustainable living spaces for residents and visitors.

5 Conclusion

The urban microclimate is a vital element within sustainable urban development, exerting a significant influence on the well-being of city dwellers and playing a pivotal role in critical environmental issues like the UHI effect and the occurrence of heatwaves. The UHI effect, influenced by urbanization, significantly impacts global warming assessments. Addressing these multifaceted challenges requires a comprehensive strategy encompassing urban planning, architectural considerations, and environmental factors.

This thesis represents a pioneering application of data-driven methodologies and techniques in addressing the intricacies of UHI effects and heatwaves. It bestows us a geoprocessed dataset with the potential to illuminate the path for future research endeavors, enriching our comprehension of urban features. In the thesis, the CLEAR model is introduced as a comprehensive and innovative framework to tackle urgent urban climate challenges, particularly addressing the UHI phenomenon. Additionally, the ML models developed through this research offer practical solutions that can be invaluable aids for architects, urban planners, and policymakers as they grapple with the multifaceted challenges presented by urban climate dynamics.

To summarize the publications this thesis comprises, a series of papers collectively explore various facets of urban microclimate research and cover the main outcomes of the thesis.

The main research outcomes are as follows:

1. The introduction of a geospatial data workflow for studying Tallinn's UHI phenomenon and incorporating meteorological data from 2014–2019 provided a comprehensive dataset for UHI analysis in Tallinn. Additionally, the geoprocessed dataset can be particularly valuable for research related to buildings, their contextual relationships, neighboring features, and other location-based data that require precise assignment (Paper I);
2. Development of ML models to assess heatwaves and the UHI effect in Tallinn involved leveraging urban, building, weather data, and UHI data. This process encompassed quantifying, building a learning algorithm, and comprehensively understanding the factors influencing the UHI effect using urban geoprocessed data. The outcome was the creation of predictive ML models utilized in transparent and explainable models, all offering valuable insights into these urban climate phenomena (Paper II);
3. Utilization of ML techniques to create transparent and explainable models. These models enhance our understanding of the built environment's features, significantly impacting the UHI phenomena. They shed light on the contributing factors and their thresholds. Therefore, considering these features and their threshold values can assist urban designers, architects, and policymakers in designing, redesigning, and improving the current state of the built environment to effectively mitigate heatwaves and the UHI effect in Tallinn (Paper III);
4. Presentation of a multi-objective optimization approach that considers the orientation of buildings and urban canyons to reduce sun exposure on the hottest summer days. This approach aims to enhance outdoor thermal comfort in the urban environment, mitigating heat-related issues. It holds significant implications for urban planning, sustainability, and the transformation of urban areas into spaces that offer greater thermal comfort and enjoyment for public use (Paper IV);

5. Introduction of a research framework for improving thermal comfort in urban areas. Focus on microclimate, human sensation, and sustainability goals, providing a holistic approach to urban climate management (Paper V);

The core objective of the CLEAR model extends beyond merely dissecting the intricacies of these climatic phenomena; it strives to emerge with insightful and actionable findings. By seamlessly integrating data-driven and model-based workflows, the CLEAR model facilitates an in-depth exploration of the root causes underpinning UHI and heatwaves in urban areas. Ultimately, the outcomes generated by the CLEAR model serve as a valuable guidepost, directing the formulation of pragmatic and sustainable solutions.

The potential applications of the results for architects and designers are as follows:

1. Data-driven approach in urban analysis:
One of the most significant potentials of applying the thesis is the insights gained from employing a data-driven approach and integrating ML into urban data analysis. This can provide architects and designers with profound insights into the intricate relationships within the built environment.

The analysis of current and historical urban data, supplemented with meteorological information, offers a holistic understanding of how the city and its elements interact under varying climatic and microclimatic conditions.

This comprehensive perspective enables architects and designers to grasp better how the built environment shapes and influences the city's microclimate, aiding in making informed design decisions.

2. Mitigation strategies for UHI effects and other climatic phenomena:
The CLEAR model goes beyond unraveling the complexities of urban climate challenges; it contributes to formulating mitigation strategies for reducing UHI effects in cities.

Architects and designers can use the insights gained from this model to design urban spaces and structures that actively address heat-related challenges while promoting sustainable urban growth and development.

3. Improving the built environment and enhancing urban conditions:
Models like the CLEAR model are potent tools for capturing, leveraging, and learning from urban data.

These models provide architects and designers with invaluable insights that can inform design solutions to enhance overall urban conditions.

By utilizing such models, architects and designers can contribute to creating more sustainable, livable, and comfortable cities for their inhabitants, effectively addressing challenges like urban heatwaves and the UHI effect.

4. Adaptability to diverse settings and relevance across locations:
The adaptability of the CLEAR model extends beyond Tallinn, encompassing diverse climatic, microclimatic, urban, and architectural scenarios.

This empowers architects, designers, and researchers to tailor the model and methodologies to different urban settings, aligning them with unique research questions and requirements.

Architects and designers can harness the benefits of the CLEAR model to address specific urban phenomena or issues in various locations.

This research highlights the pivotal role played by the CLEAR model in effectively addressing the UHI effect in Tallinn, Estonia, making it relevant to diverse urban settings

characterized by varying climatic conditions and different urban scenarios, fostering a global exchange of best practices in urban design and climate mitigation.

The main potential applications of the results for regulation makers are as follows:

1. Valuable dataset and predictive models:
This research provides regulation makers with a valuable geoprocessed dataset encompassing Tallinn's urban characteristics and meteorological data.
Equips them with predictive and interpretable ML models to help make data-driven decisions and policies.
Mitigate UHI and enhance thermal comfort and sustainable development:
The research extends strategies to enhance thermal comfort in urban areas.
Aims to reduce thermally affected urban areas, addressing challenges associated with the UHI effect.
Promotes the principles of sustainable urban development, aligning policies with long-term environmental and social goals.
2. Informed decision making:
The implications of this research empower regulation makers, urban planners, policymakers, and stakeholders to make informed decisions.
The insights and methods gained from this research can guide the development of policies and strategies to foster sustainability and enhance urban comfort.
3. Resilient urban centers:
The research equips regulation makers with tools to develop resilient urban centers.
Strategies derived from this research can aid in building cities that are better prepared to withstand the challenges of climate change and urbanization.
4. Global applicability:
This research represents a groundbreaking advancement in understanding the interplay among urban characteristics, meteorological data, and UHI phenomena.
Offers practical approaches and insights that have the potential for global applicability.
Serves as a guiding beacon for regulation makers worldwide to address UHI effects, champion sustainable urban development, and create urban environments that are both inviting and sustainable.

5.1   Future work

Looking ahead to future research, the path forward is brightly lit by the continued utilization of the findings from this thesis and the development of easily understandable ML models. These models will be instrumental in addressing issues related to the green transition and aligning with the ambitious goals of sustainable built environments.

Expanding the scope of the investigation is essential. This should involve a broader range of urban features and considerations for temporal dynamics. Incorporating climatic and microclimatic data from various dates and times will provide a more comprehensive view of urban thermal dynamics. Moreover, exploring qualitative approaches to capture data relevant to identifying thermal and non-thermal urban areas is of utmost importance.

Collaborative efforts and the cultivation of innovative methodologies will be key drivers of progress. These efforts hold the potential to deepen our understanding of

urban thermal environments and further fortify the foundations of sustainable urban development.

Future research will concentrate on harnessing data-driven methodologies and interpretable ML models. The aim is to bolster urban sustainability amid escalating environmental challenges spurred by rapid urbanization. By thoroughly analyzing geospatial data and applying ML techniques, the goal is to pinpoint the critical urban factors influencing heatwaves and to devise sustainable urban strategies that address them.

Furthermore, future endeavors may consider incorporating supplementary variables, temporal dynamics, and external influences to enhance the predictive capabilities of the models. Additionally, integrating qualitative data and user preferences could provide a holistic understanding of urban thermal comfort, facilitating user-centric design choices that promote well-being.

Continuing to employ innovative methodologies, such as Co-design and fostering collaborative initiatives in urban thermal environment research, holds immense potential for making substantial contributions to sustainable and resilient urban development.

By exploring diverse datasets and applying alternative techniques, we can advance our understanding of streamlining ML models using numerical features and uncover fresh insights into the complex dynamics of urban thermal environments. This ongoing research journey promises to pave the way for more environmentally friendly and comfortable urban spaces in our rapidly evolving world.

References

Aflaki, A., Mirnezhad, M., Ghaffarianhoseini, A., Ghaffarianhoseini, A., Omrany, H., Wang, Z. H., & Akbari, H. (2017). Urban heat island mitigation strategies: A state-of-the-art review on kuala lumpur, singapore and hong kong. *Cities, 62*, 131-145. doi:https://doi.org/10.1016/j.cities.2016.09.003

Akbari, H., Cartalis, C., Kolokotsa, D., Muscio, A., Pisello, A. L., Rossi, F., . . . M, Z. (2016). Local climate change and urban heat island mitigation techniques - The state of the art. *Journal of Civil Engineering and Management, 22*(1). doi:10.3846/13923730.2015.1111934

Aleksandrowicz, O., Vuckovic, M., & Kristina, K. M. (2017). Current trends in urban heat island mitigation research: Observations based on a comprehensive research repository. *Urban Climate, 21*, 1-26. doi:doi.org/10.1016/j.uclim.2017.04.002

Ali-Toudert, F., & Mayer, H. (2006). Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate. *Building and Environment, 41*(2), 94-108. doi:doi.org/10.1016/j.buildenv.2005.01.013.

Arnfield, A. (1990). Street design and urban canyon solar access. *Energy and Buildings, 14*(2), 117-131. doi:doi.org/10.1016/0378-7788(90)90031-D

Balogun, A. L., Tella, A., Baloo, L., & Adebisi, N. (n.d.). A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Climate, 40*(100989). doi:doi.org/10.1016/j.uclim.2021.100989

Belle, V., & Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data, 4*(688969). doi:10.3389/fdata.2021.688969

Bertinelli, L., & Black, D. (2004). Urbanization and growth. *Journal of Urban Economics, 56*(1), 80-96.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science, 39*(3), 497-524. doi:10.1111/0022-4146.00146

Building data of Tallinn. (n.d.). Retrieved from https://livekluster.ehr.ee/ui/ehr/v1/infoportal/buildingdata

Buildingdata. (n.d.). Retrieved from https://livekluster.ehr.ee/ui/ehr/v1/infoportal/buildingdata.

Car, A. (1997). *Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding.* Technical University, Vienna: PhD thesis, Department of Geoinformation.

Chakrabartty, S. (2014). Scoring and Analysis of Likert Scale: Few Approaches. *Journal of Knowledge Management and Information Technology*.

Ching, J. K. (2013). A perspective on urban canopy layer modeling for weather, climate and air quality applications. *Urban Climate, 3*. doi:doi: 10.1016/j.uclim.2013.02.001.

Chirag, D., & Ramachandraiah, A. (2010). The significance of Physiological Equivalent Temperature (PET) in outdoor thermal comfort studies. *International Journal of Engineering Science and Technology*, 2825–2828.

Cohen, B. (2004). Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. *World development, 32*(1), 23-51.

Cohen, P., Potchter, O., & Matzarakis, A. (2013). Human thermal perception of Coastal Mediterranean outdoor urban environments. *Applied Geography, 37*, 1-10. doi:doi.org/10.1016/j.apgeog.2012.11.001.

Cortes, A., Jesfel Rejuso, A., Ace Santos, J., & Blanco, A. (2022). Evaluating mitigation strategies for urban heat island in mandaue city using envi-met. *Journal of Urban Management, 11*(1), 97-106. doi:doi.org/10.1016/j.jum.2022.01.002.

Deb, C., & Ramachandraiah, A. (2010). The significance of Physiological Equivalent Temperature (PET) in outdoor thermal comfort studies. *International Journal of Engineering Science and Technology, 2*(2825-2828), 2825-2828.

Eagleson, S., Escobar, F., & Williamson, I. (2002). Hierarchical spatial reasoning theory and GIS technology applied to the automated delineation of administrative boundaries. *Computers, Environment and Urban Systems, 26*(2-3), 185-200. doi:doi.org/10.1016/S0198-9715(01)00040-0

Erell, E., Pearlmutter, D., & Williamson, T. (2012). *Urban microclimate: designing the spaces between buildings.* Urban Microclimate.

Esch, M. v., Looman, R., & Bruin-Hordijk, G. d. (2012). The effects of urban and building design parameters on solar access to the urban canyon and the potential for direct passive solar heating strategies. *Energy and Buildings, 47*, 189-200. doi:doi.org/10.1016/j.enbuild.2011.11.042

Eslamirad, N. (2023). *Geoprocess of geospatial urban data in Tallinn, Estonia, Mendeley Data*. doi:doi: 10.17632/2bm7kdf8gb.3

Eslamirad, N., De Luca, F., & Sakari Lylykangas, K. (2021). The role of building morphology on pedestrian level comfort in Northern climate. *In Journal of Physics: Conference Series, IOP Publishing, 2042*(1), 012053. doi:doi:10.1088/1742-6596/2042/1/012053

Eslamirad, N., De Luca, F., Ben Yahia, S., & Lylykangas, K. (2023). From near real-time urban data to an Explainable city-scale model to help reduce the Urban Heat Island (UHI) effect. *CISBAT 2023.* Lausane: Journal of Physics: Conference Series.

Eslamirad, N., De Luca, F., Sakari Lylykangas, K., & Ben Yahia, S. (2023). Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment. *Frontiers of Architectural Research, 12*(3), 541-555. doi:https://doi.org/10.1016/j.foar.2022.12.001.

Eslamirad, N., De Luca, F., Sakari Lylykangas, K., Ben Yahia, S., & Rasoulinezhad, M. (2023). Geoprocess of geospatial urban data in Tallinn, Estonia. *Data in Brief*, 109172. doi:doi.org/10.1016/j.dib.2023.109172.

Eslamirad, N., Sepúlveda, A., De Luca, F., & Sakari Lylykangas, K. (2022). Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus. *Energies, 15*(4). doi:doi.org/10.3390/en15041577

Farhadi, H., Faizi, M., & Sanaieian, H. (2019). Mitigating the urban heat island in a residential area in Tehran: Investigating the role of vegetation, materials, and orientation of buildings. *Sustainable Cities and Society, 46*, 101448. doi:doi.org/10.1016/j.scs.2019.101448.

Faria Neto, A., Inácio, B., Wurtz, F., & Delinchant, B. (2016). Thermal Comfort Assessment. *ELECON-ELECON – Electricity Consumption Analysis & Energy Efficiency.* doi:10.13140/RG.2.2.29416.67849

Fotheringham, A. S., & Brunsdon, C. (2010). Local Forms of Spatial Analysis. *Geographical Analysis, 31*(4), 340-358. doi:10.1111/j.1538-4632.1999.tb00989.x

Gabriele, B. D., Ocłoń, P., & De Lieto Vollaro, R. (2023). Effects of urban heat island mitigation strategies in an urban square: A numerical modelling and experimental investigation. *Energy and Buildings*, 112809.

General data of Tallinn. (n.d.). Retrieved from https://www.tallinn.ee/en/statistika/general-data-tallinn.

GeoJournal, P. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *61*(1). doi:doi: 10.1007/sgejo-004-0877-x

geopandas. (2022, 09). Retrieved from https://geopandas.org/en/stable/docs/reference/api/geopandas.GeoSeries. intersection.html.

Giampaoletti, M., Pistore, L., Zapata-Lancaster, G., Goycoolea, J., Janakieska, J., Gramatikov, K., . . . Fernandez, A. (2020). RESTORE Regenerative technologies for the indoor environment - Inspirational guidelines for practitioners.

Giridharan, R., & R, E. (2018). The impact of urban compactness, comfort strategies and energy consumption on tropical urban heat island intensity: A review. *Sustainable Cities and Society, 40*, 677-687. doi:doi.org/10.1016/j.scs.2018.01.024

Han, L., Zhao, J., Gao, Y., & Gu, Z. (2022). Prediction and evaluation of spatial distributions of ozone and urban heat island using a machine learning modified land use regression method. *Sustainable Cities and Society, 78*, 103643.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: springer.

Jing, & W Biljecki, F. (2022). Unsupervised machine learning in urban studies: A systematic review of applications. *129*, 103925.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert Scale: Explored and Explained. *British Journal of Applied Science and Technology, 7*, 396-403. doi:DOI:10.9734/BJAST/2015/14975

keskkonnateadlik hub. (n.d.). Retrieved from https://keskkonnateadlik-kaur.hub.arcgis.com/.

Kolokotsa, D., Lilli, K., Gobakis, K., Mavrigiannaki, A., Haddad, S., Garshasbi, S. H., . . . Santamouris, M. (2022). Analyzing the Impact of Urban Planning and Building Typologies in Urban Heat Island Mitigation. *12*(5), 537. doi:doi.org/10.3390/buildings12050537

Kovacs-Györi, A., Ristea, A., Havas, C., Mehaffy, M., Hochmair, H. H., Resch, B., . . . Blaschke, T. (2020). Opportunities and Challenges of Geospatial Analysis for Promoting Urban Livability in the Era of Big Data and Machine Learning. *ISPRS International Journal of Geo-Information, 9*(12), 752.

Leiqiu, H., & Brunsell, N. A. (2013). The impact of temporal aggregation of land surface temperature data for surface urban heat island (SUHI) monitoring. *Remote Sensing of Environment, 134*, 162-174.

Lin, J., Qiu, S., Tan, X., & Zhuang, Y. (2023). Measuring the relationship between morphological spatial pattern of green space and urban heat island using machine learning methods. *Building and Environment, 228*, 109910.

Lipton, Z. C. (2016). The Mythos of Model Interpretability. *Communications of the ACM, 61*(10). doi:DOI: 10.1145/3233231

Lipton, Z. C. (2016). The Mythos of Model Interpretability. *Communications of the ACM, 61*(10). doi:DOI: 10.1145/3233231

Luo, Y., He, J., & Ni, Y. (2017). Analysis of urban ventilation potential using rule-based modeling. *Computers, Environment and Urban Systems*, 13-22. doi:https://doi.org/10.1016/j.compenvurbsys.2017.07.005

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, New York, NY, USA, Association for Computing Machinery.*, 1135-1144.

Märtens, O., Pärg, R., & Sluzenikina, J. (2020). *Soojussaarte hindamine Eesti linnades 2014-2019.* Tallinn: keskkonnagentuur.

Milojevic-Dupont, N., & Creutzig, F. (2021). Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society, 64*, 102526.

Mondal, J. (2018). Debunking Thumb-rule: Orient Longer Faces of Buildings towards North-South to Minimise Solar Gain in India. *Neo-International Conference on Habitable Environments (NICHE 2018)At: School of Architecture and Design, Lovely Professional University, Delhi NCR, India.* Dehli.

Nakata-Osaki, C. M., Lucas Souza, L. C., & Souto Rodrigues, D. (2018). THIS – Tool for Heat Island Simulation: A GIS extension model to calculate urban heat island intensity based on urban geometry. *Computers, Environment and Urban Systems, 67*, 157-168. doi:doi: 10.1016/j.compenvurbsys.2017. 09.007.

Nazarian, N., Acero, J. A., & Norford, L. (2019). Outdoor thermal comfort autonomy: Performance metrics for climate-conscious urban design. *Building and Environment, 155*. doi:doi: 10.1016/j.buildenv.2019.03.028.

Neupane, B., Horanont, T., & Aryal, J. (2021). Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing, 13*(4), 808.

Onishi, A., Cao, X., Ito, t., Shi, F., & Imura, h. (n.d.). Evaluating the potential for urban heat-island mitigation by greening parking lots. Urban Forestry Urban Greening. *9*(4), 323–332. doi:doi.org/10.1016/j.ufug.20

Openshaw, S. (1977). A Geographical Solution to Scale and Aggregation Problems in Region-Building, Partitioning and Spatial Modelling. *Transactions of the Institute of British Geographers, 2*(4), 459-472. doi:doi.org/10.2307/622300

Páez, A. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal, 61*(53). doi:10.1007/sGEJO-004-0877-x

Papadakis, E., Adams, B., Gao, S., Martins, B., Baryannis, G., & Ristea, A. (2022). Explainable artificial intelligence in the spatial domain (X-GeoAI). *Transactions in GIS, 26*, 2413-2414. doi:10.1111/tgis.12996.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences. *Hydrology and Earth System Sciences, 11*(5), 1633-1644. doi:doi: 10.5194/hess-11- 1633- 2007.

Peng, M. (2009). Division of urban hierarchical grid based on hierarchical spatial reasoning. *International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, 7492*. doi:10.1117/12.838394

Renganathan, G., & Emmanuel, R. (2018). The impact of urban compactness, comfort strategies and energy consumption on tropical urban heat island intensity: A review. *Sustainable Cities and Society, 40*, 677-687. doi:doi.org/10.1016/j.scs.2018.01.024

Ribeiro, M. T., S, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, New York, NY, USA, Association for Computing Machinery*, 1135-1144.

Rizwan, A. M., Dennis, L. Y., & Chunho, L. (2008). A review on the generation, determination and mitigation of urban heat island. *Journal of environmental sciences, 20*(1), 120-128.

Rosenzweig, C. S.-L. (n.d.). *Climate change and cities: Second assessment report of the urban climate change research network.* 2018: Cambridge University Press.

Sagris, V., & Sepp, M. (2017). Landsat-8 tirs data for assessing urban heat island effect and its impact on human health. *IEEE Geoscience and Remote Sensing Letters, 14*. doi:doi: 10.1109/LGRS.2017.2765703

Shishegar, N. (2013). Street Design and Urban Microclimate: Analyzing the Effects of Street Geometry and Orientation on Airflow and Solar Access in Urban Canyons. *Journal of Clean Energy Technologies, 13*. doi:doi: 10.7763/JOCET.2013.V1.13

Steeneveld, G. J., Koopmans, S., Heusinkveld, B. G., Hove, L. W., & Holtslag, A. A. (2011). Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in The Netherlands. *Journal of Geophysical Research*. doi:116. 10.1029/2011JD015988

Ul Moazzam, M. F., Hoi Doh, Y., & Gul Lee, B. (2022). Impact of urbanization on land surface temperature and surface urban heat Island using optical remote sensing data: A case study of Jeju Island, Republic of Korea. *Building and Environment, 22*, 109368. doi:doi.org/10.1016/j.buildenv.2022.109368

Vásquez-Álvarez, P. E., Flores-Vázquez, C., Cobos-Torres, J.-C., & Cobos-Mora, S. (2022). Urban Heat Island Mitigation through Planned Simulation. *Sustainability, 14*(4), 8612. doi:https://doi.org/10.3390/su14148612

Wang, X., Song, Y., & Tang, P. (2020). Generative urban design using shape grammar and block morphological analysis. *9*(4). doi:10.1016/j.foar.2020.09.001

Wang, Y., Berardi, U., & Akbari, H. (2016). Comparing the effects of urban heat island mitigation strategies for toronto, canada. *Energy and buildings, 114*, 2-19.

Weather Underground. (n.d.). Retrieved from https://www.wunderground.com/about/our-company.

Weatherdata. (n.d.). Retrieved 04 20, 2023, from www.latlong.net. LatLong.net.

Xu, C., Chen, G., Huang, Q., Su, M., Rong, Q., Yue, W., & Haase, D. (2022). Can improving the spatial equity of urban green space mitigate the effect of urban heat islands? an empirical study. *Science of The Total Environment, 841*, 156687. doi:doi.org/10.1016/j.scitotenv.2022.156687.

Yahia, M. W., & Johansson, E. (2012). Influence of urban planning regulations on the microclimate in a hot dry climate: The example of Damascus, Syria. *Journal of Housing and the Built Environment, 28*(1). doi:28. 10.1007/s10901-012-9280-y.

List of figures

6 List of tables

Acknowledgements

I want to express my heartfelt gratitude to the following individuals whose unwavering support and guidance have been instrumental in my journey of completing this Ph.D. thesis, with each of them playing a vital role at every stage:

First and foremost, I am indebted to my dedicated supervisors, Dr. Francesco De Luca, Prof. Kimmo Sakari Lylykngas, and Prof. Sadok Ben Yahia, for their invaluable guidance, unwavering support, and insightful feedback. Your mentorship has been instrumental in shaping the course of my research. I want to express my deepest gratitude to dear Francesco. Learning from you and working alongside you has been an honor for me. We have shared visions, thoughts, and projects that I have truly cherished.

Additionally, I am fortunate to have been under the supervision of dear Kimmo. From the first day I visited him, he made me feel supported. He assured me that I could share any difficulties I encountered, and he would provide direct assistance or find a solution. Thank you for your unwavering support, guidance, and assistance. Special thanks to dear Sadok for accepting the role of my co-supervisor, for helping me, teaching me, and offering encouragement. You have been a mentor and a source of inspiration and motivation throughout this academic journey.

Special thanks to my beloved family, who encouraged me to embark on this journey and stood by me through all the ups and downs, even during moments of failure. I want to extend my heartfelt thanks, especially to my mother and father, whom I owe my life. I can only hope to repay a fraction of their boundless love. From my heart, I thank my lovely sister, Mona, for being an angel, always there to listen, assist, and provide unwavering encouragement. I am profoundly grateful to my dear brothers, Amir and Mehdi, for their constant encouragement, understanding, and love. Your belief in me has been a wellspring of strength throughout this endeavor. I love you all.

Many thanks to my valuable friend, Elnaz, for being an influential, supportive, and inspiring presence in my entire life and academic life.

I extend my heartfelt thanks to my dearest friend, Sara, who has not only been a friend but also like a kind and supportive family member. Your presence in my life has brought immense joy and positivity. I am also grateful to Francesco for his kind and generous support and to Abel, my co-author and friend, for always listening to my ideas and providing encouragement and collaboration. Thanks to other colleagues from the Architecture and Urban Studies group and the group of FINEST: Jarek Kurnitski, Ralf Martin, Fabian, Jenni, Epi, Kulli, Olli, Laura, Modupe, Dominik, and Ioannis. I would also like to express my gratitude for the administrative support I received from Triinu-Liis and Kärt. I appreciate my fellow co-author, Mahdi, whose brilliant insights and guidance have been instrumental in steering my research in the right direction.

I'm grateful to my Iranian friends who have given me a welcoming and secure environment to freely express myself and feel truly comfortable. I want to extend special appreciation to my friends Marmar, Marzieh, Navid, Vahideh, Shahram, and their lovely little one, Gisoo, for their boundless kindness and positivity.

In conclusion, I am truly humbled and grateful for the support and encouragement of all who have been part of this journey, and I am honored to have you in my life. I also extend my most profound appreciation to myself for never giving up and relentlessly pursuing my dreams, even in the face of challenges.

Nasim, October 2023

Abstract

Data-Driven Urban Modelling: The Case of Explainable Detection of Urban Heat Island

The imperative to create a more sustainable society underscores the need for a profound comprehension of the intricate environmental dynamics inherent in urban systems. As urbanization progresses, cities struggle with escalating environmental challenges that deeply impact human well-being. These changes in urban climate, accompanying the urbanization process, can lead to shifts in microclimates and the degradation of thermal environments. The situation is further worsened by Urban Heat Islands (UHI) and the rise in cities' temperatures. The adverse effects of UHIs extend to health, quality of life, and urban livability, emphasizing the critical necessity of developing effective mitigation strategies to amplify outdoor thermal comfort and elevate the overall quality of life within cities.

Moreover, this research endeavor aspires to generate crucial insights into how urban environments can be thoughtfully designed and strategically configured to enhance sustainability, livability, and outdoor thermal comfort for pedestrians. The study's findings empower researchers to identify efficacious interventions by leveraging UHI data, geoprocessing techniques, and data-driven approaches. This comprehensive evaluation serves as a guideline, directing urban planners and architects towards informed mitigation solutions for heatwaves and thermal urban areas to elevate thermal comfort within cities while also cultivating conditions aligned with approved thermal comfort levels.

At the heart of this study lies the theme of sustainable urban development. Employing generative and predictive Machine Learning (ML) models, which create new data samples, explainable ML models that provide insights into decision-making, and interpreting the models to make them easy to understand played vital roles as features in the study of UHI assessment in Tallinn. Utilizing ML empowers this research to establish meaningful connections between urban design, sustainability, and resident well-being, driving the exploration of spatial data and ML techniques to create energy-efficient and enduringly sustainable urban environments. This endeavor ultimately contributes to urban well-being and energy efficiency, highlighting ML's role in understanding urban design's profound impact on sustainability and overall quality of life, thus catalyzing the creation of more comfortable and sustainable urban spaces.

Papers I through V have meticulously covered different facets of the research journey, spanning from the initial data collection phase (Paper I) to the subsequent development of machine learning algorithms (Papers II and III). The culmination of this work is observed in the practical application of the study, where it is used to evaluate heatwaves and thermal conditions in urban environments, ultimately guiding the optimization of urban areas to enhance thermal comfort (Paper IV and Paper V). The methodology, built upon the CLEAR model, comprises four primary steps:

Data Capture (C): In this initial step (Paper I), urban data is systematically captured.

Learning (L) (Papers II and III) involves leveraging this captured data to develop learning algorithms and create various ML models. This encompasses the development of predictive and explainable models.

Rule Extraction (E): Rules are extracted from these ML models to formulate a predictive ML model as the explainable model for understanding and anticipating UHI phenomena.

Application (A): The knowledge and insights gained from the research, along with the captured dataset, are employed to propose effective mitigation strategies (Paper IV and Paper V) to reduce heatwaves and mitigate the UHI effect in urban environments.

Moreover, an iterative approach ensures that the model can be repeated and adapted for application in diverse urban scenarios and climatic conditions, enhancing its versatility as a UHI analysis and mitigation tool.

Effectively countering the UHI effect and fostering enhanced urban spaces necessitates the availability of an accurate dataset incorporating critical urban environmental features, location indices, meteorological data, and UHI levels. The case study of the thesis is Tallinn, Estonia, where an all-encompassing approach encompassing spatial data capturing (Paper I), assessment, preparation, and the development of a learning algorithm based on ML and data-driven approaches (Paper II) were undertaken to serve as a foundation for an explainable UHI model (Paper III) based on ML models. The model integrates essential building and urban features to craft strategies to mitigate UHI levels, thereby contributing to broader urban sustainability objectives (Paper IV and Paper V). Noteworthy transformations in building and urban features have emerged as influential contributors, effectively reshaping areas from UHI absence to marked presence characterized by heatwaves. These insights provide a pivotal foundation for architects, urban planners, researchers, policymakers, and city administrations involved in built environment projects.

Integrating data-driven approaches, as discussed in Papers II and III has emerged as a pivotal strategy for fostering sustainability within urban analysis. By leveraging geoprocessing and ML techniques, this study addresses challenges such as UHI, uncovering insights into configuring urban environments to enhance sustainability, livability, and outdoor thermal comfort (Paper IV and Paper V).

The study's achievements are underscored by the geospatially processed dataset, which, after preprocessing and labeling with UHI levels, serves as training data for a diverse range of ML models.

The data-driven approach and derived equation shed light on effective mitigation strategies (Paper IV) and promote sustainable urban development within Tallinn while providing valuable guidance for diverse locations characterized by varying climates. Furthermore, this study underscores the significance of designing urban canyons with optimal orientation to mitigate the UHI effect effectively, thereby maximizing outdoor thermal comfort during hot summer days. The strategic optimization of building orientation further fosters shaded and cooler pedestrian areas, ultimately leading to reduced surface temperatures and more comfortable and sustainable urban environments.

The study's findings emphasize the pivotal role of architectural and urban design choices in molding the city's thermal environment, providing valuable insights for prioritizing specific features to optimize thermal comfort and decrease the number of heated urban areas.

Lühikokkuvõte

Andmepõhine linnamodelleerimine: soojussaarte seletatav tuvastamine

Hädavajadus luua säästlikum ühiskond rõhutab sügava mõistmise olulisust seoses linnasüsteemide keeruliste keskkonnamuutuste dünaamikaga. Urbaniseerumise edenedes seisavad linnad silmitsi üha suurenevate keskkonnaväljakutsetega, mis mõjutavad sügavalt inimeste heaolu. Nii linna kliima kui ka urbaniseerumise protsessi muutused võivad viia mikrokliima muutusteni ja halvendada soojuskeskkonda. Probleemi teravdab veelgi linnade soojussaarte efekt ja linnasisese temperatuuri tõus. Linnade soojussaared avaldavad negatiivset mõju elanike tervisele ja elukvaliteedile, rõhutades tõhusate leevendusstrateegiate hädavajalikkust, et suurendada välitingimustes viibivate inimeste soojusmugavust ning parandada linnaelanike üldist elukvaliteeti.

Lisaks püüab käesolev uurimistöö luua olulist arusaamist sellest, kuidas linnakeskkondi saab terviklikult ja strateegiliselt kujundada, et suurendada jätkusuutlikkust, elukvaliteeti ning jalakäijate õues viibimise ajal tuntavat soojusmugavust. Uuringu tulemused võimaldavad teadlastel tuvastada tõhusaid sekkumisi, kasutades UHI andmeid, geotööluse tehnikaid ja andmepõhiseid lähenemisviise. See ulatuslik hindamine toimib juhendina, suunates linnaplaneerijaid ja arhitekte teadlikult kuumalainete ja termiliste linnapiirkondade leevenduslahenduste suunas, et tõsta linna soojusmugavust, samal ajal luues tingimused, mis on kooskõlas heakskiidetud soojusmugavuse standarditega.

Selle uuringu keskmes on jätkusuutlik linnakeskkond. Masinõppe (ML) generatiivsed ja ennustavad mudelid mängisid olulist rolli Tallinnas läbi viidud uuringus, mis hindas linna soojusmugavust. ML mudelid loovad uusi andmestikke, selgitavad ning tõlgendavad ML-mudelid, mis aitavad otsuste tegemisel ja mudelite tõlgendamisel nende lihtsaks mõistmiseks. ML kasutamine võimaldas selle uuringu raames luua tähendusrikkaid seoseid linnakujunduse, jätkusuutlikkuse ja elanike heaolu vahel, suunates ruumiandmete ja ML-tehnikate uurimist energiatõhusate ning jätkusuutlikkust energiatõhusalt tagava linnakeskkonna loomise suunas. See ettevõtmine aitab lõppkokkuvõttes kaasa linna heaolule ja energiatõhususele, rõhutades ML rolli linnakujunduse mõistmisel ning selle sügava mõju mõistmisel jätkusuutlikkusele ja üldisele elukvaliteedile. Selle lähenemisviisi tulemusena innustatakse looma mugavamaid ning jätkusuutlikke linnaruume.

Artiklid I kuni V on põhjalikult kajastanud erinevaid uurimise aspekte, ulatudes algse andmete kogumise faasist (Artikkel I) masinõppe algoritmide edasisse arendamisse (Artiklid II ja III). Selle töö tulemusena on praktilises uurimise rakenduses näha, kuidas seda kasutatakse kuumalainete ja termiliste tingimuste hindamiseks linnakeskkonnas, suunates lõpuks linnaalade optimeerimist termilise mugavuse parandamiseks (Artikkel IV ja Artikkel V). Metoodika, mis põhines CLEAR mudelil, koosneb neljast põhipunktist:

Andmete kogumine (C): Selles algstaadiumis (Artikkel I) kogutakse linnandmeid süstemaatiliselt.

Õppimine (L) (Artiklid II ja III) hõlmab nende kogutud andmete ärakasutamist, et arendada õppimisalgoritme ja luua erinevaid ML-mudeleid. Selle hulka kuulub ennustavate, selgitatavate ja tõlgendatavate mudelite arendamine.

Reeglite ekstraheerimine (E): Reeglid eraldatakse nendest ML-mudelitest, et sõnastada ennustava valemi loomiseks arusaamine ja UKS nähtuse ettearvamine.

Rakendamine (A): Uurimusest saadud teadmisi ja arusaamu kasutatakse koos kogutud andmestikuga tõhusate leevendusstrateegiate ettepanekute tegemiseks (Artikkel IV ja V) kuumalainete vähendamiseks ja UKS mõju leevendamiseks linnakeskkonnas.

Lisaks tagab korduv lähenemine selle, et mudelit saab korrata ja kohandada rakendamiseks erinevates linnakeskkondades ja kliimatingimustes, suurendades selle mitmekülgsust kui UKS analüüsi ja leevendusvahendit.

UHI mõju tõhusaks vastandumiseks ja paremate linnaruumide arendamiseks on vajalik täpsete andmete kättesaadavus, mis hõlmab olulisi linnakeskkonna omadusi, asukohaindekseid, meteoroloogilisi andmeid ja UHI taset. Uuringu juhtumianalüüsiks on Tallinn, Eesti, kus läbivõetud on kõikehõlmav lähenemine, mis hõlmab ruumiandmete kogumist (Artikkel I), hindamist, ettevalmistamist ja õppimisalgoritmi arendamist ML ja andmetepõhiste lähenemiste põhjal (Artikkel II) aluseks selgitatavale UHI mudelile (Artikkel III) ja ML mudelitele põhinevale tõlgendatavale UHI mudelile. Mudel integreerib olulisi hooneid ja linnakeskkonna omadusi, et kujundada strateegiaid UHI taseme leevendamiseks, andes seeläbi panuse laiematele linnajätkusuutlikkuse eesmärkidele (Artikkel IV). Hoone ja linnakeskkonna omadustes on märgatavaid muutusi, mis on tõhusad tegurid, mis muudavad piirkonnad UHI puudumisest kuumalainetele iseloomulikuks. Need teadmised pakuvad olulist alust arhitektidele, linnaplaneerijatele, teadlastele, poliitikakujundajatele ja linnajuhtidele, kes osalevad ehitatud keskkonnaprojektides.

Andmepõhiste lähenemisviiside integreerimine, nagu seda on arutatud Artiklites II ja III, on osutunud keskkonnauuringutes jätkusuutlikkuse soodustamiseks oluliseks strateegiaks. Geoprotsesside ja ML-tehnikate ärakasutamisega tegeleb see uuring väljakutsetega, nagu UHI, tuues esile teadmisi sellest, kuidas kujundada linnakeskkondi jätkusuutlikkuse, elukvaliteedi ja õues termilise mugavuse suurendamiseks (Artikkel IV and Artikkel V).

Uuringu saavutusi rõhutab geoprotsesside kaudu töödeldud andmestik, mis pärast eeltöötlemist ja märgistamist UHI tasemetega toimib mitmekülgsete ML-mudelite koolitusandmetena.

Andmepõhine lähenemine ja saadud võrrand heidavad valgust tõhusatele leevendusstrateegiatele (Artikkel IV) ja soodustavad jätkusuutlikku linnakeskkonna arengut Tallinnas, pakkudes samas väärtuslikku juhendamist erinevatele asukohtadele, mis on erinevate kliimatingimustega. Lisaks rõhutab see uuring urbanistliku kanjoni optimaalse orienteerituse tähtsust UHI mõju tõhusaks leevendamiseks, suurendades seeläbi õues termilist mugavust kuumadel suvepäevadel. Hoone orienteerituse strateegiline optimeerimine soodustab varjulisi ja jahedamaid jalakäijate alasid, viies lõpuks kaasa madalamad pinnatemperatuurid ning mugavamad ja jätkusuutlikumad linnakeskkonnad.

Uuringu tulemused rõhutavad arhitektuuri- ja linnakujunduse valikute võtmerolli linna soojuskeskkonna kujundamisel, pakkudes väärtuslikke teadmisi konkreetsete omaduste prioriteetide seadmiseks termilise mugavuse optimeerimiseks ja kuumade linnapiirkondade arvu vähendamiseks.

Appendix

Paper I
Eslamirad, N., De Luca, F., Sakari Lylykangas, K., Ben Yahia, S. & Rasoulinezhad, M. (2023). Geoprocess of geospatial urban data in Tallinn, Estonia, *Data in Brief, 48*, 109172. ISSN 2352-3409, https://doi.org/10.1016/j.dib.2023.109172

Data Article

Geoprocess of geospatial urban data in Tallinn, Estonia

Nasim Eslamirad [a,*], Francesco De Luca [b],
Kimmo Sakari Lylykangas [b], Sadok Ben Yahia [c],
Mahdi Rasoulinezhad [b]

[a] FinEst Centre for Smart Cities, Tallinn University of Technology, Tallinn, Estonia
[b] Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn, Estonia
[c] Department of Software Science, Tallinn University of Technology, Tallinn, Estonia

ARTICLE INFO

ABSTRACT

The new digital era brings increasingly massive and complex interdisciplinary projects in various fields. At the same time, the availability of an accurate and reliable database plays a crucial role in achieving project goals. Meanwhile, urban projects and issues usually need to be analyzed to support the objectives of sustainable development of the built environment. Furthermore, the volume and variety of spatial data used to describe urban elements and phenomena have grown exponentially in recent decades. The scope of this dataset is to process spatial data to provide input data for the urban heat island (UHI) assessment project in Tallinn, Estonia. The dataset builds the generative, predictive, and explainable machine learning UHI model. The dataset presented here consists of multi-scale urban data. It provides essential baseline information for (i) urban planners, researchers, and practitioners to incorporate urban data in their research activities, (ii) architects and urban planners to improve the features of buildings and the city, considering urban data and the UHI effect, (iii) stakeholders, policymakers and administration in cities implementing built environment projects, and supporting urban sustainability goals. The dataset is available for download as supplementary material to this article.

---

* Corresponding author.
  E-mail address: nasim.eslamirad@ttu.ee (N. Eslamirad).

Specifications Table

Subject	Architecture
Specific subject area	Architecture and urban planning, built environment, urban heat island (UHI) assessment
Type of data	Vector data Table
How the data were acquired	Data were acquired via geoprocessing, programming, and analysis. Applying an ascending hierarchical grid system is based on the theory of dynamic urban heterogeneity and considers data schema, features, and location. Data were processed using Python programming packages and the QGIS Tool for geoprocessing and analysis.
Data format	Raw and analyzed
Description of data collection	An extensive multidisciplinary presented dataset is collected with 34, 001 building samples (raws) and 30 features (columns) in Tallinn, including location, building characteristics, urban characteristics, UHI data, and climate data. The current work methodology proposes a framework to categorize data into homogeneous or heterogeneous, static or dynamic schemes, and then collect data considering the homogeneous grid system. Implementing the hierarchical grid system in the data collection process helps create a spatial index for each object and connects the objects to the grid system. Second, use the homogeneous ground to define urban indices mainly anchored in the heterogeneous data. The methodology uses the Python, Numpy, Pandas libraries, the Geopandas package, and QGIS Tool. The approach helps to capture urban data from GIS resources, taking into account the location, general characteristics, and other spatial properties of urban elements.
Data source location	• City: Tallinn, Country: Estonia • Latitude and longitude for collected data: 59.436962 and 24.753574 [1] • Domain size (km2): 159 km2 area [2]
Data accessibility	https://geoportaal.maaamet.ee/eng/Spatial-Data/Cadastral-Data-p310.html https://veeb.tallinnlv.ee/pilv/index.php/s/gdqlbwwbT7Ocqv3 Repository name: Mendeley Direct URL to data: https://data.mendeley.com/datasets/gwpbktrx9g/1 Direct URL to images: https://data.mendeley.com/datasets/xm92bw2f49/1 Direct URL to codes: https://github.com/maraso-TTU/Urban-Data

Value of the Data

- The usefulness of the presented dataset lies in its ability to provide a comprehensive understanding of the urban environment by incorporating various essential features related to buildings, their surroundings, and meteorological and climatic data, with a specific focus on the UHI effect.
- The data can benefit a wide range of stakeholders, including architects, urban planners, policymakers, and local authorities, who can use the information to monitor urban data, set mitigation strategies, and improve the urban quality and quality of life of future development projects.
- Moreover, the data can be reused for further insights and development of experiments in various research fields, such as environmental studies, urban sustainability, and built environment studies, to understand better the UHI effect and its impact on the urban environment. The presented workflow and method can also be replicated in other comparative studies to collect and analyze relevant spatial data.
- Overall, not only the presented dataset has significant potential to contribute to the development of sustainable and liveable cities by providing detailed information on the urban

environment, which can inform decision-making and policy development, but also the presented workflow and method allow researchers to extrapolate building and city data, including location-based features and metrological conditions with urban data and UHI intensify.

1. Objective

Geographic information systems (GIS) are now widely used, data-intensive disciplines for managing, displaying, and developing data queries on maps. One of the biggest challenges for urban planners remains the transformation of data into knowledge to facilitate planning efforts in addressing issues of complex urban systems, which requires advanced interdisciplinary analytical methods [3]. Therefore, there is a constant need for efficient methods, analytical tools, and rational design methods. GIS is characterized by various spatial and numerical interactions of geographical objects compared to other computational tools that can present data in tabular form [4].

In the data capturing method, some solutions, such as hierarchical principles, widely used in various disciplines to split complex problems into smaller ones, are applied to consider heterogeneous urban data in a homogeneous base. For the first time, the hierarchical system is defined as Hierarchical Spatial Reasoning (HSR) based on Car's spatial theory in 1997 [5]. Hierarchical principles allow the Part-Whole system to apply to all elements within the hierarchical system, forming a part of the whole [5]. Each level connects the lower and the higher levels [5]. Furthermore, spatial heterogeneity is often assumed to result from large-scale regional effects or administrative subdivisions that narrow the scope of a process [2].

2. Data Description

The initial available data are divided into three groups: Urban, Weather and climate, and UHI. The urban data are mainly downloaded from the Tallinn Land Authority Geospatial Data Portal (Maaamet), owned by the Estonian state, local governments, and legal persons governed by public laws [6]. Data formats are Spatial data (ESRI shapefile), Shapefile (.SHP), and Shape Index file (.SHX) and indicate data of buildings and land information, green areas, landscape, streets, blue bodies, and details of facilities and infrastructures in the city of Tallinn. For example, according to the Data Portal of Tallinn (Maaamet), the number of buildings located in Tallinn is 67, 113 [6].

The data relating to Surface Urban Heat Island (SUHI) is the output of the study that used Landsat-8 images suitable for the UHI effect studies and Land Surface Temperature (LST) data [7]. The authors used Landsat-8 to assess the heat wave's impact on Estonian cities and explore the extent and magnitude of the UHI effects.

Furthermore, the air temperature used to define the threshold of heat waves in Estonia is an unusually high air temperature that lasts at least several days. The maximum air temperature was measured at 30°C and over. Therefore, the dangerous air temperature was defined as more than 27°C that lasted several days in Estonia [7].

We also produce the UHI effect data for Tallinn, which can be successfully used for adaptation measures. The intensification of UHI in Tallinn is 5°C according to the study that assessed Landsat-8 images to acquire SUHI. Thus, the threshold of the UHI data is 30-35°C, 35-40°C, and more than 40°C [7].

According to the date of the UHI effect in Tallinn, the related weather data was added to the dataset.
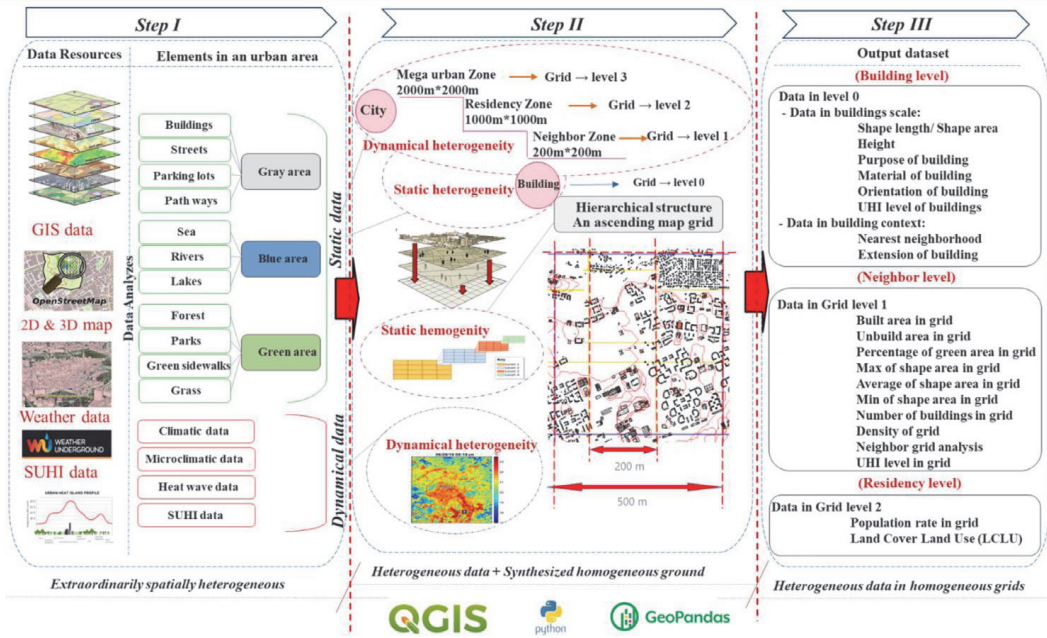
Fig. 1. The graphical abstract.

2.1. Data Location Description

The location of spatial data is Tallinn, the capital city of Estonia (Fig. 2). The latitude and longitude coordinates are 59.436962 and 24.753574 [1]. The geographical information of Tallinn is 445, 005 population and 159 km^2 area. The number of city districts in Tallinn is 8 [2]. Moreover, Tallinn is characterized by a humid continental climate with cold winters, according to the Köppen-Geiger classification Dfb [8].

The considered features of buildings are described in Tables 1 and 2. The focused features of buildings in Tallinn are general and technical characteristics. The main features of buildings are the purpose of use, material, absolute height, height, area, length, and a number of floors.

Table 1
The description of building data based on Open data information and instructions [9].

| Name of features	Description
General data of the building	
Object ID	The ID of each building is a unique combination of numbers
Purpose of use	The purpose of use with the largest area (Table 3)
General technical data of the building	
Material	The main material of the construction: Stone, wood, metal, composite material, and stone, composite material and metal
Absolute height (m)	The highest point of the highest structure
Height (m)	The largest vertical dimension of the building from the ground or pavement immediately surrounding the buildings to the highest point of the highest structure of the building, without taking into account local smaller depressions and elevations
Area(m2)	All building areas are in common use by residential and non-residential users
Length (m)	The length of the shape of the building
Number of floors	The horizontal plane in a building, on which it is possible to use the building according to its purpose
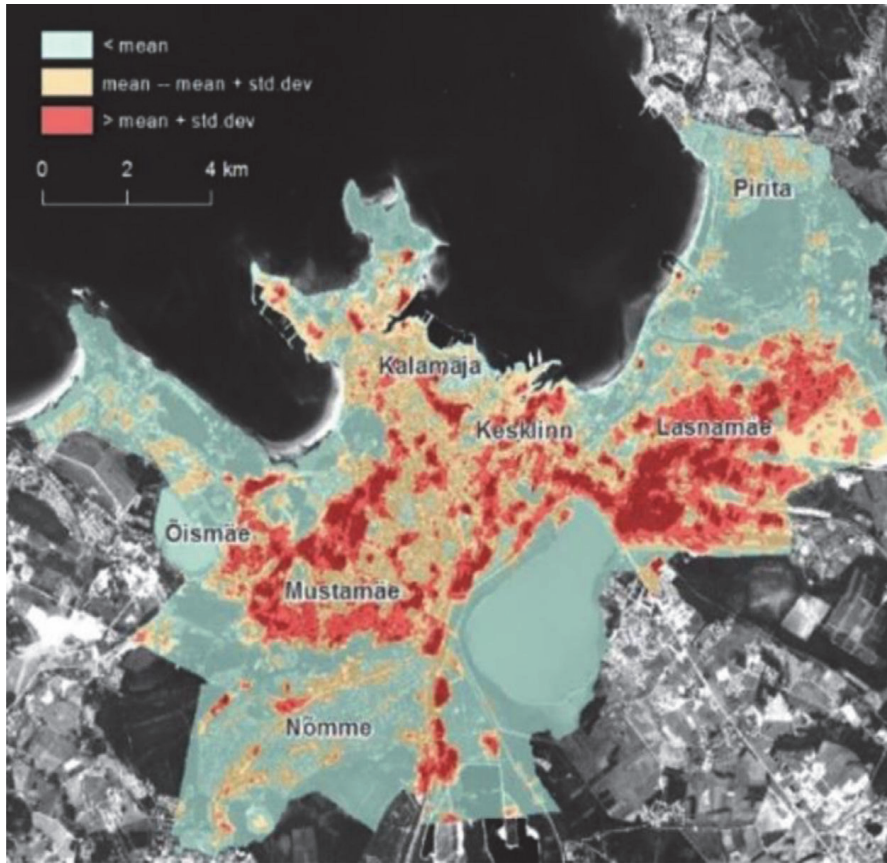
Fig. 2. SUHI map of Tallinn, July 25, 2014, 12:30 Estonian summer time (9:30 GMT). The green areas are outside of SUHI, the yellow areas are the areas over the mean SUHI, and the red areas are "inside" the SUHI (over mean + std.dev) [7].

Table 2
Purpose of use of buildings.

The purpose of buildings	Description
Residential building	Buildings for dwelling purposes
Public building	Buildings for particular public or common use (such as schools, shopping centers, banks, and offices)
Outbuilding	Buildings that are not used for residential purposes that are not industrial facilities, and not for public use
Underground building	Building with no floors above ground
Industrial facilities	Building for manufacturing and production processes, or a warehouse; a building not for public use
Underground storage space	Storage building with no floors above ground
Parking facilities	Underground garage or parking lot for cars

Furthermore, the weather data is related to the days of UHI value in the summer of 2014-2019. The weather data resource is the underground website [10]. The data of UHI downloaded from the Environment Agency is based on national monitoring data collected by the Environment Agency and analyses carried out on Landsat-8 (USA) satellite data (UHI 2014-2019) [11]. For example, Fig. 3 shows the heat map of the UHI on 25 July 2014. All UHI data used are Shapefiles in the format of (.SHP and .SHX files). The data shows UHI value in Tallinn is categorized into three levels: lower than 30°C (29°C), 30°C and 35°C.
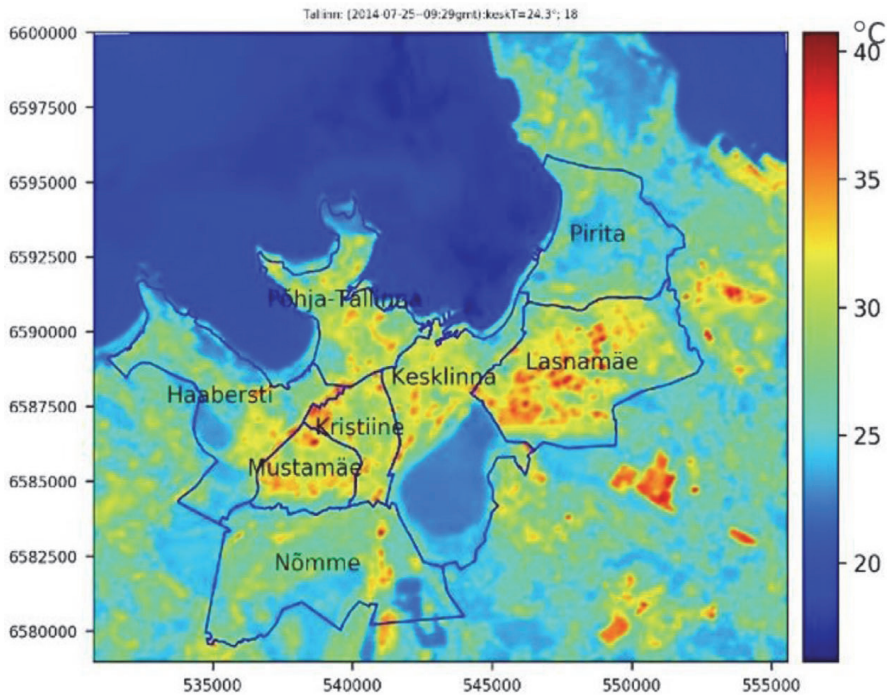
Fig. 3. Heat map of UHI, Tallinn July 25, 2014, UHI map of in Tallinn, July 25, 2014, 12:30 Estonian summer time (9:30 GMT) [7].

3. Experimental Design, Materials and Methods

The framework is tested in an urban assessment practice to collect geospatial data of buildings and their context on the one hand and data related to the UHI phenomenon to perform geoprocessing of the data on the other hand. The geodata is integrated into the latest version of the QGIS Tool (QGIS 3.22.11 'Białowieża', 3.26.x Buenos Aires, Ubuntu), and the data analysis is implemented in Python 3 environment in Jupiter notebook interface. To generate the dataset, a three-step methodology is developed:

3.1. First Step: Resources of Urban Data

The first step is to consider the available resources of urban data and the schema of all data types. This part of the process starts with defining urban data as the subject of data collection from different complex systems and resources.

The spatial information available in the public domain is downloaded and imported from the Tallinn Geoportal and Geodata [6,9,11].

3.2. Second Step: Urban Heterogeneity, Homogeneity, and the Ascending Hierarchical Grid System

The second step refers to the data recognition and categorizing of the data schema into homogeneous or heterogeneous and static or dynamic, then implementing the hierarchical grid system in the environment of the QGIS Tool.

The initial available data combine two components of the schema: static heterogeneity data and dynamic heterogeneity data.

Static heterogeneity means that elements in an urban area, like buildings, streets, and the natural landscape, remain unchanged or change little over time.

Dynamic heterogeneity refers to frequently changing data, such as climate, microclimate, and UHI data.

Here, when data collection does not consider the context of buildings, it is called static heterogeneity. However, when the data is based on urban elements at a larger scale, i.e., considering objects in an urban area in the context, such as characteristics of buildings in their neighborhood, the data is called dynamic heterogeneous data.

The aim of creating the hierarchical grid system is:

First, a spatial index is created for each object on the map, and the objects are referenced to the underlying grids. The grid system aids in analyzing extensive spatial data by effectively dividing the areas of the earth into identifiable grid cells.

Secondly, use the homogeneous ground to define urban indices mainly related to the heterogeneous data, which means accommodating both static and dynamic data in the solid hierarchical grid system.

This means that the hierarchical structure provides a solid basis for collecting data based on their locations. Thus, each grid is characterized by homogeneous static data in one layer and includes all points, geometries, and other recognized elements below the grid. For example, in the definition of built-up area, the total buildings' area located in grid level 1 is considered (Table 3).

Table 3
Defined indices, using a homogeneous hierarchical structure system and heterogeneous urban data.

Index	Grid level	Description
Building data	Grid 200 m*200m level 1	Buildings in 29°C, 30°C and 35°C UHI level
Built up area	Grid 200 m*200m level 1	B = Sum of buildings' area located in grid level 1
Density	Grid 200 m*200m level 1	D = Sum of all buildings' area located in grid level 1/40000
Green area	Grid 200 m*200m level 1	G = Sum of the area of green spaces, located in grid level 1

In this work, the hierarchical grid system in QGIS Tool represents polygons covering all parts of the map of Tallinn in three independent layers. The polygons are defined in three scales of geometric grids in a Shapefile format, as Fig. 4 shows.
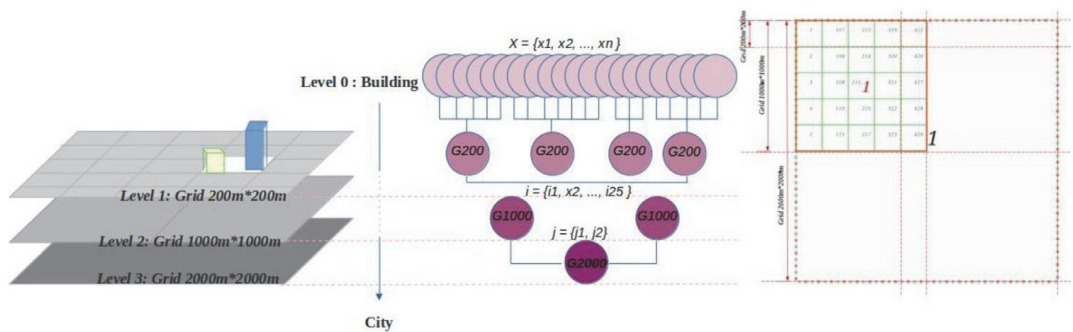

Fig. 4. The defined ascending hierarchical grid system.

As mentioned earlier, the data collection benefits from the hierarchical grid system and analyzes the spatial data of the target area on the map. Each level of the grid system has an ascending hierarchical structure, starting with the smallest scale, the square with a size of 200 meters. The second level of the grid system is a square with a dimension of 1, 000 meters in 1, 000 meters, while the largest level is a square with a dimension of 2, 000 meters in 2, 000 meters. The ascending grids are respectively called from levels 1 to 3:

Neighbourhood Zone, Residential Zone, and Mega Urban Zone, while the base level is the building level (level 0) as the smallest scale (Fig. 4)

The set X= x_1, x_2, . . ., x_n} is the building objects in the building dataset (level 0). First, this set is aggregated to the defined ascending hierarchical grid system. For example, x_i is the building located in the i th grid of level 1, so the i th grid is aggregated by the j th grid of level 2 and the k th grid of level 3. Then the spatial index of each building is appended to the dataset to determine the exact position of the building in the ascending hierarchical grid system. For example, Fig. 4 shows that each grid unit of level 3, a grid of 2, 000 meters by 2, 000 meters contain four squares of level 2 grids of 1, 000 meters by 1, 000 meters, and at the same time includes 25 squares of level 1 at 200 meters dimension.

3.3. Third Step: Object Detection and Index Definition

The last step of this section is related to the processing, arranging, and collecting data. Here we have used Python libraries and packages on the one hand and benefited from the geoprocessing attributes and expressions of the QGIS Tool, which have helped us develop methods for data collection in the different grid levels, starting from level 0 to level 3.

The hierarchical structure system is used to collect data related to the level of the grid system and defined indices mentioned in Table 3.

3.4. Experimental Methods to Capture Data

This subsection summarizes the approach taken to develop the geospatial dataset.

Data acquisition I: object detection in the hierarchical grid system

This part aims to show how the UHI value of each building is detected. The next step is completing the dataset by appending more details about buildings. As we mentioned, UHI intensity levels in the main levels of SUHI are:

Lower than 30°C (29 °C);

30 °C;

And 35 °C [7,6].

Moreover, this part's data collection process is done in Grid 2, Level 1 of the hierarchical grid system.

According to the geometric classes of objects in the QGIS Tool, we are dealing with polygons when the objects that should be detected are buildings. Moreover, objects located in the same places can be tracked based on the coordinate reference system (CRS) as a reference system to define the location of features in space [12]. When features are retrieved from a layer, the associated geometries have coordinates in the CRS layer. Therefore, objects on the map on the same CRS have intersections. In the code we developed in Python, we used this relationship to list buildings in each UHI area. The list contains buildings with unique identification codes in the specific ID of UHI. Furthermore, these two layers are in the area covered by the hierarchical grid system.

The codes are developed and executed in Python in the Jupiter Notebook interface. The used libraries are Pandas and Numpy to import the data as a Shapefile and process the imported data. Next, the Geopandas package is imported into the Notebook. The next is object detection, where we used GeoSeries.intersection attributes from the Geopandas package.

Since we need the identification code of each building to append other features in the next step, the method should use the unique code of buildings (ObjectID in the initial data), then find out the intersected geometries and return the ID of the building and the UHI map. Furthermore, in this method, we need to use more than one feature of each object. Therefore, we should consider a function to iterate pairs. For example, the zip () function in Python returns a zip

object that is an iterator of tuples where the first element in each passed iterator is paired. Then the second element in each passed iterator is paired.

The for-loop function is used to iterate over a sequence of tuples. The for loop can execute a series of statements, once for each element in the tuple. The "for loop" is repeated twice to build up the dataset and the UHI data with the zipped iterated objects.

The following codes use the intersection attributes of Geopandas, Python's zip function, and find the intersection between two called geometries of the two datasets and the ID of the building that has intersected with the UHI values. Considering the UHI value and the date, we could collect data from buildings with different heat values in the city. Moreover, the codes to determine which building is under which grid are similar to this code.

The codes are given below:

```
for r, s in zip (Buildingdata.geometry, Buildingdata.ID): for v, o
in zip (UHI.geometry, UHI.ID):
  Intersection_area = r.buffer(0).intersection(v.buffer(0)) if not
intersection_area.is_empty:
  p = gpd.GeoSeries(intersection_area) print(f"{int(o)} has
intersection with {s}")
```

Data acquisition II: the nearest neighbor

Another important feature in understanding the urban area's density and the city's compactness is the distance between buildings. We aim to find the shortest Euclidean distances between buildings at the neighborhood scale. We can use the spatial index defined in the previous step to determine the size faster than looping through the data frame and then finding the minimum of all distances when working with a large data frame.

As in the previous code, we used the zip () function to return the ID of each building and understand the pair of neighboring buildings and the for-loop function in Python, as well as the distance attributes of the Geopandas package.

According to the codes, the output will be the list of distances, starting with 0 as the minimum distance, i.e., the distance of each building to itself, and the second member of the list is the nearest building. Then we use the list of minimum distances and call the second member (element number of the second item = 1 since the first element is the 0th item in the list).

The codes can also find the distance to other boundaries for any geometry or the furthest neighbor. Furthermore, the method can be useful to find the nearest street, green space, or other location as a geometry, point, or line, even in two different data resources. Fig. 5 shows how the nearest neighbor is defined. First, each building is set as a target geometry to calculate the distance between the outer wall of the building. The distance is the shortest line between two objects. For each x_n member of the building dataset, the distance to (n-1) members of the list is calculated. If the distance between the centers of geometries is important, we should calculate the centroids for all the polygons that represent the boundaries of the buildings. The centroids can be calculated using the centroid attributes of the Geopandas package. It can be helpful when the distance between the center of two or more geometries, like cities, is needed.

To calculate distances, we invoke the distance function of the Geopandas package. In the latter, the distance between two objects is usually defined as the smallest Euclidean distance or straight-line distance among the possible distance pairs of the two geometries.

By applying distance, we can match the index values of the GeoSeries (geometries) and compare elements with the same index with align=True or ignore the index and compare elements based on their matching order with align=False [13] (Fig. 6).
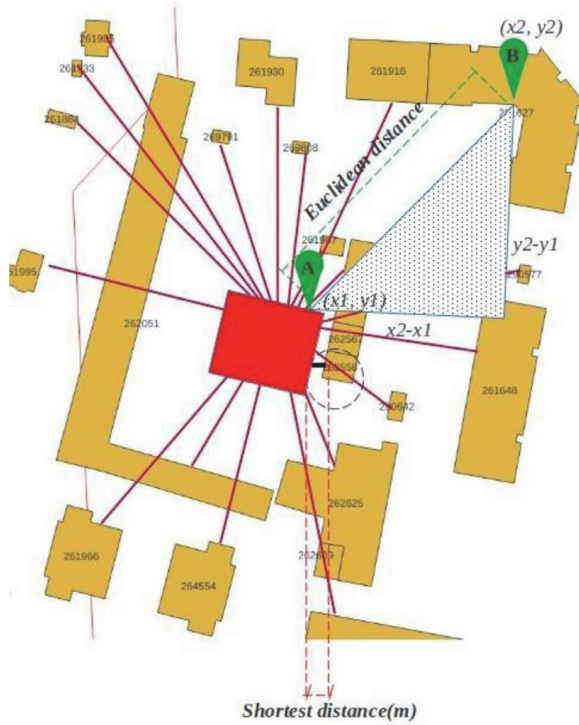
Fig. 5. The nearest neighbor to each building.

Fig. 6. The GeoSeries distance, index value True or False [13].

The code back ID of two adjacent buildings with the minimum distance in meters between them. The codes for calculating the nearest neighbor are given below:

```
for i, y in zip (DFB.geometry, DFB.objectid):
distance = DFB.distance(i, align = True) sorted_list=sorted
(distance.to_list()) min_sorted_list=sorted_list[1]
    print (f"the nearest neighbour of {int(y)} building ID is located
in min_sorted_list} meters")
```

In addition, if the maximum, mean, and average distance are needed, the following attributes of Geopandas in the codes can be applied.

```
maximum_distance = distance.max() mean_distance= distance.mean()
```

Data acquisition III: main angle and orientation of buildings

The orientation of the building is of paramount importance factor that affects the incident solar radiation and the absorbed heat wave. According to Mondal, a building with an east-west orientation (EW) has maximum solar gain, and a building with a north-south orientation (NS)
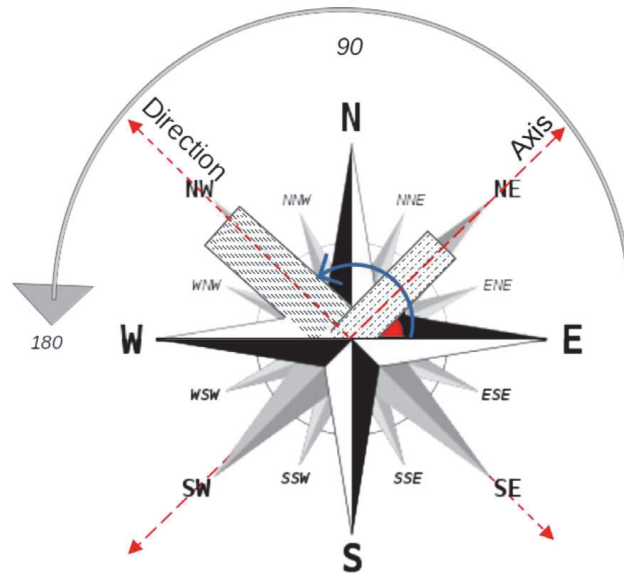
Fig. 7. Direction, angle, and orientation of buildings.

has minimum solar gain [14]. Since data collection aims to create a predictable and explainable ML UHI model, the sub-scoring of buildings oriented in NE-SW and NW-SE is crucial because it shows us how buildings are exposed to solar radiation and receives heat wave.

To determine the angle values of the building, we need to know the orientation of the longer side of the building and the direction of each building. The conventional axes denote the directions, as Fig. 7 shows. The east direction (positive x-axis) is assumed to be 0 degrees, as shown in Fig. 1. The angles are calculated in an anticlockwise direction, i.e., north is 90 ° and west is 180 ° [14]. The orientation of a building is the direction or angle to which the length is facing. Since the axis of a building is parallel to its length, or conversely, perpendicular to its orientation, the main angle of the building is located on the right side of the following figure is 45°. In contrast, the orientation of the building is 135° in the NE-SW direction. Considering the building on the left side of the figure, the angle is 135°, the orientation is 45°, and the direction is NW-SE.

In addition, Fig. 7 indicates the building direction and angle calculated using QGIS Tools.

The expression is available in the windows of the field calculator.

To determine the main angle of all buildings in the dataset, all geometries should be called with $geometry in the field calculator window of the expression part of QGIS Tools. expression main_angle ($geometry)

3.5. Completing the Dataset

The last step to prepare the dataset is appending other general and technical characteristics of buildings in different levels of the hierarchical grid system on a different scale (Tables 1, 2, and 3).

Ultimately, the dataset is created according to the characteristics of the buildings, their spatial indices on the hierarchical system grid, the UHI value of each building, the defined indexes based on buildings and different zones in the city area, and the weather data on dates that the city experienced UHI phenomena.

4. Study Limitations

This work is subject to the following limitations in exploring and acquiring urban data, especially when creating a comprehensive dataset that covers various essential features related to buildings, their surroundings, and other environmental factors in different scales.

Some of the limitations that we faced while collecting urban data are:

Lack of data sources: In data acquisition on the urban scale, there is a need for more data sources for specific features related to buildings and their surroundings.

Need for more data: Even if data sources are available, in some cases, the information may need to be completed or detailed enough to create a comprehensive dataset. For example, the available data on green areas or open spaces in the city need to be more detailed to capture all the necessary information about the vegetation cover or the quality of the environment. Data quality: Another limitation of urban data acquisition is its quality. Generally, data collecting needs to be more accurate and complete, which can affect the accuracy and reliability of the dataset. For example, there is just some general information, not detailed data, about some of the features of buildings, like the shape of the roofs, compositions of facade materials, and morphology of buildings.

In addition, there are other challenges that one might face when collecting and analyzing urban data. For example, collecting data on some features, such as the shape of the roof of a building or the features of neighbors, require on-site inspections, which are time-consuming and costly. Moreover, some data sources, such as GIS and geo data, must provide the level of detail required to capture the necessary information accurately.

Ethics Statement

The authors declare that the hereby presented data and data article fully comply with the Journal's policy regarding authors' duties, data integrity, and experimental requirements.

Credit Author Statement

Nasim Eslamirad: Conceptualization, Methodology, Programming, Data capturing, Writing – Original draft preparation; **Francesco De Luca:** Conceptualization, Supervision, Reviewing, and Editing; **Kimmo Sakari Lylykangas:** Supervision; **Sadok Ben Yahia:** Conceptualization, Supervision, Programming, Reviewing, and Editing; **Mahdi Rasoulinezhad:** Conceptualization, Programming.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Geoprocess of geospatial urban data in Tallinn, Estonia (Original data) (Mendeley Data).

Geoprocess of geospatial urban data in Tallinn, Estonia (Images) (Original data) (Mendeley Data).

Geoprocess of geospatial urban data in Tallinn, Estonia (Images) (Original data) (Github).

Acknowledgments

References

[1] www.latlong.net.LatLong.net. https://www.latlong.net/place/tallinn-estonia-5882.html.
[2] Páez. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal, 61(1), 2005. doi:10.1007/sgejo-004-0877-x.
[3] Anna Kovacs-Györi, Alina Ristea, Clemens Havas, Michael Mehaffy, Hartwig H. Hochmair, Bernd Resch, Levente Juhasz, Arthur Lehner, Laxmi Ramasubramanian, Thomas Blaschke, Opportunities and challenges of geospatial analysis for promoting urban livability in the era of big data and machine learning, ISPRS Int. J. Geo-Inf. 9 (12) (2020), doi:10.3390/ijgi9120752.
[4] Camila Mayumi Nakata-Osaki, Léa Cristina Lucas Souza, Daniel Souto Rodrigues, THIS – Tool for Heat Island Simulation: A GIS extension model to calculate urban heat island intensity based on urban geometry, Comput. Environ. Urban Syst., 67 (2018), doi:10.1016/j.compenvurbsys.2017. 09.007.
[5] Mingjun Peng, Division of urban hierarchical grid based on hierarchical spatial reasoning, International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, 7492, 2009, doi:10.1117/12.838394.
[6] General data of Tallinn. URL https://www.tallinn.ee/en/statistika/general-data-tallinn.
[7] Valentina Sagris, Mait Sepp, Landsat-8 tirs data for assessing urban heat island effect and its impact on human health, IEEE Geosci. Remote Sens. Lett. 14 (2017) ISSN 1545598X, doi:10.1109/LGRS.2017.2765703.
[8] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. 11 (5) (2007) 1633–1644, doi:10.5194/hess-11- 1633-2007.
[9] buildingdata. URL https://livekluster.ehr.ee/ui/ehr/v1/infoportal/buildingdata.
[10] Weather Underground. URL https://www.wunderground.com/about/our-company.
[11] keskkonnateadlik. URL https://keskkonnateadlik-kaur.hub.arcgis.com/.
[12] V. Janssen, Understanding coordinate reference systems, datums and transformations, Int. J. Geoinform. 5 (4) (2009).
[13] geopandas. https://geopandas.org/en/stable/docs/reference/api/geopandas.GeoSeries.intersection.html.
[14] Joy Mondal. Debunking Thumb-rule: Orient Longer Faces of Buildings towards North-South to Min- imise Solar Gain in India. 11 2018.

Paper II
Eslamirad, N., De Luca, F., Sakari Lylykangas, K. & Ben Yahia, S. (2023). Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment. *Frontiers of Architectural Research, 12*, 541-555. ISSN 2095-2635, https://doi.org/10.1016/j.foar.2022.12.001

Higher Education Press

SOUTHEAST UNIVERSITY

RESEARCH ARTICLE

Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment

Nasim Eslamirad [a],*, Francesco De Luca [b],
Kimmo Sakari Lylykangas [b], Sadok Ben Yahia [c]

[a] FinEst Centre for Smart Cities, Tallinn University of Technology, Tallinn 10319, Estonia
[b] Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn 10319, Estonia
[c] Department of Software Science, Tallinn University of Technology, Tallinn 10319, Estonia

Abstract Predicting comfort levels in cities is challenging due to the many metric assessment. To overcome these challenges, much research is being done in the computing community to develop methods capable of generating outdoor comfort data. Machine Learning (ML) provides many opportunities to discover patterns in large datasets such as urban data. This paper proposes a data-driven approach to build a predictive and data-generative model to assess outdoor thermal comfort. The model benefits from the results of a study, which analyses Computational Fluid Dynamics (CFD) urban simulation to determine the thermal and wind comfort in Tallinn, Estonia. The ML model was built based on classification, and it uses an opaque ML model. The results were evaluated by applying different metrics and show us that the approach allows the implementation of a data-generative ML model to generate reliable data on outdoor comfort that can be used by urban stakeholders, planners, and researchers.
© 2022 Higher Education Press Limited Company. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The world is experiencing rapid urbanization (Steemers and Yannas, 2000; Luo et al., 2017; Li et al., 2016). Thus, users would determine the outdoor place's vibrancy, local economy, sustainability (Teoh et al., 2020), quality of life, and vitality of the city. Moreover, proper outdoor conditions invite people to be active in urban spaces (Gehl, 2011), and

urban planners are challenged to find solutions that ensure comfort in cities (Steemers and Yannas, 2000; De Luca et al., 2018, 2021). Urban planners are constantly striving to improve the quality of life of urban dwellers (Luo et al., 2017) through various design concepts and solutions for liveable and pleasant public spaces (Kim et al., 2018). Studies on building configurations can guide decision-making to improve both existing urban areas and new developments to increase the level of outdoor thermal comfort. Recently, several researchers have investigated the factors affecting outdoor thermal comfort and the method for assessing thermal comfort (Wang et al., 2020; Lai et al., 2014; De Freitas and Grigorieva, 2015; Jamei and Rajagopalan, 2019; Chen and Ng, 2012; Lawson, 1978; Cohen et al., 2013).

Although urban planners and researchers are very interested in assessing the quality of urban open spaces (Nikolopoulou and Lykoudis, 2006), there is still a shortcoming in the existing outdoor thermal comfort studies, such as the low efficiency of simulations and incapability to apply the result of outdoor thermal comfort assessment in the initial steps of the design process. Therefore, there is unexplored potential in investigating optimal building morphologies for outdoor comfort to support the implementation of conscious design and planning practices (Marshall and Çalişkan, 2011). The ML approach is a powerful method for capturing non-linear relationships between independent and dependent variables and could also be used to map subjective responses to the physical environment (Wang et al., 2020).

Although there have been few studies on outdoor comfort based on Artificial Intelligence (AI) and ML methods, in this paper, we propose a data-driven approach to predict the outdoor comfort level in urban areas based on the results of Computational Fluid Dynamics (CFD) simulation on the sidewalk around buildings in Tallinn by applying various environmental and building feature, and weather data. The current study focuses on building features, the main ones are morphology, area, and height in Tallinn, Estonia (Lat. 59 26′N Lon. 24 45′E), which is characterized by a humid continental climate with warm summers, according to the Köppen-Geiger classification Dfb (Peel et al., 2007).

The present study is novel in two ways: firstly, it aims to create a model to assess outdoor thermal comfort, while few studies consider outdoor thermal comfort in cold climatic conditions. Secondly, the study's methodology can be generalized and applied in other locations with different urban and climatic features since the study is based on using urban data to build predictive and data-generative ML. The input data for the ML process of the current study comes from the modeling and simulation process conducted with three-dimensional building models using the ENVI-met simulation software. ENVI-met calculates the changes in a microclimate around urban structures and open spaces based on the principles of fluid mechanics, thermodynamics, radiation exchanges between urban surfaces, and temperature calculation, as well as the laws of atmospheric physics (Eslamirad et al., 2021; Park et al., 2014; Eslamirad et al., 2022).

The remainder of this paper is organized as follows: the related works are summarized in Section 2. Then, Section 3 thoroughly describes the proposed study's purpose and methodology. In Section 4, we thoroughly described the machine learning steps and the building of the ML model of the study. Section 5 looks at the results and achievements of the study and finally sketches the study's limitations and future issues. Section 6 focuses on the discussion in which we analyzed, explained, and interpreted the study results. The last section, Section 7, recalls the conclusion and take-home study outcomes.

2. Related works

Urban planning and design influence microclimate factors, which determine people's outdoor thermal comfort (Park et al., 2014). Over the past 30 years, several studies have focused on wind effects in the urban canopy layer by applying field measurements, the morphometric (geometric) method, wind tunnel experiments, and CFD simulations at the district, neighborhood, and building scales (Taleghani et al., 2015; Marshall and Çalişkan, 2011; Javanroodi et al., 2022). Therefore, researchers have developed techniques to assess the environment (Zhang et al., 2020). Taleghani et al. (2015) compared microclimatic conditions allowed by three building forms during June in the Netherlands and showed that courtyards provide the most comfortable microclimate compared to the other forms. Mittal et al. (2018) investigated methods for evaluating wind comfort at the pedestrian level by comparing different comfort criteria that significantly changed near high rise buildings. Indeed, Kim et al. (2018) considered both wind comfort and thermal comfort to assess pedestrian comfort levels and pointed out that the wind speed in the winter season could cause discomfort for pedestrians. In another study, wind speeds and direction were estimated in the ENVI-met environment based on the measured wind data and used to evaluate outdoor thermal comfort (Park et al., 2014). Pancholy et al. (2021) studied pedestrian comfort in the street canyon between parallel buildings and investigated canyon geometry on pedestrian comfort. In another study to help urban planners to measure heat stress in Melbourne, ENVI-met was used for modeling and simulation (Jamei and Rajagopalan, 2019). In 1996, John Gero described AI as a new design process (Tamke et al., 2018). Similarly, ML opens new perspectives in planning (Ramsgaard Thomson et al., 2020). ML tools deal with endowing programs with the ability to learn and adapt (Osisanwo et al., 2017). In the past years, data-driven approaches and ML have succeeded to predict the environmental performance of design solutions in the early design stages (Huang et al., 2022).

Thomsen et al. mentioned that information is advanced due to its predictive power, and urban developments can be modeled through self-organizing systems. They proposed Neural-Network Steered Robotic Fabrication method explores novel path planning approaches for robotic fabrication and it uses Machine Learning (Ramsgaard Thomson et al., 2020).

AI and ML-based methods are fully explored in urban planning. ML planners can evaluate large amounts of data, find patterns or anomalies, and ultimately use them to make better-informed decisions (Tamke et al., 2018). Okhoya (2015) states that ML can be used for a familiar

architectural task. In addition, ML can potentially be used by the engineering community to speed up complex CFD simulations. (Sebestyen and Tyc, 2020) trained an ML model comparing datasets of different complexities and energy rating prediction performance.

In addition, in a new study, a hybrid framework to perform a rapid wind assessment evaluation at the pedestrian level was applied by He et al. (2021). As the authors mentioned there is great potential in the integration of multiple computational tools such as parametric design and machine learning technique in sustainable and environmental research (Javanroodi et al., 2022) deployed a data-driven hybrid model to predict extreme microclimate conditions based on deep neural network (NN) machine learning architecture that was trained using the results of CFD simulation. Another study proposed an automated design process by applying a generative adversarial network (GNA) machine learning approach to achieve the real-time optimization of urban morphology. According to the study, the proposed method has an advantage in computational time comparing than traditional methods of outdoor environment optimization when the number of optimized samples is higher than 174 (Huang et al., 2022). Moreover, the ML approach could also be used to map subjective responses to the physical environment and to identify hidden trends (Wang et al., 2020).

This study analyzed the relationship between factors of the urban environment affecting outdoor thermal and wind comfort. The data of buildings features such as the morphological classes of the buildings, height, density of each building at the site, and orientation were collected from the cadastral map of Tallinn (Tallinn.ee) and used as variables. Then, defined samples of buildings that reflect actual buildings features in Tallinn were used for modeling, simulation, and ML approach through various interactive tools and metrics. This method helped to determine the effective design parameters to improve the outdoor thermal comfort of pedestrians and, consequently, outdoor quality of life even at the early design stage, considering patterns and structures of the urban data without the necessity of complicated and computer-intensive modeling and simulation processes.

3. Materials and methods

This study developed a predictive and data-generative model based on the ML approach to assessing outdoor comfort levels in Tallinn (cf., Fig. 1). The proposed methodology makes it possible to generate a high-accuracy learning algorithm.

3.1. The climatic condition of Tallinn in the studied date and time

The present work is based on the results of previously published research by the authors about the assessment of thermal and wind comfort around buildings in Tallinn (Eslamirad et al., 2021), conducted using the reference weather data of the springtime on 21 March, 10 a.m. in Tallinn (Table 1).

March is the month with the highest wind speeds and one of the coldest of the year (Fig. 2). The day and time were

chosen because they represent a worst-case scenario for thermal and wind comfort in spring during the morning when people are likely to spend some time in outdoor spaces. The rationale was that if people in this scenario are comfortable on the walkways of a particular building morphology, they are likely to be comfortable most of the time in spring.

3.2. Modeling and simulation process

Thanks to advanced computational techniques, the use of CFD is improving microclimate studies (Javanroodi et al., 2022). CFD simulation procedure consists of dominating equations, computational domain, boundary conditions, experimental validation, and mesh convergence analysis (He et al., 2021).

To define the geometric models, 11 morphological classifications were considered based on the most common morphology and shape of the buildings in the city plan of Tallinn: "Aggregated Blocks", "Articulated 1", "Articulated 2", "Block", "C-shape", "Court", "Free Form", "L-shape 1", "L-shape 2", "Linear" and "Series" (Table 2, Fig. 3). Then, the simplified geometries of the morphological classes were modeled in the geometric modeling process.

In the data shuffling process, the samples are defined based on their typical morphology and categorized following existing literature (De Luca et al., 2018). The approach followed in the study is to define building models based on six variables: morphological classes (11 types); several floors (1—19); the number of buildings in a site; density of the built area, floor area ratio per site (500—9500 m^2); height (3—57 m); and orientation (0, 20, 40, 60, 80°), as Table 3 shows.

The results of CFD simulation in the ENVI-met environment for thermal and wind comfort were used together to define the target used in the ML process (Eslamirad et al., 2021). Furthermore, Table 4 and Table 5, respectively, show more details about personal parameters and surface features applied in the CFD simulation and PET calculation.

As a result, the area considered for the thermal comfort assessment (green margin in Fig. 4) has a depth of 4 m for the entire building perimeter and a height of 1.1 m.

3.3. Thermal and wind comfort analyses

The data used in the study is related to the results of another published research by the authors (Eslamirad et al., 2021) that assessed outdoor thermal and wind comfort by applying the PET, Physiological Equivalent Temperature (°C) (Table 4) as the thermal comfort index and the LDDC variant of the Lawson comfort criteria (Lawson, 1978; Lawson and Penwarden, 1975; Lawson and Bristol, 1990) through the Wind Comfort Level (WCL) method (De Luca, 2019). Wind comfort was then determined using the LDDC Lawson criteria for pedestrian comfort (Table 7).

The ML method is based on the principle of learning algorithms from the discovered patterns, structures, and correlation of the labelled features and targets in the Python environment to predict outdoor comfort data in an urban context. The raw data set consisted of the attributes as independent variables in the numerical and categorical
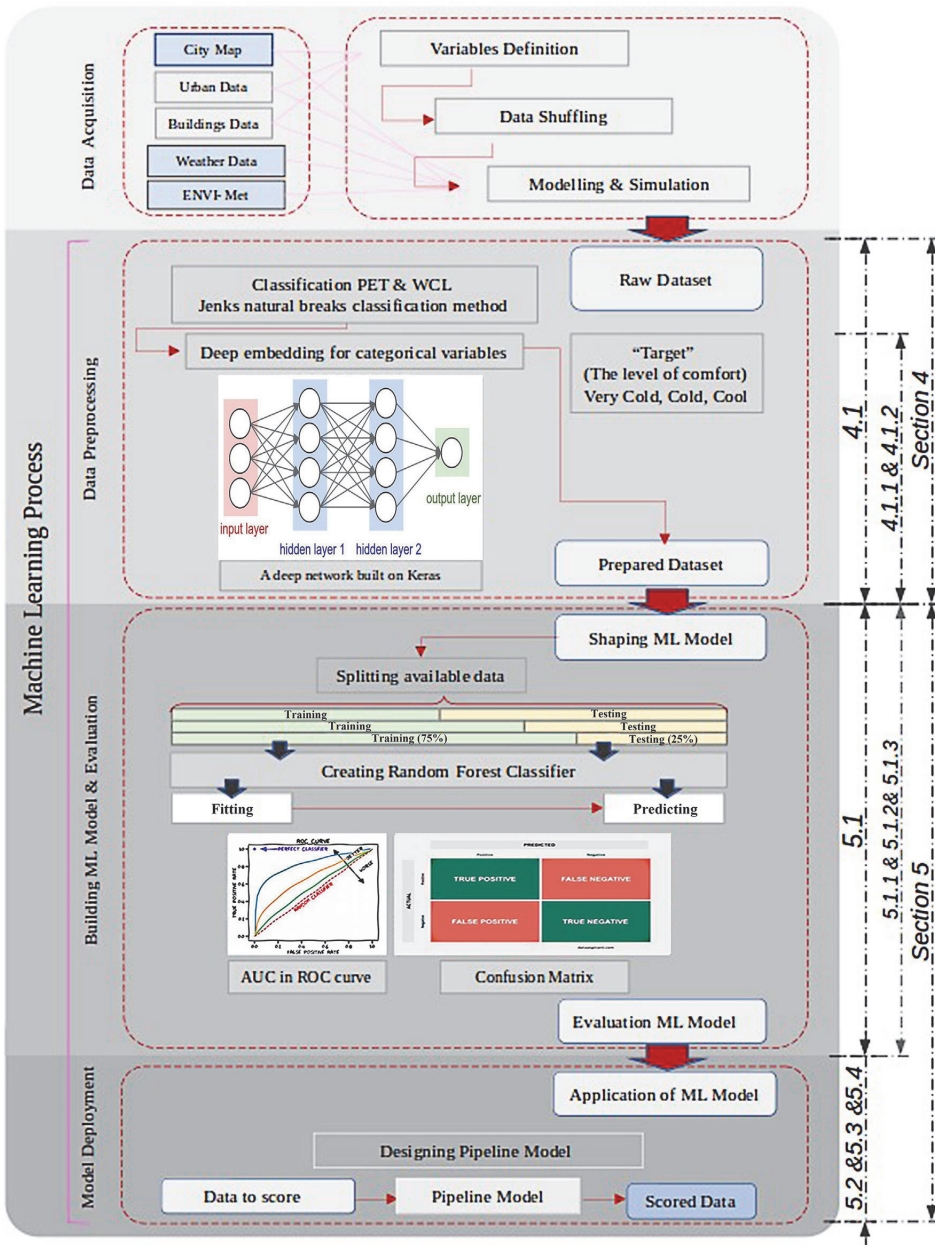
Fig. 1 The workflow developed for the study and the manuscript sections where the workflow steps are presented.

features, the PET thermal comfort metric, and the Wind Comfort Level (WCL).

The first step of the ML process is to import the data into the Python environment as Comma Separated Values (CSV) format file and create the data frame. To define the target in the data set, which indicates the outdoor comfort level of each sample based on thermal and wind comfort, the

Jenks optimizations method was applied. Based on the ranges of PET and WCL, the target comfort values were split into three groups: Very Cold (Target 1), Cold (Target 2), and Cool (Target 3).

The next step is pre-processing, which here the process benefits from the libraries in the Python environment to prepare the dataset for processing through the ML approach.

Table 1 Weather data.

Weather data, 21 March, 10 a.m.	
Wind speed (m/s)	8.40
Wind direction (degree)	202
Dry Bulb Temperature (ºC)	6.30
Dew Point Temperature (ºC)	2
Humidity (%)	74
Direct Normal Radiation (Wh/m^2)	63

Table 2 The definition and height of the morphological classification of the study.

Morphological class	The explanation of the morphological classes	Height (m)
Block	Compact square or rectangle	3—57
Aggregated blocks	The aggregation of a few or several blocks	3—57
Linear	A long rectangle	3—27
Series	The aggregation of small blocks in the main rectangle	3—27
L-shape	Two rectangles or some small aggregations	3—27
C-shape	Three rectangles with small aggregations	3—27
Articulated	Different rectangles and blocks without a central main part	3—12
Court	A block shape with an open area in the centre	3—18
Free-form	Non-rectangular, curved, aggregated shapes	3—12
Cluster	Separated or articulated buildings	3—12

First, the entire data is divided into a training set and a test set to build the ML model. The next is applying the Standard Scaler function, so the variables of the training dataset are in the same range. Then, the Random Forest Classifier trained on the training data, adjusted, and called predict to obtain predictions on the targets. Finally, the classification accuracy is evaluated to check the accuracy of the learning algorithm in the following criteria: Accuracy value, Confusion Matrix, and Area Under the Curve (AUC) in the Receiver Operator Characteristic (ROC) curve.

The last step is building the Pipeline model based on the ML learning algorithm, which comes from the Random Forest Classifier. Then, the Pipeline is trained with the new data from the other dataset. Finally, the prediction function is called only for the pipeline object based on the Random Forest Classifier to make predictions and generate new data for the target column. For example, this study evaluates the comfort level of pedestrians on the sidewalks surrounding the buildings in Tallinn. By applying the ML approach and using urban data, building characteristics, weather data, and thermal and wind comfort, the ML model will be able to generate new trustworthy results about the level of outdoor comfort in the pedestrian level in the city.

4. Machine learning process

The ML process of the study consists of two steps (Fig. 5). First, we trained a model to reach the learning algorithm. The next is using the learning algorithm in Pipeline to predict and score new input data.

The first step consists of four sub-steps to build, evaluate, and check the learning algorithm of ML before building the data generative Pipeline model.

1. Pre-processing and data exploration
2. Classification and building the ML model
3. Checking the accuracy of the classified data and the ML model

4.1. Data exploration and pre-processing

The first step in designing an ML dataset is finding the most relevant data through pre-processing. Thus, a raw tabular data set is created and characterized by the type of features, their classes, and their range. Because the ML model algorithm requires not only precision, accuracy, and minimum error to make an accurate prediction but also works better with features with higher correlation rates, the dataset needs to be evaluated in the pre-processing step to find and omit the data that are out of range. Therefore, while the feature engineering process is performing, the
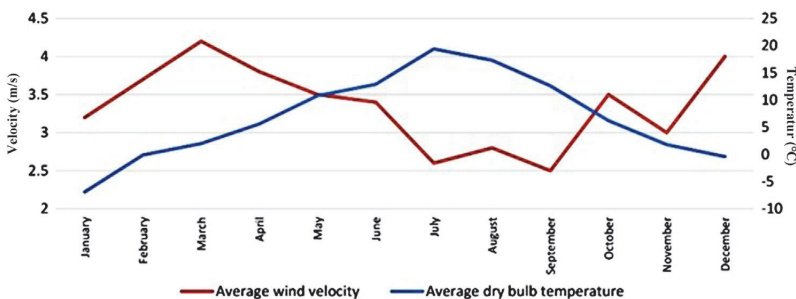

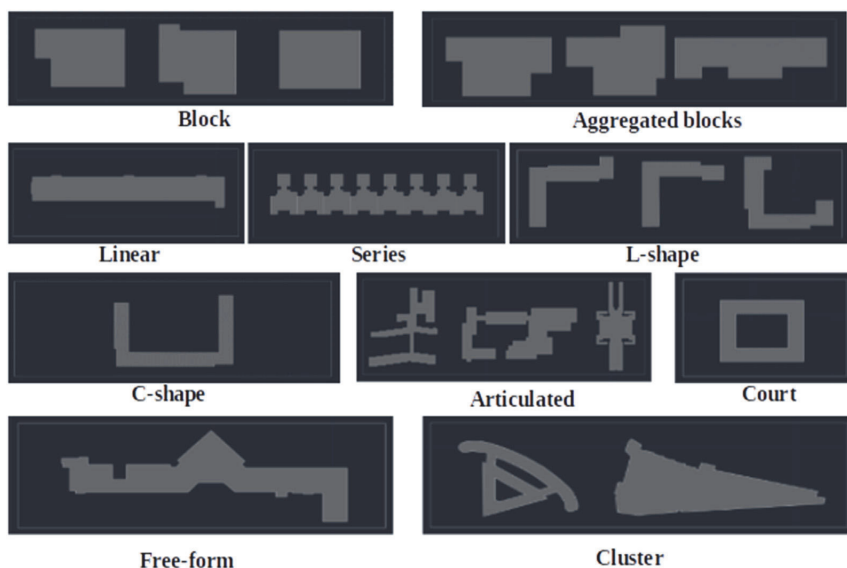**Fig. 2** Wind velocity and monthly temperature averages of Tallinn.

Fig. 3   The morphological classification and models of the study.

Table 3   The definition of samples.

Morphological class	Number of floors	Number of buildings	Density (m^2)	Tallness (m)	Orientation
Block	1—19	1—6	500—9500	3—57	0, 20, 40, 60, 80
Aggregated block	1—19	1—6	500—9000	3—57	0, 20, 40, 60, 80
Linear	1—9	1—6	1000—9000	3—27	0, 20, 40, 60, 80
Series	1—9	1—4	1000—9000	3—27	0, 20, 40, 60, 80
L shape 1	1—9	1—4	1000—9000	3—27	0, 20, 40, 60, 80
L shape 2	1—9	1—4	1000—9000	3—27	0, 20, 40, 60, 80
C shape	1—9	1—6	1000—9000	3—27	0, 20, 40, 60, 80
Articulated 1	1—4	1—6	2000—8000	3—12	0, 20, 40, 60, 80
Articulated 2	1—4	1—6	2000—8000	3—12	0, 20, 40, 60, 80
Court	1—6	1—6	1500—9000	3—18	0, 20, 40, 60, 80
Free form	1—4	1—6	2000—8000	3—12	0, 20, 40, 60, 80

Table 4   Human parameters in PET calculation.

Personal human parameters	
Age of person (year)	35
Gender	Male
Weight (kg)	75
Height (m)	1.75
Static Clothing Insulation (clo)	0.90
Work Metabolism (W)	80
Calculate from walking speed (m/s)	1.21

Table 5   The surface parameter in PET calculation.

Surface feature of buildings	
Roughness	0.010

raw input data was modified and cleaned of outlier data (that falls outside the range) and features with the lowest correlation.

The next step before building the ML model is to categorize the target columns, which indicate the comfort level of each sample based on thermal and wind comfort. Here one of the best range finder algorithms, the Jenks optimization method, is the data clustering method used to determine the best arrangement of values in different classes. Additionally, the Jenks optimization method was applied to classify PET and WCL into three classes. Table 6 shows the defined Targets based on the Jenks optimization method.

- PET class 1: 0.17 °C—2.77 °C, WCL class 1: 50.00%—72.97%
- PET class 2: 2.77 °C—4.5 °C, WCL class 2: 72.97%—87.77%
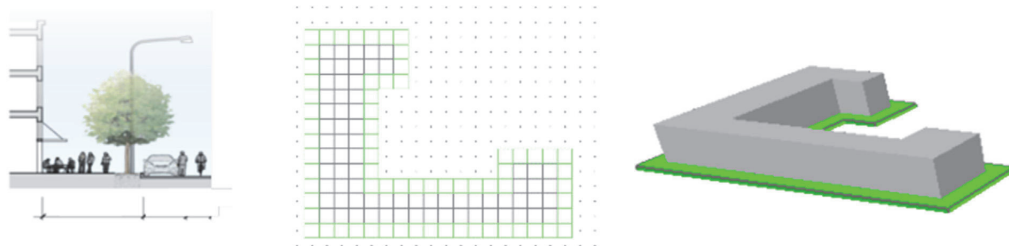
Fig. 4   The supposed sidewalk area in the simulation process.

● PET class 3: 4.5 °C−11.7 °C, WCL class 1: 87.77%−100%

In the feature selection, two variables, "Orientation" and "Density" were removed from the dataset as they had the lowest correlation with the variable "Target", which defines the outdoor comfort level on the sidewalks in Tallinn. Fig. 6 gives general information about samples in 4 different Area, Number of buildings(s), Height, and Target in Design Explorer tool. Fig. 6 samples show by filtering according to the area value; different colours represent specific areas 500 m^2, 1000 m^2, 1500 m^2, and 2000 m^2.

4.1.1. Transformation of the categorical variable

Before the data is transferred to the ML algorithm, the categorical variables of the morphological classes of the buildings are transformed into numerical features that are easier to identify in the ML process.

The transformation has a significant impact on the performance of the model. In this study, embedded layers for Deep Learning were created using Keras, which provides a complete framework. The input is the variable of each morphological class of buildings. This step results in a numerical representation of each morphological class in Tallinn in the 11 dimensions of the input set.

According to (Cerda et al., 2018) practical recommendations for encoding categorical features are an efficient choice for capturing resemblance. In this study, the output is a scaled column of the target (the degree of convenience). As categorical features differ from numerical, an operation to define a feature matrix is needed. Machine learning models that need vector data are applied after a categorical variable encoding to replace the tuple element with feature vectors. Each category is mapped to a particular vector, and the properties of the vector are adapted or learned when training a neural network. The vector space provides a projection of categories so that close correlated categories are naturally grouped. Here, the model has been trained for 50 iterations or epochs. This approach allows us to capture the relationships between categories and define three embedding variables that determine each class of building morphology in the dataset. The final result of the embedding process for each morphological class looks as follows, where each row in three elements represents one class of building morphology (Table 7).

The definition of "Target" classes is based on the classification of PET and WCL, applying the Jenks optimization method.

4.2. ML model building and classification

The learning algorithm finds patterns in the training data that map the attributes of the input data to the target and the results of an ML model that captures these patterns. First, the machine is trained with training data, and then it is checked to see how well it has learned by providing it with a test set. The test set has the same structure as the training set but does not contain any answers (Ramsgaard Thomson et al., 2020).

In this study, the Random Forest algorithm was used, one of the most popular algorithms for regression problems that
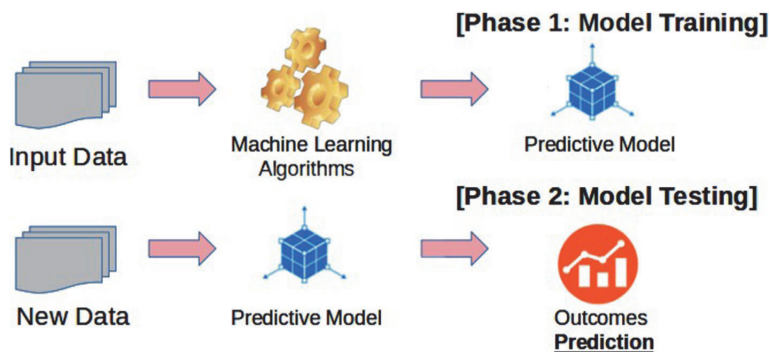

Fig. 5   Machine Learning training and scoring process.

Table 6 The range of thermal index predicted thermal perception by human beings and physiological stress on human beings (Cohen et al., 2013).

PET (°C)	Thermal Perception	Grade of physiological stress
	Very cold	Extreme cold stress
4		
	Cold	Strong cold stress
8		
	Cool	Moderate cold stress
13		
	Slightly cool	Slight cold stress
18		
	Comfortable	No thermal stress
23		
	Slightly warm	Slight heat stress
29		
	Warm	Moderate heat stress
35		
	Hot	Strong heat stress
41		
	Very hot	Extreme heat stress

Table 7 LDDC Lawson comfort criteria (Lawson and Penwarden, 1975).

LDDC Lawson comfort criteria		
Category	Comfort activity	Threshold Wind Velocity (m/s)
1	Sitting	4
2	Standing	6
3	Walking	8
4	Business walking/cycling	10
5	Uncomfortable for all activities	>10

work with both categorical and numerical input variables well. Leo Breiman proposed Random Forest in the 2000s to create a predictor ensemble with a set of decision trees that grow in randomly selected data subspaces (Wang et al., 2020).

Like the decision tree, the idea of the Random Forest Regressor comes from the ordinary tree structure, which consists of a root and nodes (the positions where branches

divide), branches, and leaves (Jehad et al., 2012) represented by the segments connecting the nodes.

This study uses the Random Forest Classifier from the Scikit Learn library. From the Sklearn model selection, the train test split function was applied to split the dataset into the train and test subsets. In this study, the classification performance was tested based on the different sizes of training and test datasets. Finally, the best classification performance was obtained with a training size of 75% and a test size of 25%. In addition, we used the stratification parameters in the splitting step to ensure that the training and test datasets contained examples of each target class in the same proportions as in the original dataset.

The next step was normalized and rescaled the training dataset in Python using the Standard Scaler function, so the variables were put in the same range.

Finally, the forest model was externalized by fitting the training subset data and calling predict to obtain predictions. In this way, we fitted and evaluated the model on separate chunks of the dataset by applying the Random Forest Classifier.

5. Results

5.1. Evaluation of classification accuracy

Three methods were applied to assess the accuracy of the classification (Table 8). This subsection describes the application of the methods to assess classification accuracy and the interpretation of the assessment results mentioned in the Results section.

5.1.1. Accuracy score
The classification accuracy score is the first method to determine the accuracy of the built learning model, which measures how many labels got right out of the total number of predictions. This metric evaluation benefits Scikit Learn by using the true labels from the test set and the predicted labels for the test set. In this study, the classification accuracy is 0.69.

5.1.2. Confusion matrix
A confusion matrix, also known as an error matrix expresses how many of a classifier's predictions were correct and, if they were incorrect, where the classifier was confused. It allows visualization of an algorithm's performance and easy identification of confusion between classes, e.g., when one class is frequently mislabelled (website of geeksforgeeks.
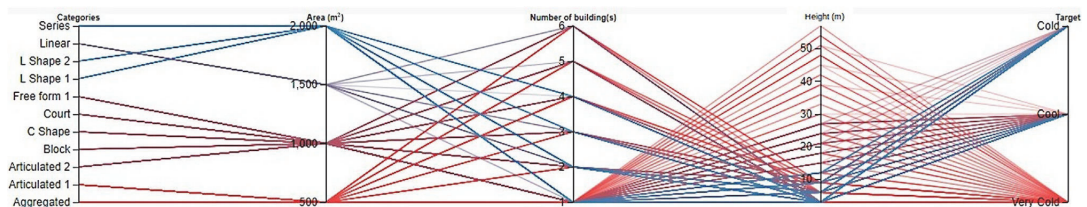

Fig. 6 The initial data before the categorical data transformation and scaling.

Table 8 The definition of "Target" classes is based on the classification of PET and WCL, applying the Jenks optimization method.

Level of outdoor comfortability		
PET	WCL	"Target"
0.17 ℃−2.77 ℃	50.00%−72.97%	Target 1
	72.97%−87.77%	Very Cold
	87.77%−100%	
2.77 ℃−4.5 ℃	50.00%−72.97%	Target 2
	72.97%−87.77%	Cold
	87.77%−100%	
4.5 ℃−11.7 ℃	50.00%−72.97%	Target 3
	72.97%−87.77%	Cool
	87.77%−100%	

org). The columns of a Confusion Matrix represent the results of the prediction classes, and the rows represent the results of the actual classes. It enumerates all possible cases of a classification problem (Zhang et al., 2020).

A confusion matrix of size $n \times n$ associated with a classifier shows the predicted and actual classification, where n is the number of different classes. Table 9 shows a confusion matrix for $n = 2$, whose entries have the following meanings.

a is the number of correct negative predictions;
b is the number of incorrect positive predictions;
c is the number of incorrect negative predictions;
d is the number of correct positive predictions.

The prediction accuracy and classification error can be obtained from this matrix (Visa et al., 2011).
Accuracy $= (a + d)/(a + b + c + d)$.
Error $= (b + c)/(a + b + c + d)$.
In the confusion matrices, the rows represent the true labels, and the columns the predicted labels. The values on the diagonal indicate the number of cases in which the predicted label matches the true label. The values in the other cells represent cases where the classifier mislabelled an observation. The column indicates what the classifier predicted, and the row indicates the correct label.

The Python library Sklearn was applied to calculate the confusion matrix and the results of the classification report. Based on the dataset, 26% of the data in the Target column are Target 1, while Target 2 and 3 account for 36% and 37.5%, respectively.

Fig. 7 shows the confusion matrix plot using labels and the colour scale, with the normalized values of all three targets. The Confusion Matrix shows Target 1 was labelled 86% correctly, i.e., 74 out of 86, while Target 2 (86 out of 120), meaning Cold label, was marked 72%, and Target 3 (68 out of 124) as Cool label was correctly labelled only 55%. Fig. 7 also shows that 14% (12 out of 86) of Target 1 instances were incorrectly labelled as Target 3, and 28% of Target 2 instances (34 out of 120) were considered Target 3 when classified. For Target 3, only 0.48% (less than 1 out of 124) were misclassified as Target 1, while 40% of instances in Target 3 were misclassified as Target 2 (50 out of 124).

5.1.3. AUC in ROC

We also used Area Under the Curve (AUC) in the Receiver Operator Characteristic (ROC) curve, also known as area under the curve, the evaluation metric used to calculate the performance of a binary classifier to find out how accurate the classification is. It is a curve that plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at different thresholds, separating the "signal" from the "noise". The AUC value measures a classifier's ability to discriminate between classes and is used as a summary of the ROC curve. The higher the AUC value, the better the model's performance in distinguishing between the positive and negative classes. Here, the AUC value in ROC indicates 0.78.

Meaning AUC in ROC shows the classification performances very well.

5.2. Designing pipeline skeleton

The next step is to create a Pipeline since we have done the necessary pre-processing and ML processing. The Pipeline creation process starts with classification and then assessment of the classification (Yahia and Johansson, 2011). Since classification and evaluation of the accuracy of the ML model were completed in the previous steps, next is fitting the algorithm to the new training data and predicting the target values in the three different groups: Very Cold,

Table 9 The embedding categorical morphological class in three sizes.

|	Name of the Morphological class	Embedding_1	Embedding_2	Embedding_3
1	Aggregated blocks	−0.07856157	0.04777364	0.12004933
2	Articulated 1	−0.09694797	0.12489636	0.140933
3	Articulated 2	0.09129719	−0.09825113	−0.11955576
4	Block	0.07574975	−0.12907775	−0.11605914
5	C Shape	0.05438302	−0.08725278	0.01638534
6	Court	0.03764237	−0.10152158	−0.06313951
7	Free form	0.05259462	−0.07555065	−0.10173931
8	L Shape 1	0.00210845	−0.02964082	−0.02180827
9	L Shape 2	0.1176304	−0.06369058	−0.04275405
10	Linear	0.06750278	−0.02232246	−0.01423943
11	Series	0.07436664	−0.0262874	−0.03133094
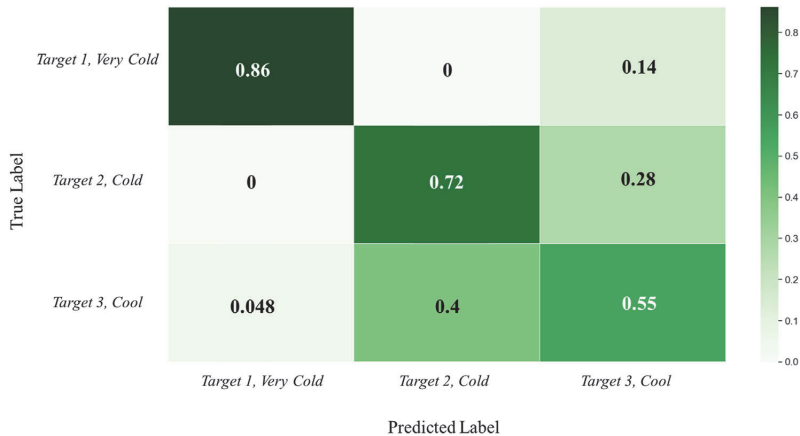
Fig. 7 The representation of the confusion matrix in this study.

Cold, and Cool. The new dataset has the same building features of Table 3, but the Target column is missing. By applying the built learning algorithm and fitting the new training data, Targets will predict through the ML Pipeline. Here the dataset is created with the samples representing the different models in the initial dataset. Therefore, the prediction function is only called for the Pipeline object to make predictions about the targets in the new test data. In this step, the new test data must first be entered into the Python environment as a CVS file using the Pandas library. Then, about 10,500 new target variables were predicted based on the learning algorithm and the model capability of ML Pipeline. The target variables correspond to the ML model in three classes representing the comfort quality of the sidewalks around each building in Tallinn in the ranges from Target 1 to Target 3, i.e., the sidewalks with different comfort levels around the building models.

The Pipeline ML model reads the data from the new dataset as training data and calls the learning algorithm based on the Random Forest Classifier. Finally, the highly accurate target values were saved in an Excel CSV file and added as the last column of the test data with about 10,500 building models showing the comfort value of the sidewalks of the building models in Tallinn based on the predictive and data-generative ML model.

5.3. The evaluation of the model

In classification problems, good classification accuracy is the primary concern (Visa et al., 2011). The accuracy of the ML model was assessed in three ways (Table 10). The first

assessment of the model's performance as an accuracy score yields 69%, which means that most of the labels in the test data set were detected very well in all three targets (Table 11).

However, since accuracy is the perfect measure of the classifier's performance when the classes are unbalanced (like the dataset in this study), the other metrics should be used to evaluate the model's performance.

The Confusion Matrix results show that the classifier and ML model perfectly understood Target 1 as a Very Cold level and Target 2 as defining the Cold level. But for Target 3, which denotes the Cool level of comfort (Cool), the model was confused and misclassified 50% of the labels as Target 2 as Cold. The results show that the model ML recognized the difference between Very Cold and other targets. Therefore, although the classification is not good enough to label Target 3, it is good enough to distinguish Target 1 (no comfort quality) from the others that correspond to comfort level.

The report of the classification contains the following values:

Precision is the number of correctly identified class members divided by the number of model predictions for that class.

Recall as the number of members of a class correctly identified by the classifier divided by the total number of members of that class.

F1 is a quick method to determine whether the classifier identifies the members of a class well.

Although the model has a high accuracy rate in all three metrics precision, Recall, and F1 score for Target 1, it mixes labels of Target 2 and Target 3. The classification report shows that the model ML and classification were

Table 10 The metrics for evaluation of classification.

The metric of evaluation	The value of the accuracy
Classification accuracy score	0.69
Confusion Matrix (Macro average)	0.71
AUC in ROC	0.78

Table 11 The confusion matrix for two-class classification problems (Visa et al., 2011).

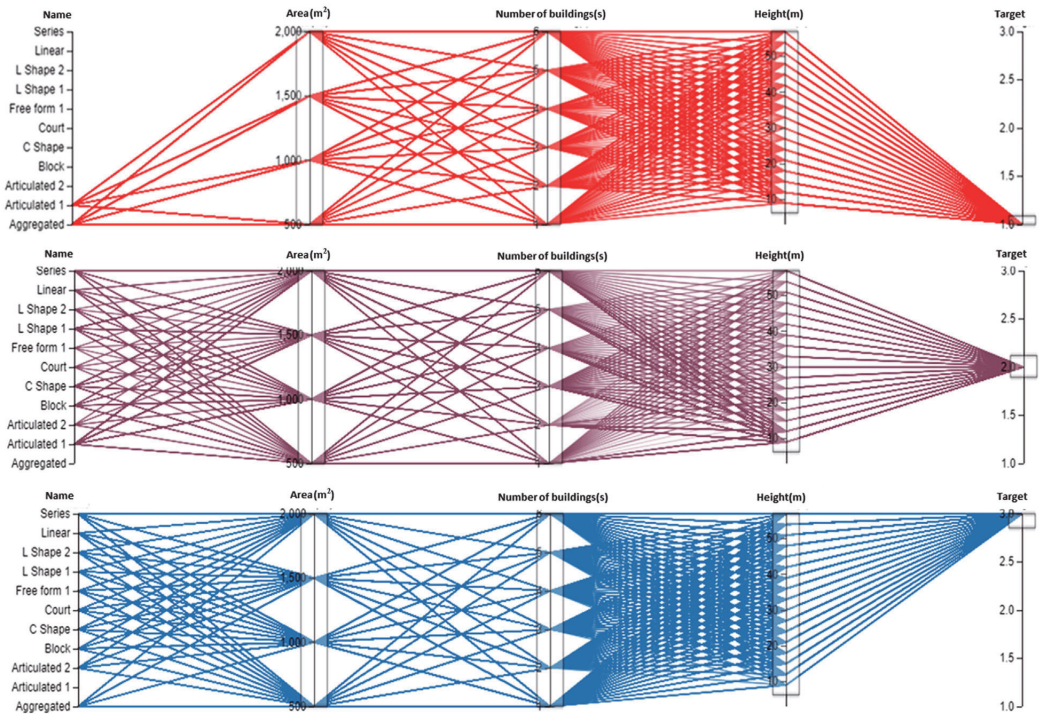|	Predicted Negative	Predicted Positive
Actual Negative	a	b
Actual Positive	c	d

Fig. 8   Using ML Pipeline, the generated data over Targets 1, 2, and 3.

able to classify the classes of the target very well, as at least 60% of the instances were classified correctly. In this study, the results of applying AUC in ROC using the Scikit-Learn libraries show 0.78. Therefore, the maximum value of the AUC in the ROC metric is 1. Then in the case of $0.5 < AUC < 1$, there is a high probability that the classifier can distinguish the positive class values from the negative class values, and the classifier can detect more numbers of True positives and True negatives than False negatives and False positives. Thus, the value of 0.78 in AUC in the ROC metric tells us that the ML model and classification work well, and the algorithm is reliable to apply in the next step.

5.4. The evaluation of the generated data

Fig. 8 shows the results of the generated target in the new dataset introduced into the ML model. The target column was filtered by "Target 1", "Target 2" and "Target 3". The generated data includes features of 500 initial geometric models.

In this study, according to the diagrams in Fig. 9, sidewalks surrounded by buildings of the morphological classes "Articulated" and "Aggregated" have the lowest level of thermal and wind comfort, i.e., the first level (Target 1, Very Cold). The remaining morphological classes provide the medium or highest level of comfort, i.e., the lowest
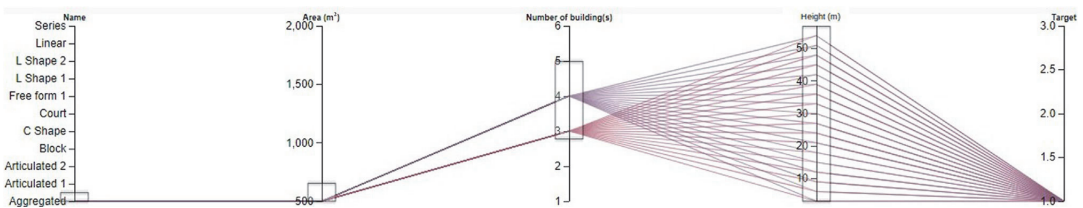
Fig. 9   The samples in the "Aggregated" morphological class cause the lowest.
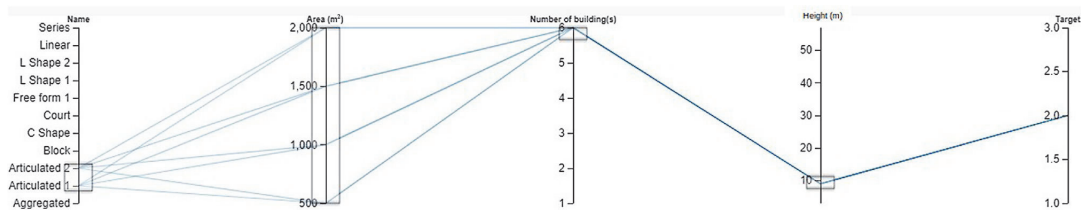
Fig. 10 The outdoor comfort level in the "Aggregated" morphological class.

Table 12 The report on the classification of the ML model.

Precision		Recall	F1-score
Target 1, Very Cold	0.93	0.86	0.89
Target 2, Cold	0.63	0.72	0.67
Target 3, Cool	0.61	0.55	0.57

level of discomfort, i.e., Cold or Cool, on the sidewalks around the buildings.

Furthermore, the results in Fig. 10 show that the buildings of the morphological class "Aggregated" with the smallest area (500 m²) and aggregated in groups of 3—5 in one place cause very cold discomfort for pedestrians on the sidewalk around the buildings. In addition, this morphological class leads to a higher level of outdoor comfort, Cold and Cool level when the variables are in the higher values.

The evaluation of two morphological classes with medium comfort levels shows that the "Articulated" class (1 and 2) causes medium, i.e., Cool, comfort level on the sidewalk of the studied cases when the number of buildings on the site is high (6). On the other hand, the height is the lowest (3 m, 6 m, or 9 m) (Fig. 10).

As Table 13 shows, most of the morphological classes in the studied cases in Tallinn provide the medium or highest comfort level (Table 12). Thus, the sidewalks around the buildings are suitable for creating a pleasant space like all kinds of activities in spring during the analyzed time and day. However, considering the effects of variables such as the area, the number of buildings on the site, and the height of the buildings on the site does not clearly indicate their independent relationships to the level of comfort around buildings. Therefore, assessing outdoor thermal

comfort depends on considering not just one of the variables but several.

6. Discussion

6.1. Analyzing data based on the ML model

Since the study aims to determine the level of outdoor thermal and wind level at the pedestrian level and understand the best morphological classes of the building that provide optimal thermal comfort, the predictive targets show us which morphological classes could offer the best comfort level to people.

Moreover, since the data were classified into three classes and the classification was assessed using the Confusion Matrix, the report shows us that the data with the labels of Target 2 and Target 3 were mixed in some cases. Therefore, ML could not identify them very well. This may be because the definition of these two classes stands for the Cold and Cool classes, while the definition of the first class, which stands for the Very Cold comfort level, is clearly different from the other comfort levels.

Therefore, evaluating the collected data on the comfort level in urban areas in the early stages of urban planning would be more helpful if the target classes were defined in binary classes (two groups), i.e., Comfort and Non-Comfort. In this way, the model of ML can better divide the samples into two groups, and the results are more accurate.

Optimizing outdoor thermal comfort at the urban scale helps to create a better environment where people engage in various activities, so architects and urban planners seek to consider the elements that contribute to maximizing comfort quality in urban areas. Collecting data on the condition of buildings in the urban area, assessing thermal

Table 13 The definition of "Target" is based on the classification of PET and WCL.

Models	Morphological class	Area (m²)	Height (m)	Density (m²)	Orientation (degree)	WS (m/s)	PET (C)	WCL	Comfort Level
M74	Block	500	45	7500	60	1.22	12.2	85.29	Cool
M86		500	54	9000	0	0.65	12.8	85.29	Cool
M226	Aggregated	500	36	6000	0	2.23	12	85.29	Cool
M236	blocks	500	42	7000	0	2.23	12	85.29	Cool
M240		500	42	7000	80	1.97	12.2	85.29	Cool
M242		500	45	7500	20	1.97	12.2	85.29	Cool
M246		500	48	8000	0	1.97	12.2	85.29	Cool
M256		500	54	9000	0	1.78	12.2	84.29	Cool

comfort and wind comfort around the buildings, and using this data to make predictions and generate comprehensive data on other buildings can be helpful steps in the pre-design and design processes in architecture and urban planning. The results of this study provide new information about the comfort level on the sidewalks around buildings in Tallinn, which can be helpful in the planning and preliminary design of buildings and urban areas. In this way, applying the generated data from the predictive ML Pipeline model can be considered a design tool to meet urban users' outdoor thermal and wind comfort needs.

6.2. Analyzing results

Since the study aims to evaluate outdoor comfort, it focuses on the sidewalk around the building to determine which morphological classification of buildings provides better thermal and wind comfort for people standing, sitting, and working around the buildings.

In this study, data from more about 500 models were derived and analyzed separately through a simulation process to obtain thermal and wind comfort in the proximity buildings to feed the ML model. Then, the data is evaluated with all variables that correlate with the concept of thermal and wind comfort conditions on the sidewalks around different buildings to create an ML Pipeline model and generate new data on the comfort quality. In the study, the correlations between climatic factors that determine outdoor thermal comfort and wind comfort was considered and applied in the ML process to predict outdoor comfort about morphological aspects and other characteristics of buildings. The results of the comfort level of new buildings show that the proposed framework has the potential to be applied in architectural and urban design studies to provide results based on both engineering and statistical modeling. This helped in finding effective design alternatives and reliable data, such as the morphological building class, the area, the number of buildings on the site and the height of the buildings, and their impact on the outdoor thermal and wind comfort and consequently, on the level of activities that people can perform in the city. The results show that, except for the "Articulated" and "Aggregated" morphological classes, other samples provide the comfort level of Cool or Cold and that the surrounding sidewalks of these patterns are better spaces for all levels of activities for people.

7. Conclusions

This article provides information on the correlation between climatic factors that are important for determining outdoor thermal and wind comfort around buildings in Tallinn, Estonia, and the ML approach to assess them. This study is subject to the following limitations.

The first limitation relates to the input data for the assessment of PET in the ENVI-Met software. Depending on the aim of the study, personal data, clothing and activity level of people, and climatic data can be changed. Moreover, though 500 building cases were generated to train and evaluate the ML model, the data was still relatively few.

The second limitation is considering weather data of a specific date and time during the CFD simulation process. Since outdoor thermal comfort depends on climatic conditions, the input data and, consequently, the thermal comfort and wind comfort will change at different seasons, dates, and times.

However, because of not considering the impact of urban elements, like green areas and vegetables, simulated models would be more or less different from realistic designs. In the same meaning, during the CFD simulation process, buildings' anthropogenic energy was not considered.

Nevertheless, in the previous study by the authors, the day and time were selected because it represents air temperature and wind conditions during winter and spring in Tallinn. The third limitation is the difficulty of considering the qualitative approach and people's thermal sensations. However, people's thermal senses can be considered when assessing the extent to which an area is perceived as comfortable or uncomfortable. To address this issue, the ML model will be based on microclimatic measurements, simulations, interviews, and surveys in the development of the work. The last limitation of the study is related to the fact that the prediction results of the machine learning method are less accurate than CFD simulation, and more on mathematics, and machine learning is based on empirical predictive models. The hybrid simulation is based on mathematics, and machine learning is based on empirical predictive models. The hybrid framework presented in this research suggests that machine learning can be applied in urban studies as a trustable assistance to CFD simulation.

For future work, the authors are considering a quantitative and a qualitative approach to assessing outdoor thermal comfort. This means that thermal comfort will be evaluated through simulation processes and people's subjective thermal sensations obtained through interviews and surveys. Thus, the ML algorithm will benefit from both aspects to generate new and more accurate data for outdoor thermal comfort assessments. In addition, future research aims to study larger pedestrian areas in urban environments during cold and hot seasons.

Declaration of interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CRediT authorship contribution statement

Nasim Eslamirad: Conceptualization, Methodology, Programming, Modeling and simulation, Machine learning model, Visualization, Writing- Original draft preparation. **Francesco De Luca:** Conceptualization, Methodology, Writing- Reviewing and Editing, Supervision. **Kimmo Sakari Lylykangas:** Conceptualization, Writing- Reviewing and Editing, Supervision. **Sadok Ben Yahia:** Conceptualization, Methodology, Writing- Reviewing and Editing, Machine learning model, Supervision.

Acknowledgments

References

Cerda, P., Varoquaux, G., Kégl, B., 2018. Similarity encoding for learning with dirty categorical variables. Mach. Learn. 107 (8), 1477–1494.

Chen, L., Ng, E., 2012. Outdoor thermal comfort and outdoor activities: a review of research in the past decade. Cities 29 (2), 118–125.

Cohen, P., Potchter, O., Matzarakis, A., 2013. Human thermal perception of Coastal Mediterranean outdoor urban environments. Appl. Geogr. 37, 1–10.

De Freitas, C.R., Grigorieva, E.A., 2015. A comprehensive catalogue and classification of human thermal climate indices. Int. J. Biometeorol. 59 (1), 109–120.

De Luca, F., 2019, April. Sun and wind: integrated environmental performance analysis for building and pedestrian comfort. In: Proceedings of the Symposium on Simulation For Architecture And Urban Design, pp. 1–8.

De Luca, F., Dogan, T., Kurnitski, J., 2018, June. Methodology for determining fenestration ranges for daylight and energy efficiency in Estonia. In: Proceedings of the Symposium on Simulation For Architecture And Urban Design, pp. 1–8.

De Luca, F., Naboni, E., Lobaccaro, G., 2021. Tall buildings cluster form rationalization in a Nordic climate by factoring in indoor-outdoor comfort and energy. Energy Build. 238, 110831.

Eslamirad, N., De Luca, F., Lylykangas, K.S., 2021, November. The role of building morphology on pedestrian level comfort in Northern climate. J. Phys. Conf. 2042 (1), 12053. IOP Publishing.

Eslamirad, N., Sepúlveda, A., De Luca, F., Sakari Lylykangas, K., 2022. Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus. Energies 15 (4), 1577.

Gehl, J., 2011. Three Types of Outdoor Activities,"Life between Buildings," and "Outdoor Activities and the Quality of Outdoor Space": from Life between Buildings: Using Public Space. Routledge, The city reader, pp. 586–608, 1987.

He, Y., Liu, X.H., Zhang, H.L., Zheng, W., Zhao, F.Y., Schnabel, M.A., Mei, Y., 2021. Hybrid framework for rapid evaluation of wind environment around buildings through parametric design, CFD simulation, image processing and machine learning. Sustain. Cities Soc. 73, 103092.

Huang, C., Zhang, G., Yao, J., Wang, X., Calautit, J.K., Zhao, C., et al., 2022. Accelerated environmental performance-driven urban design with generative adversarial network. Build. Environ. 224, 109575.

Jamei, E., Rajagopalan, P., 2019. Effect of street design on pedestrian thermal comfort. Architect. Sci. Rev. 62 (2), 92–111.

Javanroodi, K., Nik, V.M., Giometto, M.G., Scartezzini, J.L., 2022. Combining computational fluid dynamics and neural networks to characterize microclimate extremes: learning the complex interactions between meso-climate and urban morphology. Sci. Total Environ. 829, 154223.

Jehad, A., Khan, K., Ahmad, N., Maqsood, I., 2012. Random forests and decision trees. International Journal of Computer Science Issues (IJCSI) 9.

Kim, H., Lee, K., Kim, T., 2018. Investigation of pedestrian comfort with wind chill during winter. Sustainability 10 (1), 274.

Lai, D., Guo, D., Hou, Y., Lin, C., Chen, Q., 2014. Studies of outdoor thermal comfort in northern China. Build. Environ. 77, 110–118.

Lawson, T.V., 1978. The wind content of the built environment. J. Wind. Eng. Ind. Aerod. 3 (2–3), 93–105.

Lawson, T.V., Bristol, 1990. In: The determination of the wind environment of a building complex before construction. Report Number TVL 9025. Department of Aerospace Engineering, University of Bristol.

Lawson, T.V., Penwarden, A.D., 1975. The effects of wind on people in the vicinity of buildings. In: Proceedings 4th international conference on wind effects on buildings and structures. Cambridge University Press, pp. 605–622.

Li, K., Zhang, Y., Zhao, L., 2016. Outdoor thermal comfort and activities in the urban residential community in a humid subtropical area of China. Energy Build. 133, 498–511.

Luo, Y., He, J., Ni, Y., 2017. Analysis of urban ventilation potential using rule-based modeling. Comput. Environ. Urban Syst. 66, 13–22.

Marshall, S., Çalişkan, O., 2011. A joint framework for urban morphology and design. Built. Environ. 37 (4), 409–426.

Mittal, H., Sharma, A., Gairola, A., 2018. A review on the study of urban wind at the pedestrian level around buildings. J. Build. Eng. 18, 154–163.

Nikolopoulou, M., Lykoudis, S., 2006. Thermal comfort in outdoor urban spaces: analysis across different European countries. Build. Environ. 41 (11), 1455–1470.

Okhoya, V., 2015. Towards the Application of Machine Learning on Architectural Projects. https://doi.org/10.13140/RG.2.1.1164.3048.

Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J.O., Olakanmi, O., Akinjobi, J., 2017. Supervised machine learning algorithms: classification and comparison. Int. J. Comput. Trends Technol. 48 (3), 128–138.

Pancholy, P.P., Clemens, K., Geoghegan, P., Jermy, M., Moyers-Gonzalez, M., Wilson, P.L., 2021. Numerical study of flow structure and pedestrian-level wind comfort inside urban street canyons. J. Roy. Soc. N. Z. 51 (2), 307–332.

Park, S., Tuller, S.E., Jo, M., 2014. Application of Universal Thermal Climate Index (UTCI) for microclimatic analysis in urban thermal environments. Landsc. Urban Plann. 125, 146–155.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11 (5), 1633–1644.

Ramsgaard Thomson, M., Nicholas, P., Tamke, M., Gatz, S., Sinke, Y., Rossi, G., 2020. Towards machine learning for architectural fabrication in the age of industry 4.0. Int. J. Architect. Comput. 18 (4), 335–352.

Sebestyen, A., Tyc, J., 2020. Machine learning methods in energy simulations for architects and designers. In: Proceedings of the of the 38th eCAADe Conference. Berlin, pp. 613–622.

Steemers, K., Yannas, S., 2000. Architecture, City, Environment. In: Proceedings of PLEA 2000. Earthscan, Cambridge, United Kingdom.

Taleghani, M., Kleerekoper, L., Tenpierik, M., Van Den Dobbelsteen, A., 2015. Outdoor thermal comfort within five different urban forms in The Netherlands. Build. Environ. 83, 65–78.

Tamke, M., Nicholas, P., Zwierzycki, M., 2018. Machine learning for architectural design: practices and infrastructure. Int. J. Architect. Comput. 16 (2), 123–143.

Teoh, M.Y., Shinozaki, M., Said, I., 2020. Redefining the notion of outdoor thermal comfort. SEATUC journal of science and engineering 1 (1), 52–61.

Visa, S., Ramsay, B., Ralescu, A.L., Van Der Knaap, E., 2011. Confusion matrix-based feature selection. MAICS 710 (1), 120–127.

Wang, Z., Wang, J., He, Y., Liu, Y., Lin, B., Hong, T., 2020. Dimension analysis of subjective thermal comfort metrics based

on ASHRAE Global Thermal Comfort Database using machine learning. J. Build. Eng. 29, 101120.

Yahia, M.W., Johansson, E., 2011. The influence of environment on people's thermal comfort in outdoor urban spaces in a hot dry Climate. The example of Damascus, Syria. In: 27th International Conference on Passive and Low Energy Architecture (PLEA). Presses universitaires de Louvain, pp. 589—594.

Zhang, X., Weerasuriya, A.U., Zhang, X., Tse, K.T., Lu, B., Li, C.Y., Liu, C.H., 2020, December. Pedestrian wind comfort near a super-tall building with various configurations in an urban-like setting. Build. Simulat. 13 (6), 1385—1408.

Paper III

Eslamirad, N., De Luca, F., Ben Yahia, S. & Sakari Lylykangas, K. (2023). From near real-time urban data to an Explainable city-scale model to help reduce the Urban Heat Island (UHI) effect. *Journal of Physics: Conference Series*. CISBAT 2023. In print.

From near real-time urban data to an Explainable city-scale model to help reduce the Urban Heat Island (UHI) effect

Nasim Eslamirad[1], Francesco De Luca[2], Sadok Ben Yahia[3], Kimmo Lylykangas[2]

[1] FinEst Centre for Smart Cities, Tallinn University of Technology, Tallinn, Estonia
[2] Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn, Estonia
[3] Department of Software Science, Tallinn University of Technology, Tallinn, Estonia
nasim.eslamirad@taltech.ee

Abstract. Urbanization is associated with increasing the temperature of urban areas and phenomena like UHI. The UHI produces a sensible impact on people's health, quality of life, urban liveability, and even the mortality rate. As a result, reducing high temperature in cities will improve the quality of life in cities. Overall, the understanding of the relationship between UHI and urban features is not well captured due to a lack of accessibility to data and cost implications related to long-term monitoring and simulation processes. Machine Learning (ML) offers significant opportunities to develop approaches that lead to high accuracy in discovering the significant features of buildings and urban areas, improving the conditions of the inhabitants. The paper aims to describe how the actual spatial data of Tallinn, Estonia, were assessed and applied to build an Explainable ML-based model. The study outcomes include the ML-based models that help understand the importance of urban features and their value in developing strategies and solutions to mitigate the UHI effect. The study's findings show that changes in the most important features of urban areas and buildings, such as the built area in the neighbourhood zone and the area and orientation of buildings, play the most significant role in transforming an urban area from one with no UHI effect to one that experiences heat waves. The results provide essential information for urban planners who implement built environments.

Keywords: Spatial data, Urban analysis, Urban Heat Island, Machine Learning Explainable model.

1. Introduction

Urbanization has been a critical driver of development throughout history [1], which caused a significant impact on the world [2]. One of the critical challenges facing urban design in the digital age is modelling the complexities of urban environments, particularly concerning environmental issues like the UHI effect, which refers to the air temperature difference between urban and rural areas [3]. Scientists, planning authorities, and governments have increasingly recognized the impact of urban design and planning on the severity of the UHI [4]. Therefore, urban planners have employed practical measures to mitigate the UHI effect and enhance cities' sustainability and livability. These efforts aim to create more favorable conditions for human habitation while ensuring ecological balance [5, 6].

Since urban environments are becoming increasingly complex, it is essential to have efficient analytical tools to aid in planning projects. In this regard, spatial data analysis is crucial in facilitating urban planning to address the challenges [7]. Furthermore, digitalization has opened new possibilities for integrating spatial science with ML to extract valuable insights from large and complex datasets by discovering hidden patterns that would be difficult to detect using traditional methods [8].

This research implemented a novel framework to capture spatial data from different scales, buildings, neighborhood zone, and urban scales to build an ML-based transparent model that explains the importance of features impacting the UHI effect. The current study highlights the importance of ML in increasing trust and confidence in urban studies. To address these shortcomings, the key aim of this study is to develop a holistic design framework considering interrelated research questions concerning: How can spatial data be captured from different resources? Moreover, how can ML-based models show the importance of urban feature that impacts the UHI effect?

2. Background studies

Rapid urban growth and the increase in impervious surfaces in dense urban areas can have significant impacts on the surface and air temperature of metropolitan cities [9] [10] to 2–5°C than those in surrounding rural [11] [5]. Brian Stone proclaimed in 2012 that the world's big cities are escalating in temperature faster than the planet [12] and explored global warming underway in cities [13]. Urban design factors are related to the physical characteristics of the built environment [14]. As the urgency of climate change grows, researchers and practitioners increasingly recognize the need for systematic, methodically well-grounded research to upscale and implement place-specific climate solutions while respecting local variation and context [15]. Based on learning theory, ML methods can extract meaningful information and patterns from this data, allowing for the identification of practical solutions tailored to the specific location and context [15]. The importance of informing how the machine makes models and predicts features develops the idea of using the explanation of ML models. Explanations are essential to trust that the predictions made by models are correct. The needs for Explainable ML models are because black-box ML models make it hard to understand and explain the behavior of a mode [16].

3. Methodology

This research is situated in urban studies and focuses on the implementation paradigm of urban assessment practice. The study aims to collect geospatial data relating to buildings and their context and data of the UHI phenomenon within an urban area. Furthermore, this research seeks to employ ML-based models using geo-processed urban data to clarify how different urban features affect the UHI effect. The framework of the study comprises two sections, as Figure 1 shows.

The first section is related to the urban assessment practice of collecting geospatial data of buildings and their context on the one hand and data related to the UHI phenomenon, on the other hand, to perform geoprocessing on the data. The data relating to the UHI effect in Tallinn are related to the summers of 2014 and 2019. Lastly, the meteorological data recorded when the city was affected by heat waves was assigned to the processed data. The steps in the second section are:
1- Building the Random Forest Classifier model of the UHI dataset as the base ML model.
2- Finding the essential urban features (Permutation feature importance).
3- Finding the marginal contribution and threshold of the most important attributes in the ML model.
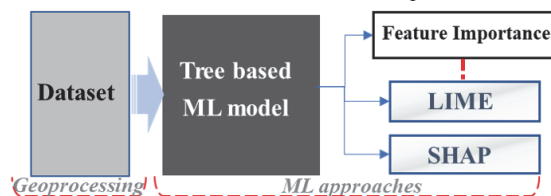
Figure 1. The schematic diagram of the framework of the study.

The geodata is integrated into QGIS Tool, and the data analysis is implemented in Python 3 in the Jupiter notebook interface. The Python libraries used are NumPy and Pandas: for data manipulation, Matplotlib and Seaborn: for visualization, and Geopandas for geodata manipulation.

3.1. Geoprocessing

The initial data used in the study are divided into three groups: Urban, Weather, and climate, and data related to the UHI effect. The urban data are based on the Tallinn Land Authority Geospatial Data

Portal [17, 18, 19] and indicate urban data in buildings and neighborhood details of Tallinn, Estonia (67,113 buildings) [20]. Moreover, to prepare the dataset, Estonian cities' assessed heat wave data that explored the extent and magnitude of the UHI effects were used [21]. The geographical information of the city is 445,005 population and 159 km^2 area [22] with a humid continental climate with cold winters, according to the Köppen-Geiger classification [23].

3.1.1. Ascending hierarchical grid system

The type of urban data was recognized as homogeneous or heterogeneous and static or dynamic types. The hierarchical grid system in the environment of the QGIS Tool was performed to capture the location-based urban data, considering their context. Each level of the grid system has an ascending hierarchical structure, starting with the smallest scale, the first grid, the square with a size of 200 meters, the second level with a dimension of 1000 meters, while the most significant level is a square with a dimension of 2000 meters [20]. The initial available data combine two components of the schema, and static heterogeneity means that elements in an urban area, like buildings, streets, and the natural landscape, remain unchanged or change little over time, and dynamic heterogeneity that refers to frequently changing data, such as climate, microclimate, and UHI data. The hierarchical grid system aims to create a spatial index for each object on the map [20].

3.1.2. Location-based features

The objects are referenced to the underlying grids and use the homogeneous ground to define urban indices mainly related to the heterogeneous data, which means accommodating static and dynamic data in the solid hierarchical grid system to detect objects on the map and define location-based indexes. According to the geometric classes of objects in the QGIS Tool, we considered polygons when the objects that should be detected are buildings [20]. Moreover, we aim to find the shortest Euclidean distances between buildings at the neighborhood scale with the spatial index to determine the size faster than looping through the data frame and then finding the minimum of all distances when working with a large data frame, that named 'NearestNeighbour' in the dataset [20]. In addition, the orientation of the building is a vital factor that affects the incident solar radiation and the absorbed heat [24] determined for all buildings in the dataset, named 'Angle' [20]. The last step to prepare the dataset is appending other general and technical characteristics of buildings in different levels of the hierarchical grid system on a different scale. Ultimately, the dataset is created according to the characteristics of the buildings, their spatial indices on the hierarchical system grid, the UHI value of each building, and the defined indexes based on buildings and different zones in the city area.

Furthermore, the meteorological data of the days during which the UHI effect was recorded in Tallinn (during the summer of 2014-2019) was assigned to the dataset, accordingly [25]. Finally, the UHI effect dataset in Tallinn, Estonia, involves 34,001 building samples from all eight districts of Tallinn, including location, building characteristics, urban characteristics, UHI data, and climate data. The UHI effect has three classes, 29 °C, 30°C, and 35°C [20] [26].

3.2. ML process

This section aims to find the importance of features that impact the UHI effect and the thresholds of the essential features in the ML model. Therefore, we first built a tree-based Random Forest Classifier model, then implemented approaches to understand the role of features in the dataset.

Feature importance of RandomForestClassifier model was performed with the permutation importance method. In the initial step to do feature engineering, we merged two classes of the UHI effect, which are 30 and 35, to make a class that shows the urban areas with the UHI effect (30°C and 35°C) and areas without the UHI effect (29°C) to make a binary classification with two values. The next was splitting the data into independent and dependent variables to indicate features and targets. The level of the UHI effect is selected as the target, and the rest are features. Next, we divided the dataset into training and testing datasets. After defining a model based on a random forest, we checked the model's performance and accuracy to ensure the model's performance was high enough. Here, the precision of the model is more critical since we are looking for the maximum performance of the model. The model's accuracy is 90%, while the precision is 82% and 92% in classification for classes 1 and 2, respectively.

The impurity-based feature importance breaks the relationship between the feature and the target. Thus, the drop in the model score indicates how much the model depends on the feature. For example, according to the bar chart in Figure 2, the numerical feature related to the 'Built Area in G200', the built area in the smallest grid of the defined ascending hierarchical grid system, is the most significant; after that, 'Angle,' 'Height,' and 'Shape Area' of the building are the most important features. The last essential features are 'Purpose of Building' and 'Material,' which are categorical.
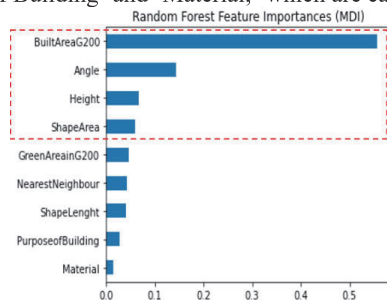
Figure 2. The feature importance in the classifier ML model.

3.2.1. The Explainable model

In the next step, we built the Explainable model based on RandomForestClassifier to analyze the performance and transparency of the model. Transparency in ML refers to the ability to understand and interpret the inner workings of a model [16]. They typically operate by relating the input of a model to its outputs without making assumptions about the internal workings of the model. The implemented Explainable ML-based models in the study are Local Interpretable Model-agnostic Explanations (LIME) & SHapley Additive exPlanations (SHAP) models. The explainable models that provide additional information about the model's predictions, used in this study, are LIME (explanation by simplification) and SHAP (feature relevance techniques). Both models aim to provide insight into how specific predictions were made [27]. Each attribute holds a certain weightage in predicting the occurrence of a class depending on a particular threshold. To explain the decisions behind the ML model, we interpreted the marginal weights of four top features ranked as the most significant attributed to the impurity-based feature importance (Figure 2).
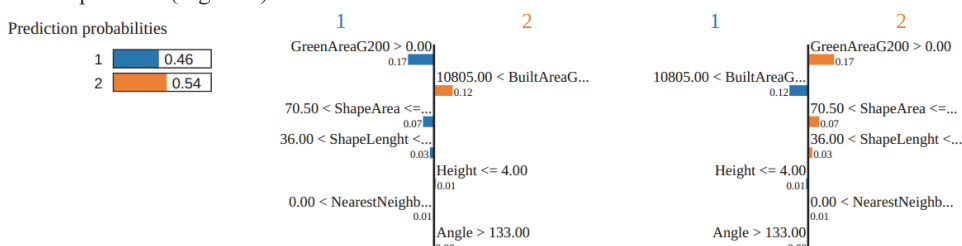
Figure 3. The threshold and value of top features in the LIME Explainable model.

According to Figure 3, LIME predicts class 1 with 46% and class 2 with 54% confidence. The first important attribute, 'BuiltAreainG200', indicates the threshold of 10805, meaning above the value increases the chance to be labelled in class 2 with a weight of 0.12, whereas below, it increases the chance of being in class 1 with a weightage 0.12. The following important attribute, 'Angle,' indicates the threshold of 133°, which means when the angle of the building is higher than 133° the UHI class is more likely to be class 2, but lower will lead the sample to label in class 1 (without UHI). The next is 'Height,' while 4m is the threshold to label a building in class 2, otherwise in class 1. In addition, the feature of 'ShapeArea', refers to the area of buildings in the dataset, with the threshold of 70.5 m^2 shows

that when the area of the building is higher than 70.5m^2, the sample has more chance to label in class 2 (UHI level=30°C); otherwise, it will be in class 1.

SHAP is the most popular explainable model, based on coalitional Game Theory (Shapley values), and involves identifying the essential features that contribute to the model's predictions to explain and interpret the coefficients as the feature's importance.

In the plotted results of the SHAP model, the red marks push the prediction higher toward the base value (about 30°C), while the blue marks do just the opposite. As Figure 4 shows, increases in the value of 'BuiltAreainG200' move the sample to be classified in UHI Class 2. The opposite changes are about the 'ShapeArea,' in the lower SHAP values.
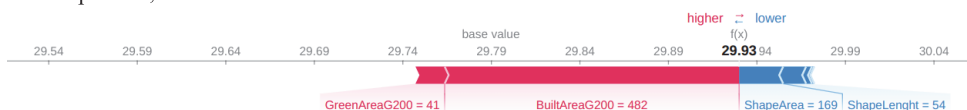
Figure 4. The performance of the SHAP Explainable model.

Moreover, according to the plot of dependency in Figure 5. a, between all most significant values in the model, the shape value increases with an increase in value of 'BuiltAreainG200' as the highest feature value in the SHAP model, in the wide range, and then 'ShapeArea' and 'Angle' as the second and third features. Also, Figure 5. b shows dots as a single prediction (row) from the dataset. In addition, the x-axis is the value of the 'BuiltAreainG200,' and the y-axis is the SHAP value for that feature, which represents how much the value of the feature changes the model's output for that sample's prediction. Moreover, the red dots represent the feature of 'ShapeArea' as a second feature that may have an interaction effect with the feature we are plotting. Therefore, as the plot indicates, the high value of 'BuiltAreainG200' and 'ShapeArea' maximizes the SHAP value.
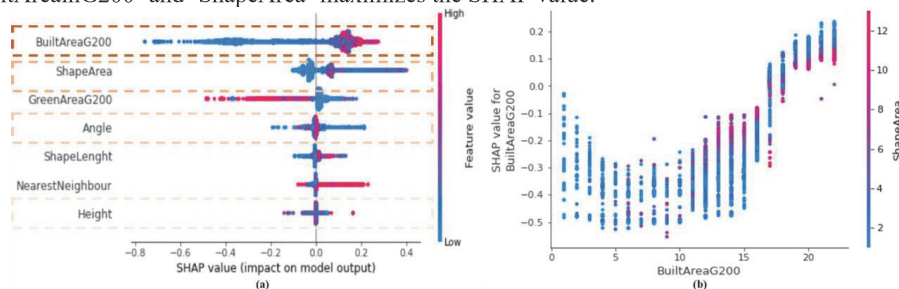
Figure 5. The plot of the dependence (a), the scatter plot of 'BuiltAreainG200' in SHAP value (b).

4. Conclusion

This study presents the spatial data and UHI data acquisition to build a dataset that is used to build an ML-based transparent model, which explains the importance of features impacting the UHI effect in urban areas. According to the impurity-based feature importance, 'BuiltAreaG200', 'Angle,' 'Height,' and 'ShapeArea' of the building are the most important features that impact the labeling of a sample in the class with the UHI effect. Furthermore, the results show that when the value of the feature of 'BuiltAreaG200' increases to more than 10000 m2, the building is more likely to be classified in the UHI class 2. It means that the building probably experiences the UHI effect more than 30°C. To check the results of feature importance, we considered the initial dataset used for the model creation. The samples show that just 5.9% of samples with a value of 'BuiltAreaG200' greater than 10000 m2 are in the UHI class 1 (lower than 30°C). Therefore, the transparent ML model results show that the feature importance in the classified model is trustable. Therefore, the Explainable model can be applied to develop mitigation strategies to improve the quality of the microclimate conditions in the urban area.

Acknowledgments

References

1. Bertinelli, Luisito, and Duncan Black, "Urbanization and growth," *Journal of Urban Economics*. 56, 1, pp. 80-96, 2004.

2. Evyatar Erell, David Pearlmutter, and Terence Williamson., Urban microclimate: designing the spaces between buildings, Urban Microclimate, 2012.

3. Muhammad Farhan Ul Moazzam, Yang Hoi Doh, Byung Gul Lee, "Impact of urbanization on land surface temperature and surface urban heat Island using optical remote sensing data: A case study of Jeju Island, Republic of Korea," *Building and Environment,* 22, 109368, 2022.

4. Or Aleksandrowicz, Milena Vuckovic, Kristina Kiesel, Ardeshir Mahdavi, "Current trends in urban heat island mitigation research: Observations based on a comprehensive research repository," *Urban Climate,* 21, 1-26, 2017.

5. Ardalan Aflaki, Mahsan Mirnezhad, Amirhosein Ghaffarianhoseini, Ali Ghaffarianhoseini, Hossein Omrany, Zhi-Hua Wang, and Hashem Akbari, "Urban heat island mitigation strategies: A state-of-the-art review on Kuala Lumpur, Singapore and hong kong," *Cities.*

6. Eslamirad, Nasim, Francesco De Luca, and Kimmo Sakari Lylykangas, "The role of building morphology on pedestrian level comfort in Northern climate," 2021.

7. Kovacs-Györi, A., Ristea, A., Havas, C., Mehaffy, M., Hochmair, H. H., Resch, B., Juhasz, L., Lehner, A., Ramasubramanian, L., Blaschke, T., "Opportunities and Challenges of Geospatial Analysis for Promoting Urban Livability in the Era of Big Data and Machine Learning," *ISPRS International Journal of Geo-Information,* 9, 12, 752, 2020.

8. Balogun, Abdul-Lateef, Abdulwaheed Tella, Lavania Baloo, and Naheem Adebisi, "A review of the inter-correlation of climate change, air pollution, and urban sustainability using novel machine learning algorithms and spatial information science," *Urban Climate,* 40, 100989.

9. D. P. a. T. W. Evyatar Erell, Urban microclimate: designing the spaces between buildings, Urban Microclimate, 2012.

10. Byung Gul Lee Muhammad Farhan Ul Moazzam, Yang Hoi Doh., "Impact of urbanization on land surface temperature and surface urban heat island using optical remote sensing data: A case study of Jeju island, republic of Korea.," *Building and Environment,* 22, 2022.

11. Akio Onishi, Xin Cao, Takanori Ito, Feng Shi, and Hidefumi Imura, "Evaluating the potential for urban heat-island mitigation by greening parking lots. Urban Forestry Urban Greening," 9, 4, 323–332.

12. J. K. Ching, "A perspective on urban canopy layer modeling for weather, climate and air quality applications," *Urban Climate,* 3, 2013.

13. B. Stone, The City and the Coming Climate, Cambridge University Press, 2012.

14. Camila Mayumi Nakata-Osaki, Léa Cristina Lucas Souza, Daniel Souto Rodrigues, "THIS – Tool for Heat Island Simulation: A GIS extension model to calculate urban heat island intensity based on urban geometry," *Computers, Environment and Urban Systems,* 67, 157-168, 2018.

15. Milojevic-Dupont, Nikola, and Felix Creutzig, "Machine learning for geographically differentiated climate change mitigation in urban areas," *Sustainable Cities and Society ,* vol. 64, p. 102526, 2021.

16. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, ""Why Should I Trust You?" Explaining the Predictions of Any Classifier," *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, Association for Computing Machinery.,* 1135–1144, 2016.

17. "General data of Tallinn,". Available: https://www.tallinn.ee/en/statistika/general-data-tallinn.

18. "Buildingdata," [Online]. Available: https://livekluster.ehr.ee/ui/ehr/v1/infoportal/buildingdata.

19. "keskkonnateadlik.," [Online]. Available: https://keskkonnateadlik-kaur.hub.arcgis.com.

20. Eslamirad, Nasim, Francesco De Luca, Kimmo Sakari Lylykangas, Sadok Ben Yahia, Mahdi Rasoulinezhad, "Geoprocess of geospatial urban data in Tallinn, Estonia," Data in Brief, p. 109172, 2023.

21. Valentina Sagris and Mait Sepp, "Landsat-8 TIRS Data for Assessing Urban Heat Island Effect and Its Impact on Human Health," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 12, pp. 2385-2389, 2017.

22. A. Páez, "Spatial statistics for urban analysis: A review of techniques with examples," GeoJournal, 61, no. 53, 2005.

23. Peel, Murray C., Brian L. Finlayson, and Thomas A. McMahon., "Updated world map of the Köppen-Geiger climate classification," Hydrology and earth system sciences, vol. 11, no. 5, pp. 1633-1644, 2007.

24. J. Mondal, "Debunking Thumb-rule: Orient Longer Faces of Buildings towards North-South to Minimise Solar Gain in India," 2018.

25. "Weather Underground.," [Online]. Available: https://www.wunderground.com/about/our-company.

26. "Mendeley reposity," [Online]. Available: https://data.mendeley.com/datasets/2bm7kdf8gb. [Accessed 23 04 2023].

27. Z. C. Lipton, "The Mythos of Model Interpretability," Communications of the ACM, vol. 61, no. 10, 2016.

Paper IV
Eslamirad, N., Sepúlveda, A., De Luca, F., Sakari Lylykangas & Ben Yahia, S. (2023). Outdoor thermal comfort optimization in a cold climate to mitigate the level of urban heat island in an urban area. *Energies, 16*, 4546. https://doi.org/10.3390/en16124546

MDPI

Article

Outdoor Thermal Comfort Optimization in a Cold Climate to Mitigate the Level of Urban Heat Island in an Urban Area

Nasim Eslamirad [1],*, Abel Sepúlveda [2,3], Francesco De Luca [3], Kimmo Sakari Lylykangas [3] and Sadok Ben Yahia [4]

[1] FinEst Centre for Smart Cities, Tallin University of Technology, 10115 Tallinn, Estonia
[2] Architecture and Intelligent Living, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany; abel.luque@taltech.ee
[3] Department of Civil Engineering and Architecture, Tallinn University of Technology, 10115 Tallinn, Estonia; francesco.deluca@taltech.ee (F.D.L.); kimmo.lylykangas@taltech.ee (K.S.L.)
[4] Department of Software Science, Tallinn University of Technology, 10115 Tallinn, Estonia; sadok.ben@taltech.ee
* Correspondence: nasim.eslamirad@taltech.ee

Abstract: Climatic and micro-climatic phenomena such as summer heat waves and Urban Heat Island (UHI) are increasingly endangering the city's livability and safety. The importance of urban features on the UHI effect encourages us to consider the configuration of urban elements to improve cities' sustainability and livability. Most solutions are viable when a city redevelops and new areas are built to focus on aspects such as optimum design and the orientation of building masses and streets, which affect thermal comfort. This research looks beyond outdoor thermal comfort studies using UHI data and geoprocessing techniques in Tallinn, Estonia. This study supposes that designing urban canyons with proper orientation helps to mitigate the UHI effect by maximizing outdoor thermal comfort at the pedestrian level during hot summer days. In addition, optimizing the orientation of buildings makes it possible to create shaded and cooler areas for pedestrians, reducing surface temperature, which may create more comfortable and sustainable urban environments with lower energy demands and reduced heat-related health risks. This research aims to generate valuable insights into how urban environments can be designed and configured to improve sustainability, livability, and outdoor thermal comfort for pedestrians. According to the study results, researchers can identify the most effective interventions to achieve these objectives by leveraging UHI data and geoprocessing techniques and using CFD simulations. This evaluation is beneficial in guiding urban planners and architects in proposing mitigation solutions to enhance thermal comfort in cities and creating suitable conditions for approved thermal comfort levels. Results of the study show that in the location used for the survey, Tallinn, Estonia, the orientation of West-East offers the optimum level of comfort regarding thermal comfort and surface temperature in the urban environment.

Keywords: urban climate changes; outdoor thermal comfort; Urban Heat Island (UHI); surface temperature; mitigation strategies of UHI effect

1. Introduction

According to the European Environment Agency (EEA) [1], the number of megacities has nearly tripled since 1990 [2]. Recognizing the paramount importance of providing secure, healthy, and comfortable housing for individuals' overall well-being, urban design plays a critical role in mitigating the adverse effects of climate change on cities [2]. The rapid expansion of urban areas brings about significant alterations in surface temperatures, particularly in densely populated regions characterized by impermeable surfaces that absorb substantial solar radiation, resulting in heat retention within buildings [3]. By effectively managing the factors that influence the urban microclimate, the quality of life

for city residents can be significantly improved [2]. Conversely, inadequate urban design exacerbates the impacts of climate change in urban areas [1].

The UHI effect is a specific phenomenon associated with urban environments, leading to substantially higher temperatures than surrounding rural areas [2]. UHI contributes to a 2–5 °C temperature rise in urban areas [3]. The low level of thermal comfort in cities emphasizes the urgent need for urban planners to prioritize sustainability and reevaluate their approaches. This is particularly significant due to the far-reaching impacts of urban warming on health, well-being, human comfort, and the local atmosphere [4,5], as well as the economic and social systems of cities [1]. Brian Stone highlighted in 2012 that major cities worldwide are experiencing temperature increases faster than the rest of the planet [6].

Consequently, there is a pressing need to present mitigation strategies to address the exponential growth of UHI and heat waves associated with rising urban temperatures. As a result, scientific interest in mitigating UHI has increased, reflecting an increased awareness among scientists, urban planners, and governmental organizations [7,8]. This is primarily due to UHI's direct impact on urban residents' health [4,9] and its implications for their well-being, human comfort, and the local atmosphere [8].

Research on the UHI phenomena often focuses on the canopy layer and investigates it at micro and local scales, such as single-street canyons and neighborhoods [10]. The design configuration of urban areas, including optimized building and street geometry and orientation, plays a crucial role in influencing solar radiation and airflow within an urban canyon [11]. Among the various measures used to assess UHI, ambient temperatures, including air and surface temperatures, are particularly important [4]. In urban street canyons, the amount of solar radiation directly impacts solar access and, consequently, the thermal comfort experienced by pedestrians [11]. Hence, incorporating these strategies into urban development plans could create more sustainable, resilient, and livable cities [12].

However, there are still limitations in studies concerning outdoor thermal comfort, mainly due to the inefficiency of simulations and the challenge of applying assessment results during the early stages of the design process [13]. Furthermore, UHI mitigation strategies still face several hurdles, such as the complexity of execution and the difficulty in effectively communicating scientific knowledge to municipal governments and urban planners [7].

While various strategies for mitigating the UHI effect have been acknowledged, there is still a need to bridge the gap between the accumulated knowledge and the practical implementation of UHI mitigation measures. One approach to address this is to consider the influence of outdoor comfort levels and surface temperatures during the design process of urban elements and building extensions. Furthermore, Computational Fluid Dynamics (CFD) and numerical simulations allow researchers to assess design scenarios and compare their effectiveness in mitigating UHI and improving outdoor thermal comfort in urban areas.

Furthermore, rather than retrofitting existing urban areas, it is advantageous to incorporate UHI mitigation strategies right from the start when developing new cities. Since thermal comfort is a crucial factor in design considerations [14], analyzing the geometry and orientation of urban canyons and surrounding buildings enables researchers to determine the optimal configuration that maximizes shading, airflow, and solar radiation. This optimization approach helps minimize the UHI effect and enhance outdoor thermal comfort.

This study offers a unique contribution through its innovative approach of sampling urban areas using geoprocessed urban data. It extensively analyzes the relationship between building orientation, outdoor thermal comfort, urban surface temperature, and the UHI effect, specifically in Tallinn, Estonia. Additionally, the study introduces a novel method for analyzing the Physiological Equivalent Temperature (PET) value, utilizing a scaling evaluation method. This study provides a valuable opportunity to gain insights into the intricate interplay between urban features, microclimate conditions, and human well-being through the implementation of optimization research. The findings obtained

from this research can inform and guide decision-making processes in urban design and planning. This study offers valuable insights that can contribute to developing informed and effective strategies for mitigating the UHI effect. By incorporating these insights, urban planners, architects, and policymakers can make informed decisions and create design guidelines that promote sustainable and livable urban environments.

The essence of this research as an optimization urban research has three main objectives:

Objective 1: To evaluate the outdoor thermal comfort at the pedestrian level in different scenarios of building mass orientation in the urban canyon. This objective holds significance as it directly impacts the well-being of city residents and aids in identifying the most favorable and sustainable design solutions for residential areas.

Objective 2: To evaluate the surface temperature of different building mass orientation scenarios. Surface temperature plays a pivotal role in the UHI effect, and comprehending its behavior in different scenarios is essential for identifying effective UHI mitigation strategies and making informed design decisions.

Objective 3: To propose solutions that improve the quality of life regarding outdoor thermal comfort to help reduce the effect of UHI in Tallinn. This objective holds great importance in light of the growing significance of urbanization and climate change as global concerns.

After the Introduction, the paper is divided into five sections:

Section 2 provides an overview of the literature review and related works based on an extensive review of published academic research studies. Section 3 discusses the material and methods applied in the study. Section 4 presents the numerical analysis and CFD simulation to assess the outdoor thermal comfort and surface temperature of case studies in different scenarios. Section 5 is related to the assessment results. The discussion follows in the next section with the interpretation of the results and the study's conclusion.

2. Background

2.1. UHI Effect and Surface Temperature Studies

Rapid global urbanization has resulted in extensive urban development, with a significant portion of the global population residing in cities. This trend is projected to increase to around five billion people, or 61 percent of the global population, by 2030 [2]. In Europe, the percentage of the population living in cities is currently around 73%, with an expected rise to 82% [1]. Due to urbanization and global climate change, urban areas experience higher temperatures than non-urban regions. This trend is expected to continue throughout the 21st century, leading to higher temperatures in urban areas than in non-urban regions due to the UHI phenomenon [15]. The phenomenon of UHI resulting from urbanization was first observed in 1818 by Howard [5]. The occurrence of heatwaves and the UHI effect presents significant climate risks to cities, leading to extensive research efforts to explore various methods to mitigate their impacts [6]. The intensity of the UHI is a measure of the additional heat introduced into the atmosphere by urban areas [16]. Changes in the urban thermal environment substantially impact the energy balance within urban areas, affecting boundary meteorology and climatology [3]. These factors have socio-economic implications, including increased energy consumption, heightened vulnerability to heat-related illnesses, and higher mortality rates [17]. Consequently, it is essential to investigate UHI mitigation strategies sustainably [4]. Accurate analysis and understanding of UHI's spatial and temporal variations of UHI are crucial for effective management and mitigation efforts [17].

Mitigating the UHI phenomenon through physical environmental modifications is crucial for altering the urban microclimate. Implementing large-scale mitigation measures encompassing the entire urban environment can significantly impact the urban microclimate [7]. The research conducted by Akbari, Rosenfeld, and Taha at the Lawrence Berkeley National Laboratory played a crucial role in popularizing the concept of UHI mitigation [7]. UHI mitigation is essential for improving human thermal comfort and creating better living environments in urban residential areas. However, limited attention has been given to understanding the combined effects of UHI mitigation strategies on human thermal

comfort [18]. Akbari suggests various strategies to reduce UHI and enhance thermal comfort in cities, highlighting that the intensity of UHI is influenced by urban characteristics, micro-climatic conditions, urban materials, and green spaces within urban areas [1]. The arrangement of buildings and the land's topography play crucial roles in determining temperature distribution within a city [19]. Additionally, urban geometry plays a pivotal role in controlling the retention and release of heat, emphasizing its significance in the UHI phenomenon. Therefore, it is vital to understand how urban geometry influences these factors to implement effective strategies for UHI mitigation [17].

In 2009, Giguère compiled a comprehensive inventory of UHI mitigation strategies, categorizing them into four groups. These categories include vegetation and cooling measures such as selective tree planting, greening of parking lots, and implementing green roofs. The second category is sustainable urban infrastructure, focusing on designing buildings and roads with UHI mitigation in mind. Third, sustainable water management is another category that utilizes trees, green roofs, permeable surfaces, and retention ponds to manage water in urban areas. Lastly, reducing anthropogenic heat involves controlling heat production, reducing vehicle numbers, and implementing efficient air conditioning systems [7].

Dynamic numerical approaches are the most reliable and satisfactory method for assessing the effectiveness of UHI mitigation strategies [5]. Similarly, in a study conducted in Toronto, various UHI mitigation strategies were evaluated in different urban neighborhoods using numerical simulations with the ENVI-met software to gauge their impact on reducing UHI effects. The study revealed that urban form significantly influences the duration of direct sun and mean radiant temperature, which are crucial factors in determining urban thermal comfort [20].

In a study conducted in the Sydney metropolitan area, researchers examined the impact of various urban design factors on ambient and surface temperatures in open spaces. Factors such as building height, street width, aspect ratio, built area ratio, orientation, and dimensions of open spaces were analyzed. Using the ENVI-met simulation tool, the study developed fourteen precincts to simulate different scenarios, both with and without mitigation measures. The results demonstrated a strong correlation between the gradient of temperature decrease along the precinct axis (GTD) and the average aspect ratio of the precincts, regardless of whether mitigation strategies were implemented. As a result, the study suggests that implementing urban design interventions that modify the aspect ratio of buildings and streets can effectively mitigate the UHI effect and improve thermal comfort in open spaces [21].

Additionally, Xu et al. conducted a study investigating the potential of using the spatial equity of green areas in cities and land surface temperature to mitigate UHI effects. Their findings indicated that increasing the amount of urban green spaces can be beneficial in reducing the average urban temperature and mitigating UHI effects [22]. Similarly, another study focused on improving UHI in Mandaue, Philippines, through various mitigation measures, including increasing vegetation, adding open spaces, employing green roofs, and combining these strategies. The study considered changes in air temperature, surface temperature, and thermal comfort in the study areas to understand the impact of altering green areas and implementing green roofs on reducing the UHI effect [23].

Using ENVI-met simulations, another study investigates how the built environment impacts microclimate parameters. It confirms the existence of the UHI phenomenon in Chennai, India, emphasizing the importance of urban planning in designing neighborhoods that prioritize thermally comfortable outdoor spaces to enhance pedestrian comfort. These findings have significant implications for urban planners, underscoring the need to consider thermal comfort when designing outdoor areas [24]. Similarly, the study explores the influence of urban form parameters on pedestrian thermal comfort in the arid climate of Mashhad, Iran. By employing the ENVI-met software, the researchers analyze these parameters to predict outdoor thermal comfort conditions in current and future urban contexts. To assess outdoor thermal comfort, the study proposes an alternative approach for cities,

advocating using UHI zoning to replace traditional urban form zoning. This alternative method proves particularly advantageous in large cities where gathering data on the urban form is challenging due to limited resources and time constraints. By incorporating UHI zoning, urban planners can effectively evaluate and enhance outdoor thermal comfort in urban areas [25].

In another study focusing on hot climate conditions, Farhadi et al. assessed various strategies for mitigating the UHI effect and improving thermal comfort in Tehran, which experiences urban warming. Their findings revealed a strong correlation between lower surface temperatures and the reduction of the UHI effect, leading to improved thermal comfort [26]. Similarly, Arnfield's research indicated that the orientation of streets plays a significant role in determining the amount of solar energy absorbed by walls [27]. Likewise, Van Esch et al. examined the effects of street width and orientation, as well as building parameters such as roof shape and building envelope, on solar access to the urban canopy and the viability of passive solar heating strategies in residential buildings [28]. Furthermore, evaluations of the UHI effect in numerous cities and villages across the Netherlands demonstrated a significant UHI in most Dutch cities. The 95th percentile of the UHI is strongly correlated with population density [29]. Additionally, the design of streets, the orientation of urban canyons, and the presence of trees had a remarkable impact on ground surface temperatures and outdoor thermal comfort, consequently influencing the UHI effect [30].

2.2. Outdoor Thermal Comfort Studies

The potential for reducing outdoor air temperatures in a square in Rome was studied using a numerical model created using the ENVI met tool to simulate different mitigation scenarios to reduce warming in urban areas. The study found solutions such as using grass pavers to provide the most significant advantages that could enhance the thermal conditions of the air and reduce outdoor air temperatures [31].

Fazia Ali-Toudert et al. discuss the role of street design, such as aspect ratio, and orientation, towards developing pedestrian-level comfortability. The study benefits of the three-dimensional numerical model ENVI-met, simulating microclimatic changes within urban environments in a high spatial and temporal resolution in Ghardaia, Algeria. The study analyzed the symmetrical urban canyons with various height-to-width ratios and different solar orientations (i.e., East–West, North–South, North East–South West, and North West–South East). In addition, the study assessed outdoor thermal comfort value in the physiologically equivalent temperature (PET) index. The results show contrasting patterns of thermal comfort between shallow and deep urban streets and the various orientations. Moreover, the results prove that PET at the street level depends strongly on aspect ratio and street orientation [32]. Another study evaluated the potential for UHI mitigation of greening parking lots and the relationships between land surface temperature (LST) and land use/land cover (LULC) in different seasons in Nagoya. The results show that different LULC types play different roles in different seasons and times, and using more green areas slightly reduced the LST for the whole study area in spring or summer [3]. The other study that used qualitative and quantitative approaches to assess outdoor thermal comfort as a mixed method identified which urban areas need more improvement during the summer. The results of thermal comfort assessment through the PET index and subjectively perceived thermal sensation using ENVI-met environment to do CFD simulation and thermal comfort assessment [33].

In the study by Giridharan, the author defines urban compactness as a combination of various urban design factors, including the building area-to-volume ratio, aspect ratio (height to width), sky view factor, distance to the nearest wall, width of the street, built-up area, green areas, albedo, water surface size, roads, open areas, and distance to a heat sink [10]. These factors influence the urban microclimate, and their combination can affect the level of UHI and outdoor thermal comfort in urban areas. In addition, Aleksandrowicz et al. outline the physical features of the urban environment, such as the

density of buildings, the area of land used and unoccupied areas, and the type of materials in urban components, which all affect UHI level and strength [7].

The background studies and literature review show that heat waves and the UHI effect are significant climate risks affecting cities. There are many ways in which urban design can be modified to mitigate the UHI effect in cities, such as increasing green spaces, using reflective or high albedo materials, modifying the built environment, reducing the anthropogenic heat, and optimizing building and urban canopy orientation and layout. These modifications can help decrease surface temperatures in urban areas, which can significantly impact the UHI effect and the level of thermal comfort experienced by people living and working in these environments. Additionally, by evaluating the orientation and extension of buildings and urban canyons that affect the amount of solar radiation, which impacts local temperatures and, consequently, the surface temperature and thermal sensation in the urban area, we can help to lower temperatures and improve thermal comfort at the pedestrian level, the surface temperature in the urban canyon, and UHI effect in the urban area during hot summer days. Furthermore, these modifications can also provide other benefits, such as enhancing the comfort and livability of urban areas and creating more sustainable and resilient environments that promote the health and well-being of all residents. The study's findings will be valuable to urban planners and designers as they influence decisions about designing and configuring urban elements to create more sustainable and comfortable environments.

2.3. Novelty of This Investigation

The literature review highlights the significant influence of city layout and structure on heat waves and the UHI effect. Previous research emphasizes dynamic numerical approaches to assess UHI, outdoor thermal comfort, and surface temperatures, enabling the identification of mitigation strategies.

This study suggests designing urban canyons with proper orientation as an effective method to mitigate the UHI effect and enhance pedestrian comfort during hot summer days. Optimizing building orientation creates shaded and cooler areas, reducing surface temperatures and heat concentration. This approach promotes more sustainable urban environments with lower energy demands and decreased health risks.

The study's novelty lies in the innovative sampling technique using geoprocessed UHI urban data from Tallinn. Extensive numerical analysis establishes the relationship between building orientation, outdoor thermal comfort, urban surface temperature, and the UHI effect. In addition, a new method for analyzing the Physiological Equivalent Temperature (PET) value is introduced. By prioritizing resident well-being, addressing heat waves, and proposing UHI mitigation strategies, this research contributes to the development of livable and sustainable urban spaces.

3. Methodology

Using three specific case studies, we examined the space between a target building and its nearest neighboring building across the street. Each case study involved a geometric model of an actual residential building in Tallinn, including the urban canopy, the nearby neighbor building(s), and the street in between. These case studies aimed to determine the optimal orientation of the building mass that would ensure the highest level of outdoor thermal comfort and the lowest surface temperature in the analyzed area during hot summer days.

The building direction often describes the orientation of the canyon axis (e.g., North–South, East–West) or (North West–South East, North East–South West) [2]. In the definition of scenarios and the simulated models of the study, the orientation of the canyon axis represents the direction of an elongated space, measured (in degrees) as the angle between a line running North-South and a significant axis running the length of a street or other linear area, measured counterclockwise. Figure 1 shows mentioned orientations on the axis of four main directions.
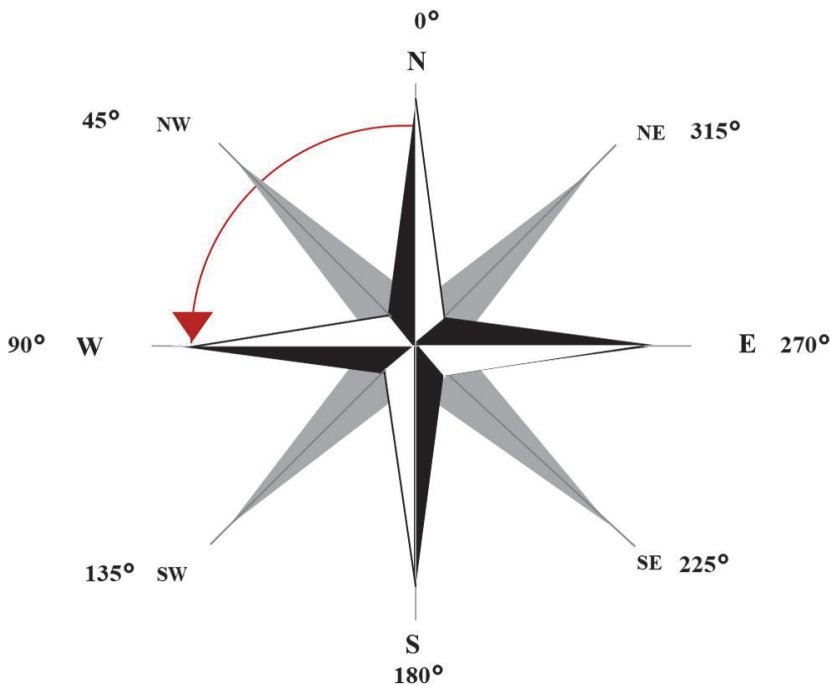
Figure 1. Main eight orientations of the urban environment used in the study.

In this study, we examine real buildings in their current orientation as well as hypothetical orientations. The orientations of 0°–180° correspond to the extension of the urban canyon in the North-South and South-North directions, while 270°–90° represent the extension in the East-West direction. Additionally, the orientations of 45°–225° indicate the North West-South East extension, and 135°–315° refer to the South West-North East extension. The specifications of each case study are summarized in Table 1.

Table 1. Features of the main building in case studies.

Sample	Height (m)	Floors	Length (m)	Total Area (m²)
Case study 1/M1	50	17	291.3	2258.4
Case study 2/M2	45.4	14	208.5	2093.3
Case study 3/M3	37.3	9	202.8	2470.6

Figure 2 displays three selected residential buildings in Tallinn as case studies.

Table 2 outlines the characteristics considered when defining different scenarios during modeling and simulations for case studies 1 through 3.

Case study 1: XY coordination based on Google earth: 59°24′51″N24°44′18″E
Harju county, Tallinn, Kesklinna district, Pärnu mnt 110

Case study 2: XY coordination based on Google earth: 59°26′08″N24°49′15″E
Harju county, Tallinn, Lasnamäe district, Pae tn 68

Case study 3: XY coordination based on Google earth: 59°25′57″N24°46′45″E
Address: Harju county, Tallinn, Kesklinna district, Vesivärava tn 50

Figure 2. Case studies, three residential buildings in Tallinn, Estonia.

The methodology of the study is designed in five steps, starting with step 0 to acquire data and finishing with step 4 to make an inventory and explain the application of the study.

According to Figure 3, the steps are in the following order:

- Step 0: Capturing data
- Step 1: Sampling
- Step 2: Simulation
- Step 3: Assessment
- Step 4: Application

Table 2. The general features of the simulated models.

Scenarios of Simulated Case Studies in the Different Extensions of the Canopy			
Model	**Case Study**	**Orientation (°)**	**Extension**
M1	Cs1	347	NE-SW
M2	Cs2	22	N-S
M3	Cs3	325	NE-SW
M1-1	Cs1	0	N-S
M1-2	Cs1	45	NW-SE
M1-3	Cs1	90	W-E
M1-4	Cs1	135	SW-NE
M1-5	Cs1	180	S-N
M1-6	Cs1	225	SE-NW
M1-7	Cs1	270	E-W
M1-8	Cs1	315	NE-SW
M2-1	Cs2	0	N-S
M2-2	Cs2	45	NW-SE
M2-3	Cs2	90	W-E
M2-4	Cs2	135	SW-NE
M2-5	Cs2	180	S-N
M2-6	Cs2	225	SE-NW
M2-7	Cs2	270	E-W
M2-8	Cs2	315	NE-SW
M3-1	Cs3	0	N-S
M3-2	Cs3	45	NW-SE
M3-3	Cs3	90	W-E
M3-4	Cs3	135	SW-NE
M3-5	Cs3	180	S-N
M3-6	Cs3	225	SE-NW
M3-7	Cs3	270	E-W
M3-8	Cs3	315	NE-SW

Figure 3. The framework for the study.

3.1. Step 0 "Urban Data Geoprocessing"

This research concentrates on the residential buildings with the greatest volume (maximum height and area) that went through an intense heatwave and UHI in 2014, 2018, and 2019. Moreover, case studies were chosen by sampling and utilizing the histogram to locate the more critical and more severe residential buildings from the Tallinn UHI dataset [34]. Thus, each case study is an actual building in Tallinn, Estonia (Lat. 59°26′ N Lon. 24°45′ E); the country has a humid continental climate with mild summers, as mentioned in the Köppen–Geiger classification in Dfb class [35] as Figure 4 shows.
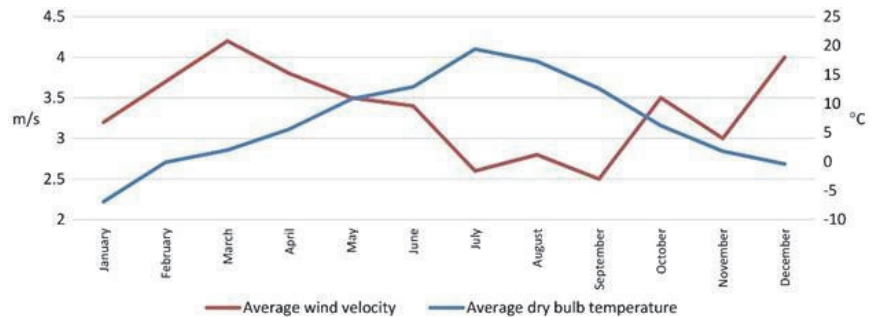
Figure 4. Monthly wind velocity and temperature averages of Tallinn, Estonia.

An extensive multidisciplinary presented dataset is collected with 34,001 building samples (rows) and 30 features (columns) in Tallinn, including the characteristics of the buildings, for example, Material, Height (m), Absolute Height (m), Number of Floors Above Ground, Shape Length (m), Shape Area (m^2), the spatial indices of buildings on the hierarchical system grid, such as Built-up area (G200, level 1), Urban Density D1 (G200, level 1), Urban Density D2 (G1000, level 2), Urban Density D1 (G2000, level 3), Average Building Area in G200 (m^2), Max Area in G200 (m^2), Number of Buildings in G200, the defined indexes based on buildings and different zones in the city area, the Nearest Neighbour (m), Green Area in G200 (m^2), the Ratio of Green Area/Grid Area (G200), Purpose of Building, Main Angle, Orientation, Height to Width (G200), the weather data on dates that the city experienced UHI phenomena, and the UHI value of each building [34,36].

The methodology to acquire data in the geoprocessed UHI dataset proposes a framework to categorize data into homogeneous or heterogeneous, static, or dynamic schemes and then collect data considering the homogeneous grid system [34]. Capturing data is the implementation of the hierarchical grid system in the data collection process:

- First, create a spatial index for each object and connect the objects to the grid system.
- Second, use the homogeneous ground to define urban indices mainly anchored in the heterogeneous data. The methodology uses the Python, Numpy, and Pandas libraries, the Geopandas package, and QGIS Tool. The approach helps to capture urban data from Tallinn GIS resources [37], taking into account the location, general characteristics, and other spatial properties of urban elements [34,37].

3.2. Step 1: Sampling, Finding Critical Urban Canyons

Our initial analysis investigated the relationship between the urban dataset and the Urban Heat Island (UHI) effect. The findings, depicted in Figure 5, revealed significant correlations and dependencies between various features of the dataset and UHI levels. Notably, we identified the size of the building, its height, and its area as influential factors affecting the UHI value. With this understanding, we focused our attention on the area and height of the buildings to identify large-volume structures in Tallinn using the collected data.
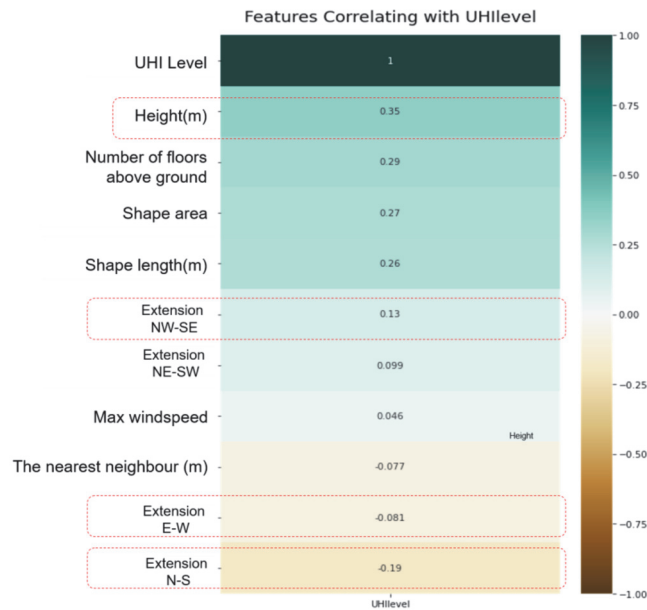
Figure 5. Dependency values between the features of the dataset and UHI level.

Moreover, our study placed particular emphasis on residential buildings, recognizing them as critical cases in the city, as they are susceptible to higher temperatures based on UHI data. To streamline our analysis, we initially filtered the dataset by prioritizing residential buildings, narrowing our focus.

As we delved into the analysis, we specifically examined the footprint area of the buildings. However, we discovered that the data exhibited a random distribution without any discernible patterns. Figure 6a visually represents this observation, indicating that buildings with a footprint area exceeding 2000 m^2 were less common in the dataset. Consequently, we conducted a more comprehensive investigation of the entire dataset, specifically targeting buildings with an area greater than 2000 m^2 to identify the infrequent occurrence of taller structures within this subset.
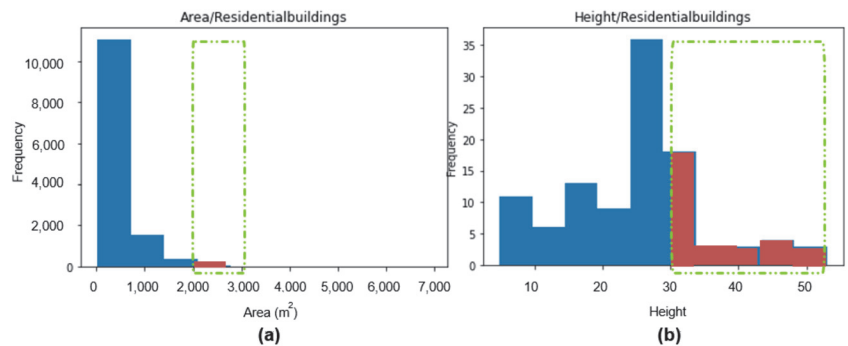
Figure 6. Sampling by using a histogram to choose case studies from the urban UHI dataset in Tallinn. (**a**) The highest value of the area in the residential buildings, (**b**) The highest value of the height in the residential buildings.

Thus, as the histogram shows, filtering the dataset to the highest volume buildings helps us to find which buildings are the critical cases to study. In addition, Figure 6a shows samples in the highest area and height of residential buildings in the UHI dataset (more correlated features with the UHI effect) with an area of more than 2000 m^2 and a height of over 30 m. Conversely, the graph shows fewer samples with an area higher than 2000 m^2 and a height higher than 30 m. Consequently, Figure 6b reveals samples that meet the research question's goal, pointing to some residential buildings in the UHI dataset with an area greater than 2000 m and a height above 30 m. The selected samples with an area of more than 2000 m and a height over 30 m are shown in red.

3.3. Step 2: Simulation in Building and Urban Scale

In step 2, we focused on the geometric modeling and simulation process of case studies in different scenarios (Table 2) to verify outdoor thermal comfort and solar access. Based on the sampling results and the UHI dataset, we have chosen three urban canyons with high-volume residential buildings in Tallinn, Estonia. The urban canopy layer refers to the space enclosed by the vertical boundaries of urban buildings, extending up to their rooftops [38].

To conduct CFD simulation, a three-dimensional computational model of fluid dynamics and energy balance, designed to simulate the microclimate of the study at the street level, was employed. The simulation uses ENVI-met, a software package specifically developed for urban microclimate modeling.

To evaluate outdoor thermal comfort, we employed the Physiological Equivalent Temperature (PET) index, which considers various factors that impact human thermal comfort, including air temperature, humidity, wind speed, and radiation. The PET index utilizes a range of thermal perceptions and physiological stress levels experienced by humans [39]. The PET index is based on the Munich Energy-balance Model for Individuals (MEMI). This two-node model simulates the thermal balance of the human body in a physiologically relevant way [40]. By utilizing this index, we were able to assess the thermal comfort of the outdoor environment across different scenarios and identify any potential issues related to heat stress and discomfort. The thermal perception and corresponding ranges of PET for each thermal comfort class are presented in Table 3. Additionally, we examined surface temperature, as it plays a crucial role in understanding the potential effects of absorbed solar radiation, which can contribute to increased surface temperatures, UHI effects [41], heat waves, and the overall thermal performance of buildings and outdoor spaces.

Table 3. The range of thermal index predicted thermal perception by human beings and physiological stress on human beings [39].

Thermal Perception Grade of Physiological Stress		
PET (°C)	Thermal Perception	Grade of Physiological Stress
	Very cold	Extreme cold stress
4		
	Cold	Strong cold stress
8		
	Cool	Moderate Cold stress
13		
	Slightly cool	Slight cold stress
18		
	Comfortable	No thermal stress
23		
	Slightly warm	Slight heat stress
29		
	Warm	Moderate heat stress
35		
	Hot	Strong heat stress
41		
	Very hot	Extreme heat stress

3.4. Step 3: Assessment and Results

Through the CFD simulation and analysis, output data related to the thermal comfort of people at the pedestrian level and the surface temperature of the urban area under different scenarios were acquired.

To evaluate thermal comfort, we used a comprehensive approach to evaluate the thermal performance of the studied urban area under different scenarios (Table 2). The first thermal comfort analysis is about finding the non-uniform spatial distribution of PET in each particular scenario in the urban canopy between the target building and the nearest neighbor.

The second analysis aims to create a model based on the scoring system to show uniform or normalized spatial distribution. The normalized PET of the urban canyon in each scenario is the weighted PET value by considering the quality and quantity of PET data. The highest value of the weighted PET leads us to find the optimal degree of orientation in the urban canyon.

The scoring system we implemented considers the thermal comfort level at each point in the canopy area. The scoring system allows us to calculate the overall level of comfortability at the pedestrian level by combining the scores of all the individual points. In addition, the surface temperature assessment of the studied areas was performed using the results of the CFD simulation. Overall, the orientations lead to the lowest surface temperature highlighted as the optimum building mass extension and taken into account in the final assessment to determine the best urban environment orientation to ensure comfortability.

3.5. Step 4: Application

Output data related to the optimal building mass orientation is helpful to suppose more sustainable solutions in cities. Furthermore, by improving outdoor thermal comfort and surface temperature in urban areas and leveraging UHI data, the study provides valuable insights into the thermal performance of the studied urban area.

The evaluation is particularly beneficial to guide urban planners and architects in proposing mitigation solutions to enhance thermal comfort in cities and create suitable conditions for achieving approved thermal comfort levels with complementary solar access in the city area. With this information, planners and architects can make more informed decisions about the design of new buildings, the placement of green spaces and other urban elements, and the use of shading devices and other technologies to reduce heat gain and improve outdoor thermal comfort.

Overall, the study's findings highlight the importance of considering outdoor thermal comfort and solar access in urban design and planning to create more livable and sustainable cities.

4. CFD Simulation and Numerical Analysis

The section is related to the CFD simulation and the numerical analysis to assess the outdoor thermal comfort and surface temperature in the studied areas. CFD simulation is used to simulate the microclimate of the studied area at the street level, considering the influence of various factors on thermal, such as air temperature, humidity, wind speed, and radiation.

By combining CFD simulation and numerical analysis, we can comprehensively understand the studied area's thermal performance under different scenarios. This information can then identify potential heat stress and discomfort issues and propose mitigation solutions to improve outdoor thermal comfort.

Table 4 and Figure 7 show more detailed information about the areas focused on CFD simulation and the area of the urban canyon in which PET results were evaluated.

Table 4. Detailed information about the CFD simulated areas of each case study.

Case Studies	Canopy Area (m^2)	Simulated Area (m^2)
Case study 1	20,650	5000
Case study 2	15,000	770
Case study 3	12,700	1400
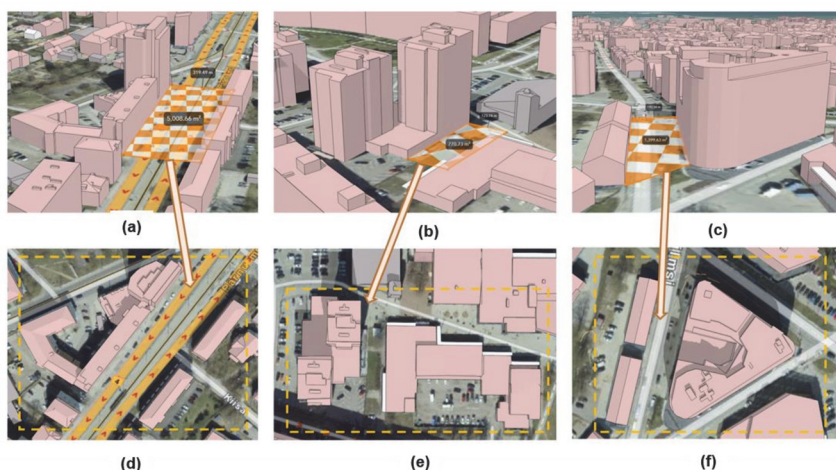
Figure 7. The canopy areas (the hatched rectangular shapes) and the whole simulated areas of each case study (the dashed rectangular shapes). (**a**) The canopy area, cases study 1, (**b**) The canopy area, cases study 2, (**c**) The canopy area, cases study 3, (**d**) The whole simulated area, case study 1, (**e**) The whole simulated area, case study 2, (**f**) The whole simulated area, case study 3.

Different scenarios of case studies were simulated based on the specific orientation of the canopy extension, including east-west (E-W), north-south (N-S), southeast-northwest (SE-NW), and northeast-southwest (NE-SW), as shown in Table 2. Additionally, Table 4 provides details on the simulation area and the corresponding canopy coverage area for each model. Moreover, the canopy area and simulated area for each scenario were determined based on the information presented in Table 4. The geometry modeling process took into account the real physical attributes of the case studies described in Table 1.

In order to perform the CFD simulation using ENVI-met, specific inputs are required, including information about the surrounding environmental features and meteorological data. Additionally, the construction system of the buildings and the surface materials within the simulated urban areas are defined using the material library provided by ENVI-met. For instance, the envelope of the models was defined using a concrete material with moderate insolation properties, which was available in the ENVI-met software's database. Moreover, precise calculations of solar reflectivity and radiation values were conducted, taking into account the specific date, time, and location of the case studies, in order to determine the sun's position during the simulation.

4.1. Meteorological Setting

The input data for the simulation models are the physical properties of the studied urban areas (buildings, soil, and vegetation) and geographic and meteorological data [30].

This study conducted the CFD simulation on 25 July 2014, during a high UHI and heat wave period [37]. The simulated period lasted from 16:00 to 17:00, including the maximum air temperatures during a summer day of 28 °C. The outdoor thermal comfort assessment was conducted at 17:00 and evaluated at 1.80 m, the average human height.

Simple forcing was used in all scenarios to adjust for the meteorological conditions, creating a 24-h weather data cycle that defined the meteorological boundary conditions for the ENVI-met simulation.

Table 5 gives information about the weather condition of Tallinn on the supposed date and time that was chosen for the CFD simulation.

Table 5. The input meteorological data during the CFD simulation by ENVI-met.

Date, 25 July 2014. Time: 17:00	
Air temperature (°C)	Max 28/Min 17
Max relative humidity (%)	Max 75/Min 45
Wind speed at inflow border (m/s)	2.00
Wind direction at inflow (°)	90.00
Roughness length (m)	0.010
Specific humidity in 2500 m (g/kg)	8.00

4.2. Outdoor Thermal Comfort Assessment, PET

PET, expressed in °C, is based on the human energy balance model MEMI and includes the physiological thermoregulatory processes of human beings to adjust to a climatic situation outdoors. The thermal comfort zone for the PET index was initially defined as 18–20 °C [30]. The other classes of thermal comfort are mentioned in Table 3.

In this section, the authors listed all the parameters used in the CFD simulation. In addition, during the simulation, the building's indoor temperature was set to a constant value of 20 °C. Therefore, the outside microclimate did not influence the building temperature. Overall, using the PET index and CFD simulations is a useful approach to assess the thermal comfort of the studied area. Furthermore, by taking into account the physiological thermoregulatory processes of human beings and using advanced simulation techniques, a more accurate and comprehensive understanding of outdoor thermal comfort in urban areas will achieve.

Personal Parameters

Thermal comfort is a subjective concept that depends on personal features and describes a person's state of mind regarding whether they feel comfortable [42]. Thus, once the meteorological data and environmental characteristics are added to the input data used in the CFD simulation, thermal comfort in PET indices needs to set the individual personal data that are supposed as the users of the urban areas. In this study, PET is taken as the outdoor thermal comfort assessment and calculated just for a male pedestrian wearing very light summer clothes standing with a walking speed of 1.2 m/s. For a simple PET assessment process, just male pedestrians wearing unique clothing values with normal body parameters were considered. Table 6 shows other personal parameters used in PET evaluation.

Table 6. Personal parameters applied in PET assessment in ENVI-met simulations.

Basic Personal Parameters	
Age of the person	35
Weight (kg)	75
Height (kg)	1.75
Surface area of the body (sm^2)	1.91
Clo	0.10
Metabolic work (W)	164.70

5. Results

5.1. Surface Temperature

The assessment of the surface temperature of the urban canopy in each scenario is a valuable approach to understanding the impact of changes in canopy orientation on the urban temperature. By analyzing the minimum, maximum, and median values of surface temperature, we can identify the optimum orientation of the canopy for maximizing thermal comfort and decreasing temperatures of urban surfaces. Furthermore, surface temperature is a critical measure in assessing the UHI effect, as it indicates the level of heat absorbed by the surfaces of the urban environment. Thus, by reducing surface temperature, it is possible to mitigate the UHI effect and improve thermal comfort for pedestrians.

Through the surface temperature analysis, we can determine the impact of canopy orientation on surface temperature and identify the optimal orientation that reduces surface temperature and maximizes thermal comfort. This information can inform urban planning and design strategies prioritizing thermal comfort and sustainability. It is important to consider these orientations as they can cause more heat on urban surfaces and potentially result in lower thermal comfort levels for pedestrians.

According to Figure 8a,b, in case studies 1 and 2, scenarios M1-8, M1-6 (24.9 and 28.3 °C) and M2.8, M2.6 (24.3 and 24.4 °C) have the lowest median surface temperatures when oriented at 315° and 225°, respectively. This suggests that these orientations can provide the highest thermal comfort for pedestrians in the case studies. Likewise, Figure 8c shows in case study 3, the median surface temperature data is observed in M3-5 and M3-4 (22.4 and 23.6 °C) with orientations of 180° and 135°, respectively. The finding indicates that these orientations can also provide high thermal comfort for pedestrians in this case study. Furthermore, in case study 1, the orientations with the highest median of surface temperature are M1-1, M1-2, and M1-3 with orientations of 0°, 45°, and 90°, respectively. In case study 2, the orientations with the highest median of surface temperature are M2-1, M2-4, M2-2, and M2-3 with orientations of 0°, 135°, 45°, and 90°, respectively. Finally, in case study 3, the orientations that cause the highest median of surface temperature in the analyzed areas are M3-2, M3-1, M3-3, and M3-8 with orientations of 45°, 0°, 90°, and 315°, respectively.
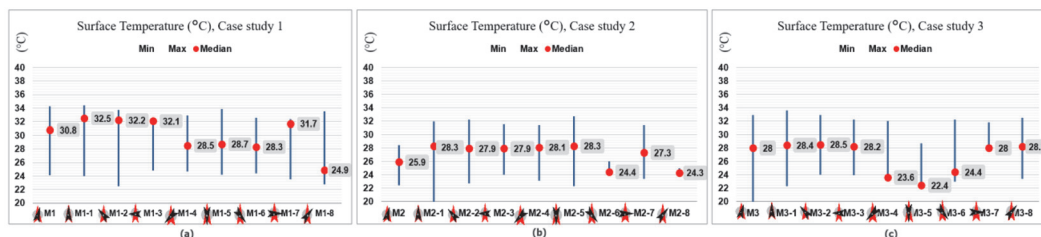
Figure 8. Results of surface temperature (°C) in the urban canopy of different scenarios. (**a**) Case study 1, (**b**) Case study 2, (**c**) Case study 3.

This section presents the results of outdoor thermal comfort, expressed in terms of the PET index and surface temperature in degree centigrade (°C), to choose the best orientations in each case study to lead to the highest comfort level. The spatial distribution of outdoor thermal comfort in terms of the metric PET was calculated via simulation for 27 scenarios of three case studies. The thermal comfort assessment results are explained in two forms, non-uniform and normalized spatial distribution of PET.

5.2. Non-Uniform Spatial Distribution of PET

Figures 9–11 show the initial results of PET assessed for case studies in a different scenario considering the input setting, meteorological data, and parameters during the CFD simulation in the defined canyon orientation, according to Table 2.
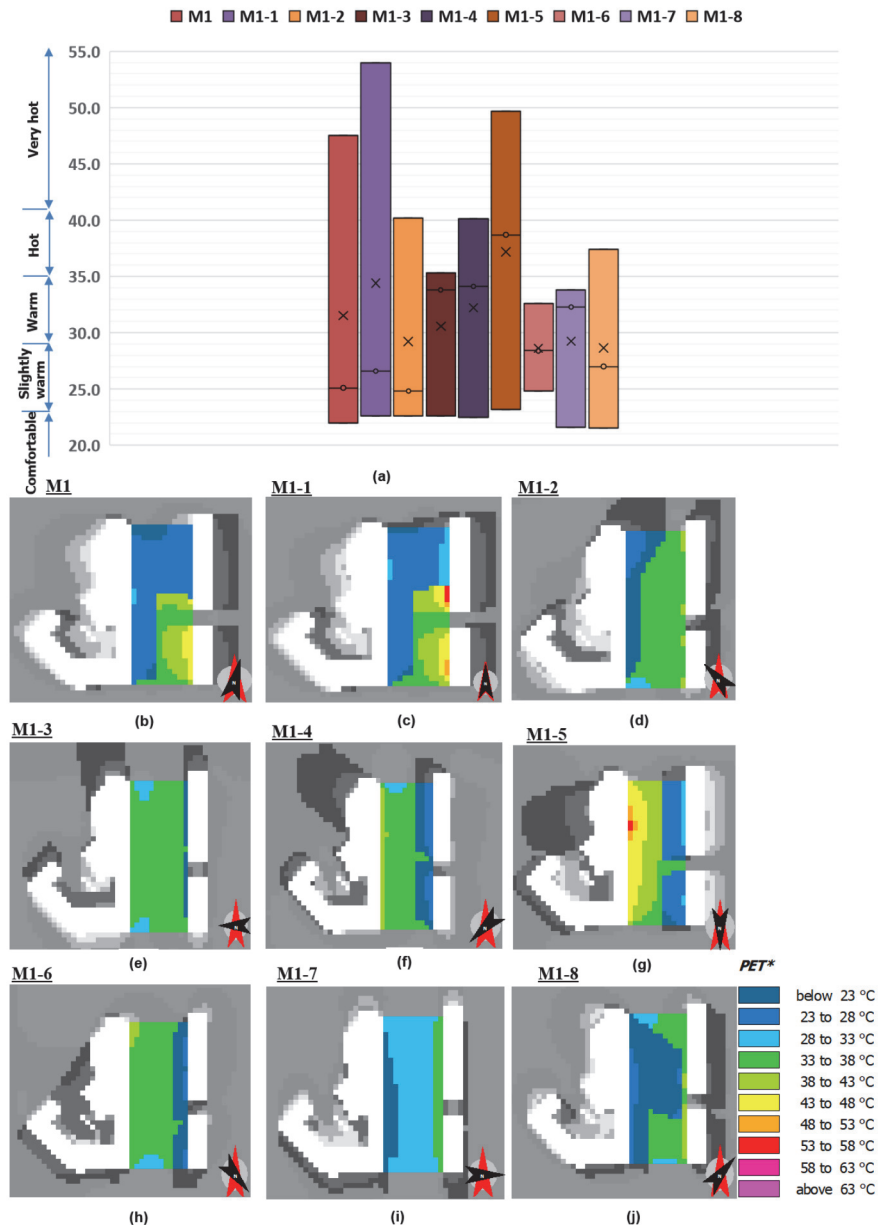
Figure 9. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 1, Scenarios: M1, M1-1 to M1-8. The minimum, maximum, and average scores stand out prominently within the box plots, resembling a star. For the configuration shown in (**a**) The box plot shows the minimum, average, median, and maximum of PET in each scenario, spatial distribution of PET, (**b**) M1 (**c**) M1-1, (**d**) M1-2, (**e**) M1-3, (**f**) M1-4, (**g**) M1-5, (**h**) M1-6, (**i**) M1-7, (**j**) M1-8.
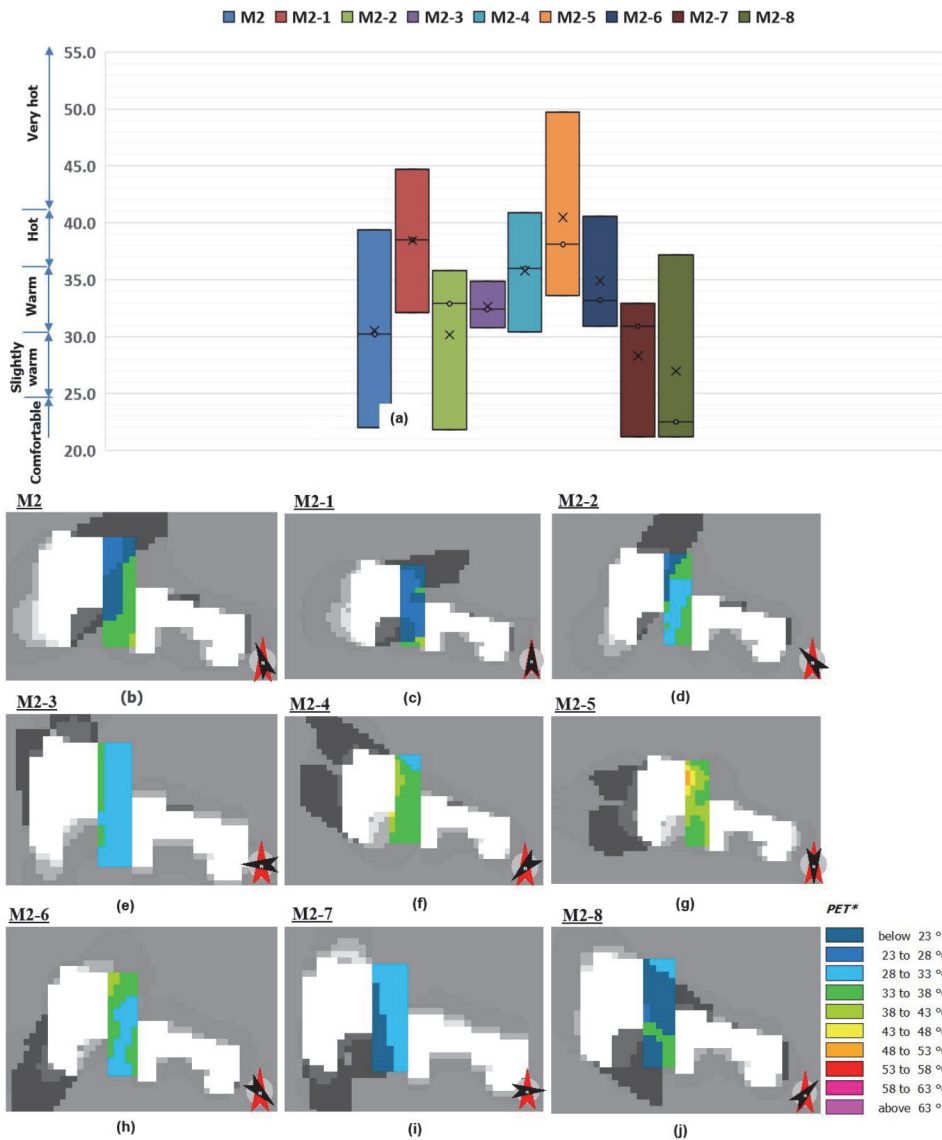
Figure 10. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 2, Scenarios: M2, M2-1 to M2-8. The minimum, maximum, and average scores stand out prominently within the box plots, resembling a star. For the configuration shown in (**a**) The box plot shows the minimum, average, median, and maximum of PET in each scenario, Spatial distribution of PET, (**b**) M2 (**c**) M2-1, (**d**) M2-2, (**e**) M2-3, (**f**) M2-4, (**g**) M2-5, (**h**) M2-6, (**i**) M2-7, (**j**) M2-8.
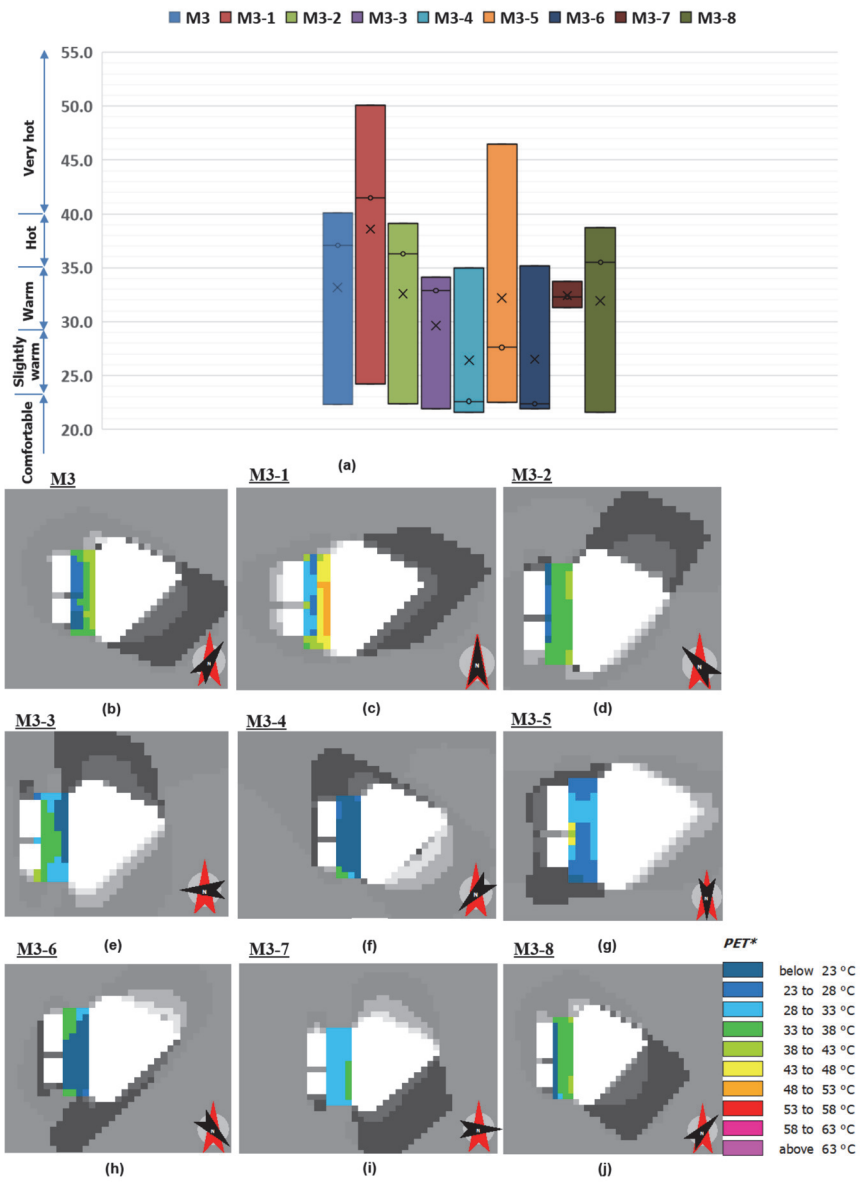
Figure 11. Graphical distribution of PET in different canyon orientations, obtained from CFD simulation, cases study 3, Scenarios: M3, M3-1 to M3-8. The minimum, maximum, and average scores stand out prominently within the box plots, resembling a star. For the configuration shown in (**a**) The box plot shows the minimum, average, median, and maximum of PET in each scenario, spatial distribution of PET, (**b**) M3, (**c**) M3-1, (**d**) M3-2, (**e**) M3-3, (**f**) M3-4, (**g**) M3-5, (**h**) M3-6, (**i**) M3-7, (**j**) M3-8.

For example, in Figure 9, the initial results of PET assessment in different scenarios of case study 1 are demonstrated. Furthermore, Figures 10 and 11 show the results of the PET assessment of case studies 2 and 3, respectively.

It can be seen in the box plot in Figure 9. However, the minimum value of PET of all scenarios is almost the same; scenarios have different values in the average, median, and maximum of PET. Likewise, M1, with the original orientation of 347° (North East-South West), has the lowest minimum value of PET while the other values are even higher than others. Moreover, M1-1 and M1-5, with the orientation of 0 and 180°, have the highest value of the maximum and median of PET. Likewise, a comparison of all different orientations in the urban canopy of case study 1 indicates M1-2 with the orientation of 45° has the lowest median of PET value rather than other scenarios.

In conclusion, the non-uniform spatial distribution of PET does not give us comprehensive data to interpret the results and understand two meanings: (1) in which scenario does the urban canyon offer a better level of thermal comfort in PET at the pedestrian level? (2) in which scenario most of the area is in the comfort zones of PET, such as comfortable (18–23 °C) and slightly warm (23–35 °C).

Thus, there are different measures of PET, such as the minimum, maximum, median, and average values, which can vary depending on the scenario. However, based on the initial evaluation of PET of the simulated scenarios, it is impossible to conclude which orientation offers a higher level of thermal comfort. Therefore, PET's non-uniform spatial distribution can make it challenging to interpret the study results and draw definitive conclusions about which scenario offers a better level of thermal comfort in PET at the pedestrian level.

5.3. Normalized Spatial Distribution of PET

According to Joshi et al., measuring subjective experiences or phenomena can be challenging, as they are often difficult to quantify using conventional measurement techniques [35]. Thus, to obtain the most comprehensive analysis and find the optimum orientation of the urban canopy in the scenarios, we need to consider both the frequency of PET data in each level and the maximum and minimum PET, as well as the average in each urban canyon.

Therefore, evaluation scales can be presented in various graphical ways, with different levels of detail, and no standard gives specifications on the choice of the most suitable configuration; thus, the selection is often a matter of the specifications of the study [43].

In the study, Nazarian et al. used the continuous Outdoor Thermal Comfort Autonomy (OTCA) scale as a metric to measure outdoor thermal comfort. According to the authors, OTCA considers the percentage of time an outdoor space is within the desired thermal comfort range, including periods where the thermal comfort level is below the threshold. It is an extension of Spatial OTCA, defined as the percentage of outdoor space within the desired thermal comfort range at least half of the occupied time (over a year or a prescribed period of use) [44].

Here, we designed a weighted scale to consider the level of thermal comfort in studied areas and rank the data of PET based on the frequency of data in each PET class. As a widespread scale used in different areas such as psychology, sociology, health care, marketing, attitude, preference, customers' quality perceptions or expectations, and subjective well-being in health care, Likert scales have wide applications in different science [45]. In addition, Likert scales are examples of such scales in psychometrics used widely in social science & educational research [46]. Therefore, we applied the Likert scale to weigh the PET classes and ranked the importance of data in each thermal comfort level.

Figure 12 shows the Likert scaling system applied in the study.
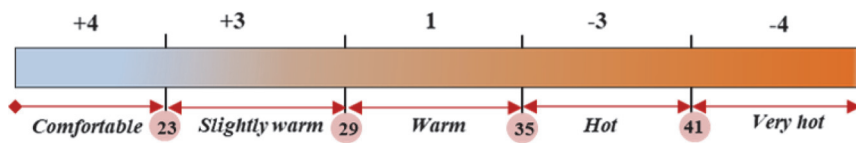
Figure 12. Five-point Likert scale was used in the PET analysis of scenarios.

Each item in the Likert scale usually has an odd number of response categories, to five or seven levels [45], and is named the five-point or seven-point Likert scale. Here, we applied five points Likert scale by considering the comfortability of the area that the Likert scale should measure and assigning the highest indicator equals +4 to the PET class of Comfortable, +3 to Slightly warm, +1 to Warm, and the negative scores to the worst classes of PET cause a high level of discomfort in the urban area, meaning −3 and −4 to Hot and Very hot classes.

Statistical Methods and Exploration of Data

In this section, to better interpret the results, calculating the overall thermal comfort level of each scenario, not only considering the arithmetic and mathematical average of PET but also taking into account the frequency of data in each level of PET, is essential. Therefore, it is an excellent approach to consider the frequency of data in each level of PET to better interpret the results of each scenario's overall thermal comfort level. Accordingly, at first, we stored the results of the PET assessment of each scenario as the experimental data in a matrix and split them into five levels of thermal perception.

To describe a process of analyzing the results of a study on thermal comfort levels in different urban scenarios, we should mention that, at first, the results of the PET assessment of each scenario as the experimental data in were split into five levels of thermal perception based on five-point Likert scale. To better understand the distribution of PET data in each thermal comfort level, we created an experimental matrix and pie charts that show the percentage of PET data distribution of scenarios in different classes of PET, Figures 13–15.
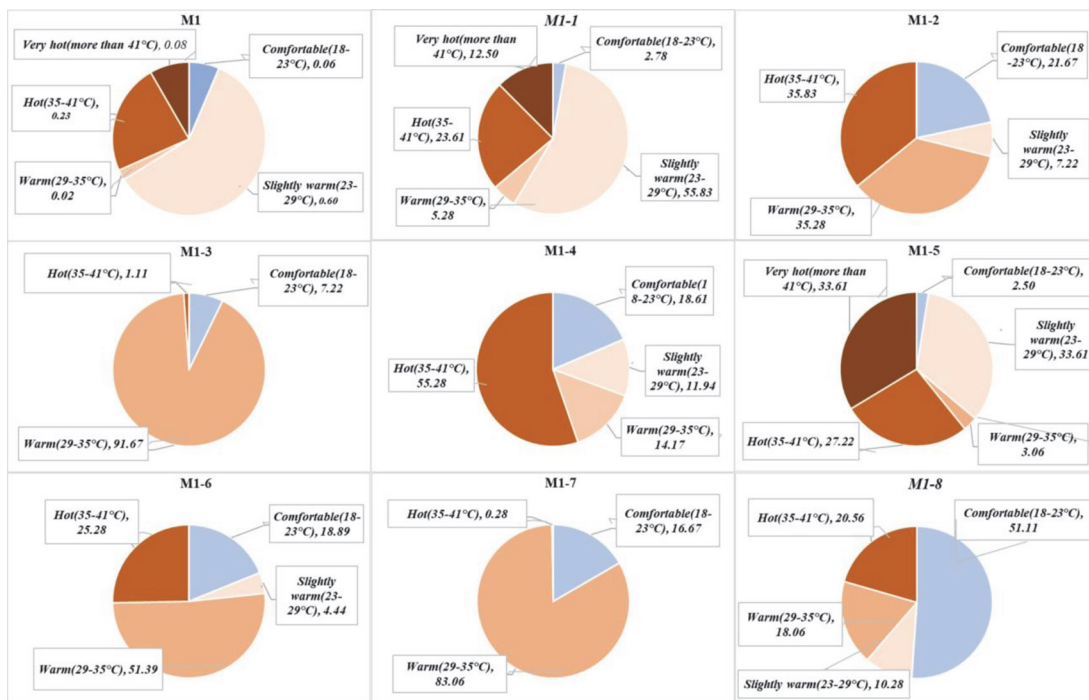
Figure 13. The distribution of PET data in each class of PET/Case study 1 (The total of all segments in every pie chart amount to 100%).
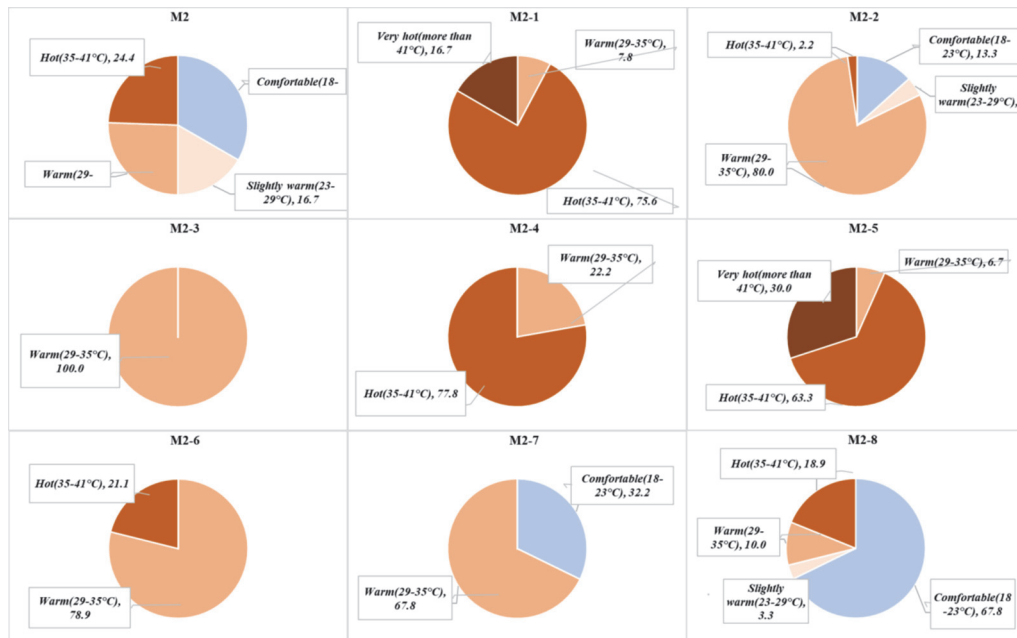
Figure 14. The distribution of PET data in each class of PET/Case study 2 (The total of all segments in every pie chart amount to 100%).
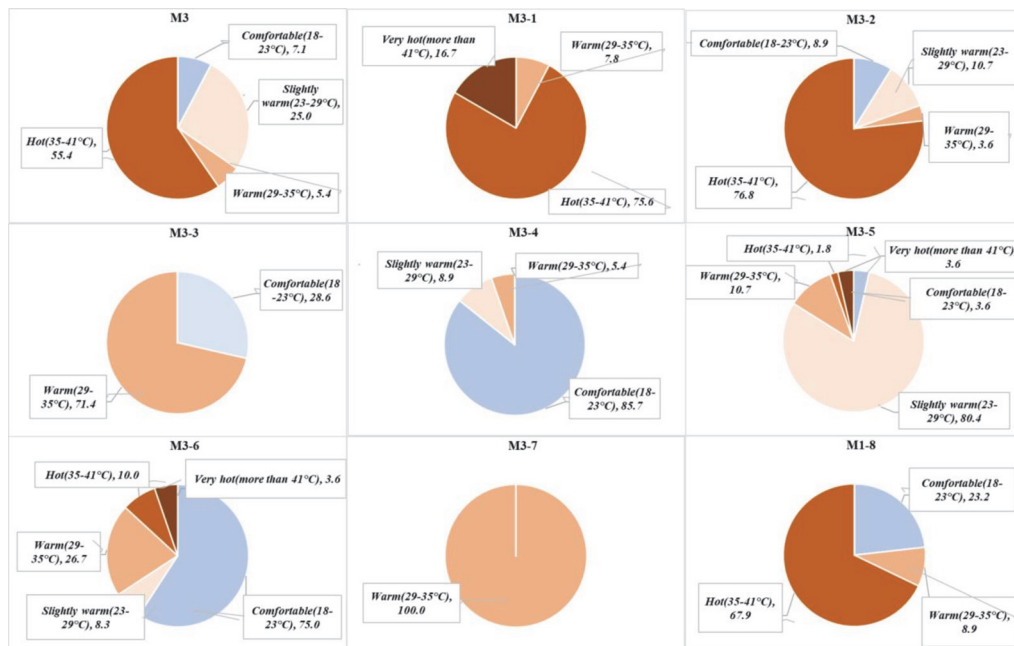
Figure 15. The distribution of PET data in each class of PET/Case study 3 (The total of all segments in every pie chart amount to 100%).

The experimental matrix contains the thermal perception of all scenarios and PET's arithmetic average (mathematical average) in each thermal comfort level, starting from comfortable to very hot. Moreover, the quantity of PET data in each thermal comfort level is counted to understand how much each urban area in the urban canopy offers the considered thermal comfort level. To further refine the analysis, we applied the Likert scaling system to assign a score to each thermal comfort level, which reflects its importance. Combining the arithmetic average of PET in each thermal comfort level with the weight or value of each thermal perception level and the quantity of PET data in each level makes it possible to obtain a more accurate and meaningful measure of each scenario's overall thermal comfort level.

Thus, we defined a statistical method that uses the arithmetic average of PET in each thermal comfort level, starting from comfortable, PET lower than 23° as the comfort zone to greater than 41° to as the zone with the very hot comfort level.

The next step is calculating the statistical average of each PET level by multiplying the arithmetic average of each PET level by the weights of the Likert scale and the respective count of data in each PET level. The following formula shows the weighted mean of PET used in the PET exploration method.

$$\text{Weighted Mean (Wm)} = \sum_{ni=1}^{n} (xi * wi) / \sum_{ni=1}^{n} wi$$

$$Wm = w_1x_1 + w_2x_2 + \ldots + w_nx_n / w_1 + w_2 + \ldots + w_n$$

where: $\sum$ denotes the sum
w is the weights, and x is the value of PET
In cases where the sum of weights is 1,

$$Wm = \sum_{ni=1}^{n} (xi * wi)$$

Figure 16 is the schematic diagram of the model to evaluate PET in each scenario.
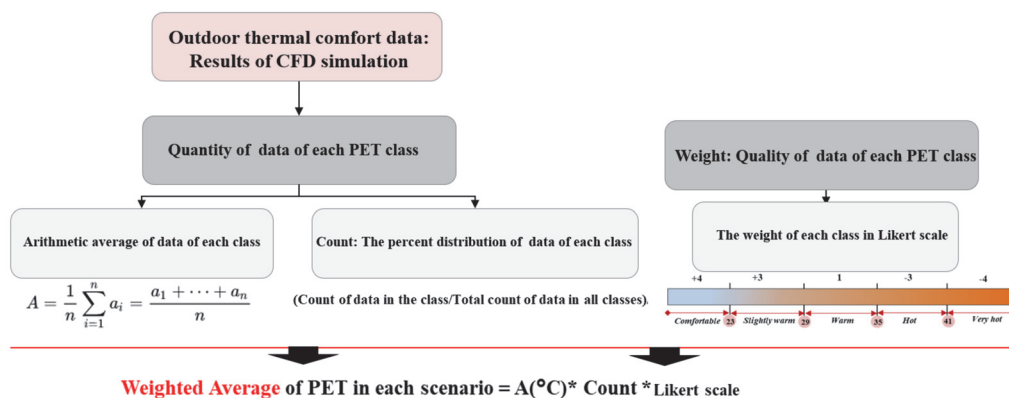
Figure 16. The evaluation model of PET value.

5.4. Analyzing Data Based on the Evaluation Method

The next step was calculating the statistical average of each PET level. For example, to calculate the weighted mean of PET (Wm-PET) for scenario M1, we multiplied the arithmetic average of each PET level (22.5, 24.6, 34.5, 38, and 43.4) by the weights of each response in the Likert scale (4, 3, 1, −3, and −4), and then multiplied each of those values by the count of responses (0.064, 0.597, 0.022, 0.233, and 0.083).

For example, for PET level 1 (Comfortable), the weighted score (Ws) would be:

$$(22.5 * 4 * 0.064) = 5.76$$

We repeated this process for each PET level and then summed up the weighted scores for all PET levels to obtain the overall weighted mean for PET in scenario M1. The calculation for the overall weighted mean is:

$$Wm\_PET, M1 = (Ws\_PET\_level1) + (Ws\_PET\_level2) + \ldots + (Ws\_PET\_level5)$$
$$= (22.5 * 4 * 0.064) + (24.6 * 3 * 0.597) + (34.5 * 1 * 0.022) + (38 * (-3) * 0.233) + (43.4 * (-4) * 0.083)$$
$$= 5.7 + 44.1 + 0.8 - 26.6 - 14.5 = 9.5$$

Therefore, the weighted mean of PET in scenario M1 is 9.5, which represents the overall level of perceived exertion for this scenario, taking into account both the quantity of data in PET classes and the weight of each class as well as the arithmetic average of data of PET in each level. To report the result regarding the optimal orientation for achieving the highest thermal comfort in different scenarios, the summarized weighted average of each scenario is shown in Figure 17.
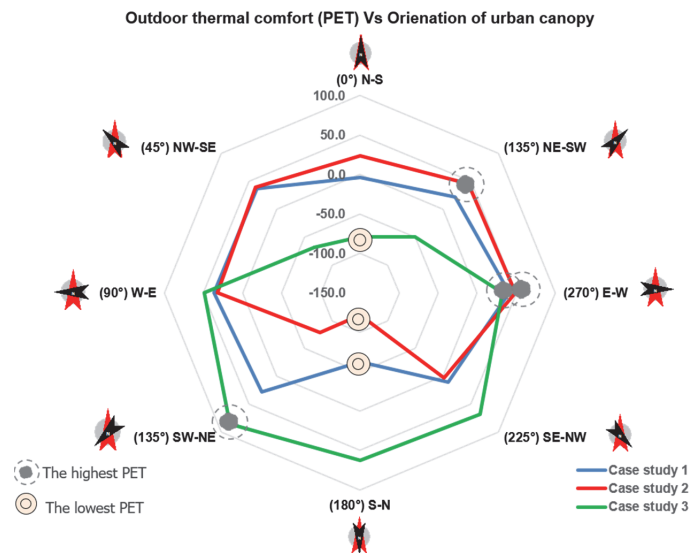
Figure 17. The results of weighted average of PET of each case study in different orientations of the urban canopy.

According to the data presented in Figure 17, the PET index indicates that the highest level of thermal comfort in case study 1 is observed when the urban canyon is oriented in the East-West direction (270°) in scenario M1-7. In case studies 2 and 3, the highest PET levels are seen in scenarios M2-7 and M3-4, respectively, with orientations of 270° (East-West) and 135° (South West-North East) for the extension of the urban canyons. The assessment of outdoor thermal comfort in the studied case studies reveals that the canopy extension in the North-South direction yields the lowest thermal comfort value in all three cases. This is attributed to the sun's high angle during the day, which can create hot and uncomfortable conditions in these orientations. It is important to note that the optimal orientation for achieving outdoor thermal comfort varies depending on the specific characteristics of each case study, as Figure 17 shows.

6. Discussion

Figure 18 presents the study's findings, indicating that the West-East and East-West orientations offer the best outdoor thermal comfort and lowest surface temperatures across all examined case studies and simulation scenarios. However, it is crucial to consider the location of the tallest wall within the urban canyon, as this significantly affects thermal comfort and surface temperatures. Furthermore, the study reveals that the North East-South West orientation provides optimal thermal comfort for case studies 1 and 2 at the pedestrian level. Still, it yields lower thermal comfort for case study 3. Conversely, the South West-North East orientation offers optimum outdoor thermal comfort for case studies 1 and 3. In contrast, the South East-North West orientation only provides good thermal comfort for case study 3. Consequently, the optimal orientation of canopies varies depending on the specific characteristics of each case study, the surrounding environment, and the level of sun exposure and shading on the surfaces. When a tall building is present on one side of a canyon or street, the orientation of the canopy becomes crucial in determining the thermal comfort levels in shaded areas. For instance, in case studies 1 and 2, extending the canopy from North East to South West does not offer sufficient shading due to the orientation of the taller building on the left side of the canyon, potentially resulting in discomfort at the pedestrian level. In contrast, in case study 3, where the taller wall of the canopy is located on the right side, extending the canopy from South West to North East provides better shading and enhances thermal comfort. Therefore, it is vital to consider the specific morphological features of buildings and the surrounding area when designing urban canopies.
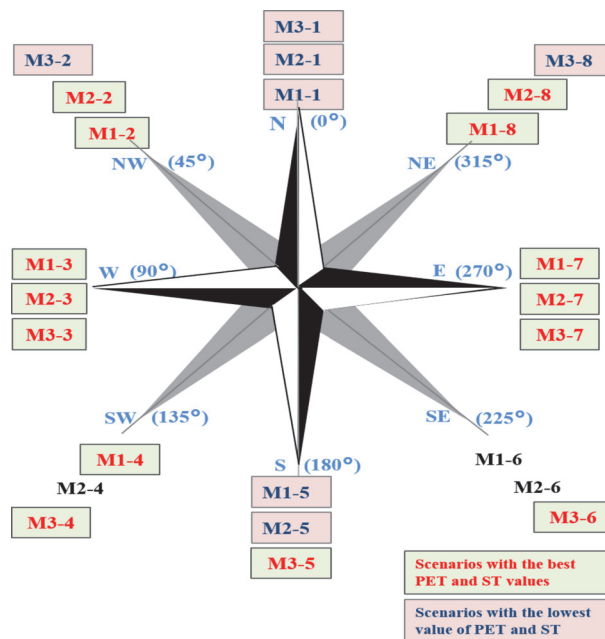
Figure 18. Final results regarding optimum thermal comfort and surface temperature of each scenario in different urban environment orientations.

Step 4: Application of the Study

The study underscores the significant influence of urban canyon orientation on thermal comfort and surface temperature within urban settings. By considering the orientation of urban canyons, urban planners and designers can optimize urban environment design to prioritize the well-being of residents and visitors. The findings offer insights into de-

termining the optimal extension of the canopy layer to maximize thermal comfort, thus providing more comfortable areas for pedestrians. Overall, these findings are valuable for urban planners and designers seeking to develop more sustainable and livable urban environments that prioritize the well-being of individuals. Moreover, considering the specific characteristics of the urban canyon enables planners to implement targeted strategies, including shading, ventilation, and other measures, to enhance outdoor thermal comfort and mitigate the effects of phenomena including the UHI effect.

7. Conclusions

This investigation focused on the impact of urban canopy orientation on outdoor thermal comfort and surface temperature in areas with significant UHI effects. Numerical analysis and CFD simulations were employed to evaluate these factors in urban environments. The orientation of urban environments, buildings, and streets within urban canyons is crucial in determining thermal comfort levels. The level of thermal comfort and surface temperature is influenced by factors such as building mass, canopy wall height, and arrangement. The study's findings underscore the significance of the canopy orientation in determining thermal comfort levels in shaded areas when there is a tall building present on one side of a canyon or street. The research highlights that the optimal orientation of canopies depends on the unique characteristics of each case study, the surrounding environment, and the degree of sun exposure and shading on surfaces. It emphasizes the importance of considering the buildings and elements surrounding them when designing and positioning canopies to achieve optimal thermal comfort in shaded areas affected by tall buildings. By incorporating these factors into the design and planning processes, urban planners and designers can create comfortable and sustainable living spaces. The assessment of outdoor thermal comfort within the urban canopy provides valuable insights for optimizing the extension of the canopy layer, enhancing thermal comfort, controlling the UHI effect, and improving residents' quality of life. The findings have broad applicability and can assist in early-stage city design, redevelopment, and renovation projects. Prioritizing outdoor thermal comfort in urban design offers social, environmental, and economic advantages.

While the study's findings are specific to Tallinn, the methodology and approach we employed for assessing thermal comfort and surface temperatures have broader applicability and can be implemented in various locations and climates. This can be achieved by considering the unique characteristics of the built environment and meteorological conditions of each area. However, further research is needed to formulate effective policies and planning codes that can adequately tackle the variations in Urban Heat Islands, manage excessively hot urban regions, and enhance thermal comfort. It is essential to base mitigation strategies for UHI on the specific thermal attributes of the particular urban area rather than relying solely on distinctions between urban and rural zones. Moreover, when adapting thermal comfort indices to diverse climates and countries, cultural differences should also be taken into consideration.

Limitations of the Study

It is crucial to acknowledge and consider the limitations of any study to ensure that the results are not overgeneralized. In this study, several limitations should be taken into account. To begin with, the geometric models used for the CFD simulations were simplified by omitting detailed information regarding the systems, materials, and albedo of the models. This simplification has the potential to impact the accuracy of the obtained results. The thermal comfort assessment in the study was also limited to a specific subset of the population, namely males in light clothing and seated position. This limitation calls for future studies to include a more diverse population, especially vulnerable groups such as the elderly and children, and gather data on their physical characteristics, clothing choices, and activity levels to provide a more comprehensive analysis of the impact of thermal comfort on different individuals.

Furthermore, it is important to note that thermal comfort is just one aspect related to the UHI effect in urban areas. Future studies should aim to provide a more holistic understanding of UHI by considering other strategies and elements of the urban environment and their configurations. Lastly, it should be acknowledged that the optimal orientation for achieving thermal comfort in an urban canyon can vary depending on various factors, such as the local climate, building materials used, and the layout of the urban canyon itself. Considering all these factors will contribute to more accurate and reliable results.

References

1. Akbari, H.; Cartalis, C.; Kolokotsa, D.; Muscio, A.; Pisello, A.L.; Rossi, F.; Santamouris, M.; Synnefa, A.; Wong, N.H.; Zinzi, M. Local climate change and urban heat island mitigation techniques—The state of the art. *J. Civ. Eng. Manag.* **2016**, *22*, 1–16. [CrossRef]
2. Erell, E.; Pearlmutter, D.; Williamson, T. *Urban Microclimate: Designing the Spaces between Buildings*; Urban Microclimate; Routledge: London, UK, 2012. [CrossRef]
3. Onishi, A.; Cao, X.; Ito, T.; Shi, F.; Imura, H. Evaluating the potential for urban heat-island mitigation by greening parking lots. *Urban For. Urban Green.* **2010**, *9*, 323–332. [CrossRef]
4. Aflaki, A.; Mirnezhad, M.; Ghaffarianhoseini, A.; Ghaffarianhoseini, A.; Omrany, H.; Wang, Z.-H.; Akbari, H. Urban heat island mitigation strategies: A state-of-the-art review on Kuala Lumpur, Singapore and Hong Kong. *Cities* **2017**, *62*, 131–145. [CrossRef]
5. Mirzaei, P.A.; Haghighat, F. Approaches to study urban heat island–abilities and limitations. *Build. Environ.* **2010**, *45*, 2192–2201. [CrossRef]
6. Sagris, V.; Sepp, M. Landsat-8 tirs data for assessing urban heat island effect and its impact on human health. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2385–2389. [CrossRef]
7. Aleksandrowicz, O.; Vuckovic, M.; Kiesel, K.; Mahdavi, A. Current trends in urban heat island mitigation research: Observations based on a comprehensive research repository. *Urban Clim.* **2017**, *21*, 1–26. [CrossRef]
8. Vásquez-Álvarez, P.E.; Flores-Vázquez, C.; Cobos-Torres, J.-C.; Cobos-Mora, S.L. Urban Heat Island Mitigation through Planned Simulation. *Sustainability* **2022**, *14*, 8612. [CrossRef]
9. Rizwan, A.M.; Dennis, L.Y.; Chunho, L. A review on the generation, determination and mitigation of urban heat island. *J. Environ. Sci.* **2008**, *20*, 120–128. [CrossRef] [PubMed]
10. Giridharan, R.; Emmanuel, R. The impact of urban compactness, comfort strategies and energy consumption on tropical urban heat island intensity: A review. *Sustain. Cities Soc.* **2018**, *40*, 677–687. [CrossRef]
11. Shishegar, N. Street Design and Urban Microclimate: Analyzing the Effects of Street Geometry and Orientation on Airflow and Solar Access in Urban Canyons. *J. Clean Energy Technol.* **2013**, *1*, 52–56. [CrossRef]
12. Eslamirad; Nasim; De Luca, F.; Lylykangas, K.S. The role of building morphology on pedestrian level comfort in Northern climate. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 2042, p. 012053. [CrossRef]
13. Eslamirad; Nasim; De Luca, F.; Lylykangas, K.S.; Yahia, S.B. Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment. *Front. Archit. Res.* **2023**, *12*, 541–555. [CrossRef]
14. Eslamirad; Nasim; Kolbadinejad, S.M.; Mahdavinejad, M.; Mehranrad, M. Thermal comfort prediction by applying supervised machine learning in green sidewalks of Tehran. *Smart Sustain. Built Environ.* **2020**, *9*, 361–374. [CrossRef]
15. Rosenzweig, C.; Solecki, W.D.; Romero-Lankao, P.; Mehrotra, S.; Dhakal, S.; Ibrahim, S.A. (Eds.) *Climate Change and Cities: Second Assessment Report of the Urban Climate Change Research Network*; Cambridge University Press: Cambridge, UK, 2018. [CrossRef]
16. Martilli, A.; Krayenhoff, E.S.; Nazarian, N. Is the Urban Heat Island intensity relevant for heat mitigation studies? *Urban Clim.* **2020**, *31*, 100541. [CrossRef]
17. Kim, S.W.; Brown, R.D. Urban heat island (UHI) variations within a city boundary: A systematic literature review. *Renew. Sustain. Energy Rev.* **2021**, *148*, 111256. [CrossRef]
18. Wang, X.; Li, H.; Sodoudi, S. The effectiveness of cool and green roofs in mitigating urban heat island and improving human thermal comfort. *Build. Environ.* **2022**, *217*, 109082. [CrossRef]

19. Montávez; Juan, P.; Rodríguez, A.; Jiménez, J.I. A study of the urban heat island of Granada. *Int. J. Climatol. A J. R. Meteorol. Soc.* **2000**, *20*, 899–911. [CrossRef]
20. Wang, Y.; Berardi, U.; Akbari, H. Comparing the effects of urban heat island mitigation strategies for Toronto, Canada. *Energy Build.* **2016**, *114*, 2–19. [CrossRef]
21. Dionysia, K.; Lilli, K.; Gobakis, K.; Mavrigiannaki, A.; Haddad, S.; Garshasbi, S.; Mohajer, H.R.H.; Paolini, R.; Vasilakopoulou, K.; Bartesaghi, C.; et al. Analyzing the Impact of Urban Planning and Building Typologies in Urban Heat Island Mitigation. *Buildings* **2022**, *12*, 537. [CrossRef]
22. Chao, X.; Chen, G.; Huang, Q.; Su, M.; Rong, Q.; Yue, W.; Haase, D. Can improving the spatial equity of urban green space mitigate the effect of urban heat islands? An empirical study. *Sci. Total Environ.* **2022**, *841*, 156687. [CrossRef]
23. Cortes, A.; Rejuso, A.J.; Santos, J.A.; Blanco, A. Evaluating mitigation strategies for urban heat island in mandaue city using envi-met. *J. Urban Manag.* **2022**, *11*, 97–106. [CrossRef]
24. Rajan, E.H.S.; Amirtham, L.R. Urban heat island intensity and evaluation of outdoor thermal comfort in Chennai, India. *Env. Dev Sustain.* **2021**, *23*, 16304–16324. [CrossRef]
25. Darbani, S.; Elham; Parapari, D.M.; Boland, J.; Sharifi, E. Impacts of urban form and urban heat island on the outdoor thermal comfort: A pilot study on Mashhad. *Int. J. Biometeorol. Vol. Int. J. Biometeorol.* **2021**, *65*, 1101–1117. [CrossRef] [PubMed]
26. Farhadi, H.; Faizi, M.; Sanaieian, H. Mitigating the urban heat island in a residential area in Tehran: Investigating the role of vegetation, materials, and orientation of buildings. *Sustain. Cities Soc.* **2019**, *46*, 101448. [CrossRef]
27. Arnfield, A. Street design and urban canyon solar access. *Energy Build.* **1990**, *14*, 117–131. [CrossRef]
28. van Esch, M.; Looman, R.; de Bruin-Hordijk, G. The effects of urban and building design parameters on solar access to the urban canyon and the potential for direct passive solar heating strategies. *Energy Build.* **2012**, *47*, 189–200. [CrossRef]
29. Steeneveld, G.J.; Koopmans, S.; Heusinkveld, B.G.; van Hove, L.W.A.; Holtslag, A.A.M. Quantifying urban heat island effects and human comfort for cities of variable size and urban morphology in The Netherlands. *J. Geophys. Res.* **2011**, *116*, D20129. [CrossRef]
30. Yahia, M.; Johansson, E. Influence of urban planning regulations on the microclimate in a hot dry climate: The example of Damascus, Syria. *J. Hous. Built Environ.* **2012**, *28*, 51–65. [CrossRef]
31. Battista, G.; de Lieto Vollaro, E.; Ocłoń, P.; de Lieto Vollaro, R. Effects of urban heat island mitigation strategies in an urban square: A numerical modelling and experimental investigation. *Energy Build.* **2023**, *282*, 112809. [CrossRef]
32. Ali-Toudert, F.; Mayer, H. Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate. *Build. Environ.* **2006**, *41*, 94–108. [CrossRef]
33. Eslamirad, N.; Sepúlveda, A.; De Luca, F.; Lylykangas, K.S. Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus. *Energies* **2022**, *15*, 1577. [CrossRef]
34. Eslamirad; Nasim; De Luca, F.; Lylykangas, K.S.; Yahia, S.B.; Rasoulinezhad, M. Geoprocess of geospatial urban data in Tallinn, Estonia. *Data Brief* **2023**, *48*, 109172. [CrossRef]
35. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [CrossRef]
36. Eslamirad, N. Geoprocess of Geospatial Urban Data in Tallinn, Estonia. Mendely Data, V3, 4 02 2023. Available online: https://data.mendeley.com/drafts/2bm7kdf8gb (accessed on 11 May 2023).
37. General Data of Tallinn. Available online: https://www.tallinn.ee/en/statistika/general-data-tallinn (accessed on 5 April 2023).
38. Jamei, E.; Rajagopalan, P.; Seyedmahmoudian, M.; Jamei, Y. Review on the impact of urban geometry and pedestrian level greening on outdoor thermal comfort. *Renew. Sustain. Energy Rev.* **2016**, *54*, 1002–1017. [CrossRef]
39. Cohen, P.; Potchter, O.; Matzarakis, A. Human thermal perception of Coastal Mediterranean outdoor urban environments. *Appl. Geogr.* **2013**, *37*, 1–10. [CrossRef]
40. Deb, C.; Ramachandraiah, A. The significance of Physiological Equivalent Temperature (PET) in outdoor thermal comfort studies. *Int. J. Eng. Sci. Technol.* **2010**, *2*, 2825–2828.
41. Moazzam, M.F.U.; Doh, Y.H.; Lee, B.G. Impact of urbanization on land surface temperature and surface urban heat island using optical remote sensing data: A case study of jeju island, republic of korea. *Build. Environ.* **2022**, *22*, 109368. [CrossRef]
42. Neto, A.F.; Bianchi, I.; Wurtz, F.; Delinchant, B. *Thermal Comfort Assessment*; ELECON, Electricity Consumption Analysis and Energy Efficiency: Gujarat, India, 2016. [CrossRef]
43. Giampaoletti, M.; Pistore, L.; Zapata-Lancaster, G.; Goycoolea, J.P.F.; Janakieska, M.M.; Gramatikov, K.; Kocaman, E.; Kuru, M.; Andreucci, M.; Calis, G.; et al. *RESTORE Regenerative Technologies for the Indoor Environment—Inspirational Guidelines for Practitioners*; ELECON, Electricity Consumption Analysis and Energy Efficiency: Gujarat, India, 2020.
44. Nazarian, N.; Acero, J.A.; Norford, L. Outdoor thermal comfort autonomy: Performance metrics for climate-conscious urban design. *Build. Environ.* **2019**, *155*, 145–160. [CrossRef]
45. Chakrabartty, S. Scoring and Analysis of Likert Scale: Few Approaches. *J. Knowl. Manag. Inf. Technol.* **2014**, *1*, 31–44.
46. Joshi, A.; Kale, S.; Chandel, S.; Pal, D. Likert Scale: Explored and Explained. *Br. J. Appl. Sci. Technol.* **2015**, *7*, 396–403. [CrossRef]

Paper V

Eslamirad, N., Sepúlveda, A., De Luca, F. & Sakari Lylykangas, K. (2022). Evaluating outdoor thermal comfort using a mixed-method to improve the environmental quality of a university campus. *Energies, 15*, 1577. https://doi.org/10.3390/en15041577

Article

Evaluating Outdoor Thermal Comfort Using a Mixed-Method to Improve the Environmental Quality of a University Campus

Nasim Eslamirad [1,*]**, Abel Sepúlveda** [2]**, Francesco De Luca** [2] **and Kimmo Sakari Lylykangas** [2]

1   FinEst Centre for Smart Cities, Tallinn University of Technology, 19086 Tallinn, Estonia
2   Department of Civil Engineering and Architecture, Academy of Architecture and Urban Studies,
    Tallinn University of Technology, 19086 Tallinn, Estonia; absepu@taltech.ee (A.S.);
    francesco.deluca@taltech.ee (F.D.L.); kimmo.lylykangas@taltech.ee (K.S.L.)
*   Correspondence: nasim.eslamirad@taltech.ee; Tel.: +372-58-188-087

Abstract: Thermal comfort in cities is increasingly becoming a concern and comfortable places can be highly valuable for a variety of activities. Our investigation aims to explore how to improve the quality of cities by considering the relationship between microclimatic conditions, thermal sensation, and human preferences. The case study conducted in the open areas of Tallinn University of Technology (TalTech) campus, which is quite populated by visitors, staff, and students. We used a mixed-methods approach to assess outdoor thermal comfort, based on qualitative and quantitative findings of the relationships between the measured weather conditions and the results of thermal comfort assessment through the PET index and subjectively perceived thermal sensation. In the qualitative part, data was collected through semi-structured interviews. The main conclusions from the interviews were used to design a survey and the samples. Based on the results, it was possible to identify places that offer different levels of thermal comfort. Thus, the study helps to improve thermal comfort at the campus, which is one of the goals of the Green Transition project to make the campus fully sustainable. Moreover, the methodology is applicable in different urban areas to improve urban health and sustainability and create resilient urban environments.

Keywords: survey; semi structured interview; outdoor thermal comfort optimization; urban simulation

1. Introduction

In the context of urban planning, it is interesting how thermally comfortable urban environments influence people's behavior, use of outdoor spaces, and the quality of life in cities [1]. Sustainable cities are designed with their environmental impact in mind [2]. Moreover, the number of users of outdoor spaces would determine the vibrancy of the place, the local socio-economy, and sustainability of the city [3]. In addition, outdoor microclimate or related outdoor thermal comfort is an important factor influencing the quality of urban spaces [3]. As weather and location invite people to be active in urban spaces where many outdoor activities take place, activity levels in cities are influenced by a range of urban conditions [4]. However, in contrast, uncomfortable and low-quality outdoor spaces cause people to rush home [5]. According to Gehl, since voluntary activities are more likely to take place when outdoor space is of good quality, the number of social activities that depend on the presence of others in public space also tends to increase significantly [4]. In addition, pedestrians are directly exposed to the immediate environment in the form of variations in sun/shade and shade, wind speed [1], and other weather features during recreation and leisure in the outdoor spaces [1,6]. Therefore, the quality of outdoor spaces affects the quality of life of residents [3], and microclimatic conditions have significant impact on urban development [7]. Accordingly, one of the main objectives in the study of thermal comfort is to find out the relationship between the thermal environment and the thermal perception of the inhabitants [8].

Urban planners are increasingly concerned with the impact of climate on urban planning [9]. They are constantly striving to maintain and improve the quality of life of city dwellers by creating comfortable and pleasant environment [10]. Indeed, outdoor quality of life is one of the essential parameter to assess the quality of the urban microclimate [11] and can be improved through various design concepts and solutions to make public spaces more livable and pleasant [12]. Similarly, studies on thermal comfort at an urban scale create a close link between urban and landscape planners to pay more attention to pedestrians (biometeorology) and climate (climatology) [13]. In addition, outdoor thermal comfort has received great attention since the start of the new millennium [14]. Nevertheless, due to the great complexity of the outdoor environment in terms of temporal and spatial variability and the variety of activities that people engage in, there have been very few attempts to understand outdoor comfort conditions [12]. Therefore, there is a lack of studies trying to find out the relationship between people's thermal perceptions and the ranges of thermal comfort in urban areas and the cause-effect elements.

Thanks to advances in urban climatology and biometeorological techniques, some detailed microclimatic analyses and assessments of thermal comfort have been conducted in recent decades [1]. A criterion for qualitative assessment would also be helpful in creating comfortable urban spaces [6]. In addition, there is a lack of understanding of urban features that provide a comfortable urban space to urban dwellers, considering both qualitative and quantitative features. To further explore these subjectivities, this paper aims to explore (i) the objective and subjective elements in defining outdoor thermal comfort, (ii) the use of both qualitative and quantitative research methods in assessing outdoor thermal comfort and the level of comfortability in each area, and (iii) how and to what extent all personal factors, weather conditions, and urban features influence the sense of outdoor thermal comfort.

1.1. Background of Outdoor Thermal Comfort Evaluation

Since 1920, studies on the human thermal environment and various thermal indices based on air temperature and relative humidity have been developed. A classical concept to describe thermal perception was developed by Fanger who described "thermal comfort" as "man's satisfaction with his thermal environment" [5,15]. In 1960, Fanger started to study thermal comfort [16] and defined this concept for indoor spaces and also developed a physiological index Predicted Mean Vote (PMV) to describe "thermal comfort" quantitatively [5], while his first results were published in 1967 [16]. Fanger's equation was the basis for ISO 7730 [17] and ASHRAE 55 [18]. The main thermal comfort standard [17] was based on PMV and PPD that considers the criteria like ISO 8996 (metabolic rate) and ISO 9920 (clothing) and describes the PMV and PPD indices and specifies acceptable conditions for thermal comfort [19].

Numerous thermal comfort biometeorological indices such as standard effective temperature (SET), PMV, physiological equivalent temperature (PET), universal thermal climate index (UTCI), perceived temperature (PT), and outdoor standard effective temperature (OUT_SET) have been developed to describe human thermal comfort by establishing a link between local microclimatic conditions and human thermal sensation [1,5]. PET is based on the Munich energy balance model for individuals (MEMI) as an index with temperature dimension expressed in degrees Celsius (°C), which makes its interpretation understandable even for people without much knowledge of meteorology [1]. Table 1 shows PET in the different thermal perception by human. PET is defined using the concept of equivalent temperature: it is the indoor air temperature of an isothermal environment that produces the same core and skin temperature as the actual complex outdoor conditions. In this typical room the ambient conditions are homogeneous, the air is calm (<0.1 m/s), and the vapor pressure is 1200 Pa (50% relative humidity at 20 °C). Thus, PET allows a layman to compare the integrated effects of complex outdoor thermal conditions with his own experience indoors [13]. Another study, Lin used PET as the primary thermal index [20]. Other studies used PET to determine neutral temperature in different climatic regions [21–23].

PMV and SET are based on the physically based heat balance and heat transfer model [8]. OUT_SET and PET are both based on steady-state energy balance models of the human body. Their application is limited to situations where people spend long periods of time outdoors [24].

Table 1. Ranges of the thermal index of PET for different grades of thermal perception by human [13].

PET (°C)	Thermal Perception (Internal Heat Production: 80 W, Heat Transfer Resistance of the Clothing: 0.9)	Grade of Physiological Stress
	Very Cold	Extreme cold stress
4		
	Cold	Strong cold stress
8		
	Cool	Moderate cold stress
13		
	Slightly cool	Slight cold stress
18		
	Comfortable	No thermal stress
23		
	Slightly warm	Slightly heat stress
29		
	Warm	Moderate heat stress
35		
	Hot	Strong heat stress
41		
	Very hot	Extreme heat stress

The UTCI metric was developed by 45 scientists from 23 countries to standardize applications in the most important areas of human biometeorology [14]. The UTCI metric is expressed as the equivalent ambient temperature of a reference environment that produces the same physiological response in a reference subject as the actual environment [25]. The UTCI has been called for by various disciplines as a physiological response-based assessment index that is valid for a wide range of outdoor climatic conditions, including weather extremes.

Due to the different indices of thermal comfort, assessment methods, and procedures used in the different studies, the results may not be comparable. Table 1 shows PET in the different thermal perception by human. Another study, Lin [20] used PET as the primary thermal index, and Ng et al. [21], Kántor et al. [22], and Kruger et al. [23] used PET to determine neutral temperature in the climatic regions they studied. In addition, the UTCI was developed by 45 scientists from 23 countries [14] to standardize applications in the most important areas of human biometeorology.

1.2. Literature Review

There has been significant effort to investigate the influence factors of outdoor thermal comfort and assessment methods. Unlike indoor environments, urban microclimates, are dynamic and the changing sunlight, wind and shading from trees make the environment volatile [9]. In addition to the climatic aspects of thermal comfort, a number of physical and social factors come into play that influence people's perception of urban space when they are outdoors [1]. In 1971, Gehl studied the influence of microclimate on outdoor activities and showed that sunny or shady local conditions significantly influence people's desire to either stay or leave [4].

Nikolopoulou et al. were pioneers addressing human behavior, research frameworks and analytical procedures [1]. They asked people, within the context of recreational areas in Cambridge, UK, about their subjective thermal sensation, which were given on a five-point scale from too cold to too hot. They also considered environmental characteristics (air temperature, sunlight, etc.) and individual characteristics (age, gender, clothing, etc.) [1]. Nikolopoulou et al. showed that the thermal environment is indeed of prime importance

for users of urban spaces, but psychological adaptation, i.e., available choices, environmental stimulation, thermal history, memory effect, and expectations are also of great importance [26].

The environmental stimulus (i.e., local microclimatic conditions) is the most important factor influencing thermal sensation and human assessments of comfort [1]. These assessments are dynamic in the sense that adaptation to a thermal environmental condition is progressive and that thermal sensation is primarily influenced by experience, and subjective, which implies that the assessment of a thermal comfort condition does not always correspond to objective climatic or biometeorological conditions [1]. Kenz et al. reported significant influences of weather parameters and personal factors on participants' perceived and subjective evaluations of outdoor urban areas [27].

Lenzholzer et al. used qualitative methods to design thermally comfortable urban space, linking thermal and spatial information from to people's perceptions [5]. Zacharias et al. investigated seven corporate plazas and public spaces in the city center of a North American city (Montreal) to determine the relationship between the local microclimate and the level of use, measured as the degree of presence of people and activities (sitting, standing, and smoking) [1]. The other study integrates theoretical findings on outdoor thermal comfort, weather perception, and emotional experience related to travel behavior, and collected verbal responses. Mechanisms of thermal and mechanical comfort lead to more pleasant emotions during travel [28]. Moreover, Auliciems described the physiological responses of the human body to thermal conditions as "thermal sensation" and argued that the common use of the term "thermal comfort" in the literature is inadequate to describe unpleasant thermal stimuli to which humans are frequently exposed. He proposed a neutral and comprehensive term to describe physiological and psychological influences together: "thermal perception" [5].

In other study, long-term perception was introduced through the terms 'short-term' and 'memory' [29]. The aim of this study was to investigate correlations between the influences of the respective urban space on thermal perception. The method qualitative techniques (e.g., interviews) and quantitative studies (e.g., micrometeorological measurements of physical parameters and numerical modeling) to obtain a balanced view of the objective and subjective aspects of the thermal perception. Thus, the spatial and material properties of the environment have an influence on thermal perception [5]. The study examined people's perceptions using a structured interview that included questions about their perception of the place in terms of microclimate and perception of the place. Klemm et al. asked participants about thermal perception and spatial perception on a five-point Likert scale from "very cold" to "very hot" to obtain specific information about the descriptors of "ambience" in relation to thermal perception for Dutch urban squares. In this study, people's thermal perceptions correlate with the measurements and simulation results [30].

Another study determined people's individual thermal sensations in spatial zones located in Cairo, Egypt, by conducting interviews in parks during summer and winter. The survey results were presented through descriptive and correlative statistics between PET and the voices on thermal sensations. The results showed a close relationship between the thermal voices and the influence of the sky visibility factor, wind speed, and albedo, based on the microclimatic influence of landscape elements such as the presence of vegetation and fountains [31].

In other seminal study, Ahmad assessed outdoor comfort based on field research through a survey, in the humid tropics. The study discovered factors affecting comfort outdoors for Dhaka and a comfort range of environmental parameters [6].

From an urban design perspective, Zacharias et al. tried to find investigated the relationship between climatic parameters and the human thermal sensation and what significance these parameters have for people's behavior. In this study, a quantitative relationship between microclimate and the use of urban open spaces was described. The results show that, among the microclimatic factors, temperature is the most important variable influencing human well-being [32].

Nikolopoulou et al. showed that microclimatic parameters cannot fully explain the large differences between objective and subjective comfort ratings, although they strongly influence the perception of warmth [33].

In another study based on large-scale interviews in within the context of Cambridge City Centre, Nikolopoulou showed that the study of thermal comfort in an urban context does not require a quantitative approach because outdoor comfort conditions are not adequately described. The results of this study show that the design of public spaces requires an understanding of dynamic human parameters and that psychological adaptations such as available choice, environmental stimulation, thermal history, memory effect, and expectations have a greater influence than the thermal environment [26].

In the study, which was conducted in the Nordic city, of Gothenburg, Sweden, Kenz et al. investigated the psychological mechanisms in the evaluation of outdoor places and weather, as well as the significant influences of weather parameters and personal factors such as environmental attitudes and age on the perception and evaluation of urban places [27]. A study in another Nordic city, Tallinn, Estonia investigated optimal commercial building cluster layouts to improve outdoor thermal comfort, assessed using the UTCI index, and indoor thermal comfort and cooling energy use needed to maintain comfort during the warm season [34].

Eliasson et al. used a multidisciplinary and interdisciplinary approach, by focusing on assessing the impact of weather and microclimate on people in outdoor urban environments. The result supports the concept of climate-sensitive planning in future urban design and planning projects [29]. The other study looked at thermal comfort in both dynamic and subjective terms. The study was a four-level assessment framework: physical, physiological, psychological, and social/behavioral. The results led to a general framework for assessing outdoor thermal comfort based on behavioral aspects and outdoor thermal comfort planning [1].

Similarly, Lai et al. assessed outdoor thermal comfort and space utilization in a residential community in Wuhan, central China. The method of the study is both observation of the microclimate through interviews with the residents and observation of the amount and type of activities of the residents [3]. The TSV was used to assess thermal comfort in this study, providing valuable information for the design of outdoor areas in residential communities [3].

1.3. Originality and Aims of the Study

Thanks to advances in urban climatology and biometeorological techniques, some detailed microclimatic analyses and thermal comfort assessments have been conducted in the last decade [1]. Therefore, qualitative methods that provide an explicit combination of thermal and spatial information have been developed to link thermal and spatial information about people's perceptions. Based on these findings, new insights can be gained for the design of thermally comfortable urban spaces [5]. There is a lack of understanding of personal parameters, human preferences, microclimatic conditions, as well as urban features that provide comfortable urban space to urban dwellers. The novelty of this study is applying mixed approaches to assess and improve the quality of urban spaces. Thus, the aims of this investigation are as follows:

- To study the role of objective and subjective elements in defining comfort in an urban area;
- To apply qualitative and quantitative research methods in assessing outdoor thermal comfort;
- To explore how and what extent personal factors, weather conditions, environmental conditions, and phisical features of the area can affect comfort sensation.

2. Methodology

As different models and tools with varying degrees of complexity have been developed to address the problem of thermal comfort in an urban context, providing a general, com-

prehensive, insight requires analysis at different levels of complexity and overwhelming engagement with them.

Usually campuses are complex organisms characterised by different activities of people living and working there [35]. We used a mixed-method methodology in this investigation. On one hand, we studied people's thermal preferences during summer in the Tallinn University of Technology (Taltech), Estonia through semi-structured interviews and surveys. On the other hand, we used the well-known ENVI-met CFD software [36,37] to assess outdoor thermal comfort in terms of the PET metric, which is one of the most robust outdoor thermal comfort indices nowadays [9,13].

Figure 1 shows the general framework followed in this study. The qualitative research approach used in the study consists of semi-structured interviews with people who use the campus during the summer season. The quantitative part involves conducting an online survey designed from interviews analyses, as well as observing data on campus, collecting measured climate data on campus, and finally modeling the urban area and conducting CFD simulations. The results of the online survey were used to set simulation case studies: people's personal characteristics, activity levels, and thermal preferences. The measured microclimatic data of the case study area and the city climate data were considered to provide the data for thermal comfort assessment for each sample of the study on each day based on the assumed scenario. The final step in the study is the conclusion and visualization so that all the results and achievements of the study can be used to improve the quality of urban space and provide people with a better-quality urban environment.
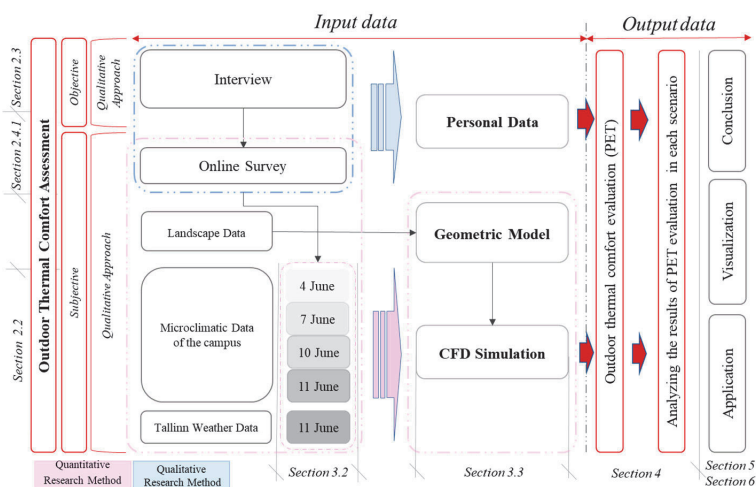
Figure 1. The framework proposed in this investigation.

2.1. Case Study

The current case study is located in Tallinn, Estonia (Lat. 59°26′ N Lon. 24°45′ E). Specifically, the Taltech University campus is located in the residential neighborhood of Mustamäe and consists of interconnected buildings surrounded by paved areas, car parks, and green spaces (Figure 2). Tallinn is categorized humid continental climate according to Köppen-Geiger classification Dfb [38]. The medium-density neighborhood is populated by concrete housing blocks with heights from 6 to 10 storey concrete flat blocks, industrial and commercial buildings, and a network of major access roads. In addition, the neighborhood also has green spaces with trees between the buildings but not along the streets [39].
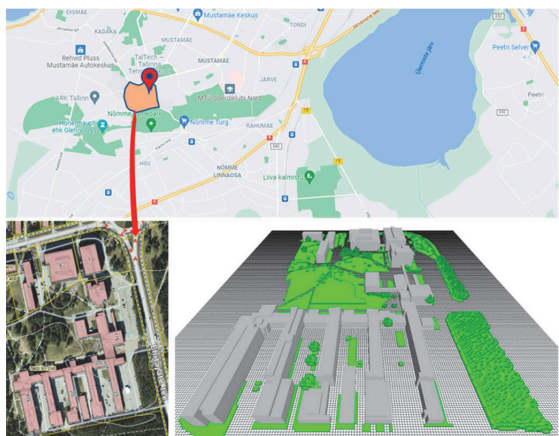
Figure 2. The location of Taltech main campus in Tallinn city map [39], top view of the buildings in the GIS map and the geometric 3D model in ENVI-met.

Thus, the case study encompasses a variety of micro-environments, including buildings, green spaces, parking lots, corridors, sculptures, monuments, benches, lighting elements, etc. In fact, visitors have many opportunities to enjoy the urban environment and engage in different types of activities at different levels. Although the case study hosts people all year round, whether they just enter the buildings or use the surrounding area, people tend to use the campus more in summer when the weather is nice.

2.2. Climatic Data of the Study

Tallinn weather data in June 2021 that Figure 3 shows are dry bulb temperature (Min: 7.2 °C, Average: 19.5 °C, Max: 32.2 °C), relative humidity (Min: 20%, Average: 66%, Max: 100%), and wind speed (Min: 0 m/s, Average: 7.7 m/s, Max: 24 m/s) [38].
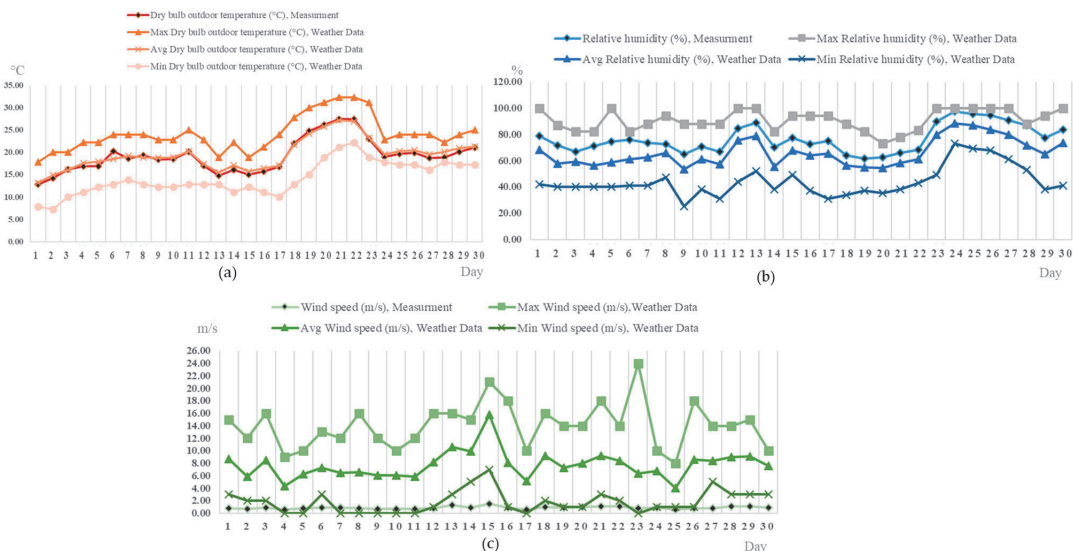
Figure 3. Tallinn, Estonia Weather History, June 2021 [40]. (**a**) Dry bulb temperature, (**b**) relative humidity, and (**c**) wind speed.

In addition, remote data loggers from nZEB Taltech technological test facility [41] were used to record the environmental profile of the outdoor microclimate and environmental features. Mean radiant temperature T_{mrt} (°C) represents the radiative heat load received by a standing human [15]. ENVI-met software that we used for PET assessment is one of the common software for the simulation of Tmrt. Mean radiant temperature T_{mrt} (°C) was calculated by the ENVI-met model by summing the contribution of short wave and long wave and long wave radiant fluxes, from the sun, sky, surrounding buildings surfaces, and ground considering the human body view factors of the flux sources. The software needs the obstacle structure of buildings and the global radiation over the whole sky to calculate T_{mrt} in two dimensions [42].

The microclimatic data collected in June 2021 (Figure 4), considering the weather conditions in the studied area. The measured microclimatic data are: dry bulb temperature (Min: 12.8 °C, Average: 19.2 °C, Max: 27.5 °C), wind direction (Min: 123.3°, Average: 201.3°, Max: 257.8°), wind speed (Min: 0.5 m/s, Average: 0.9 m/s, Max: 1.5 m/s), and relative humidity (Min: 61.6%, Average: 76.5%, Max: 97.4%).
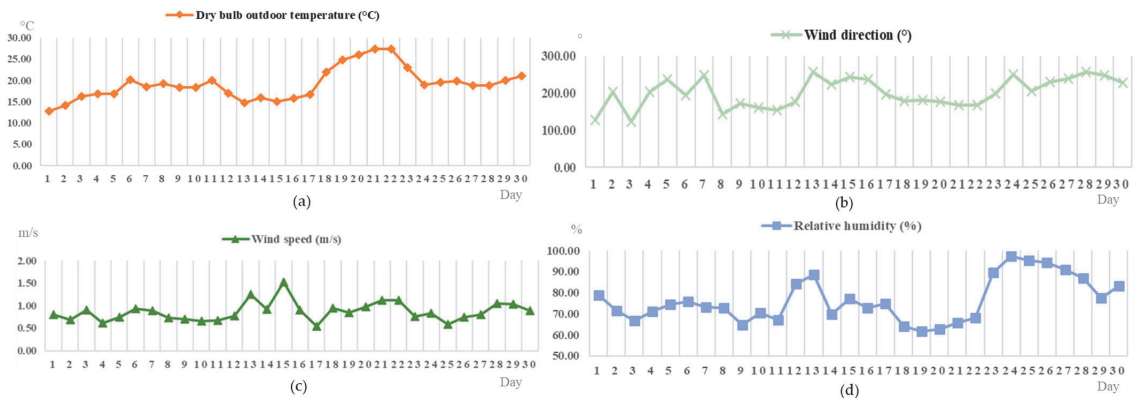
Figure 4. Microclimatic data of the campus, June 2021. (**a**) Dry bulb temperature, (**b**) Wind direction, (**c**) Wind speed, and (**d**) Relative humidity.

2.3. Qualitative Study Based on Semi-Structured Interviews

The present part of the study is grounded on a constructivist methodology because the author accepts that the outdoor thermal sensation of outdoor areas is a complex phenomenon that requires a holistic approach to study it since it depends on each user' construction of reality and how they perceive it [39]. Specifically, the authors used the method of the phenomenological semi-structured interview since it could give detailed insights not only of the subjective perception of local weather and outdoor areas of the Tal-Tech Campus but how this is related to the development of outdoor activities and level of clothing [40,41]. We did not include pre-set questions related to interviewees' thermal sensation but indirect/open questions that would lead us to understand their preferences in terms of outdoor areas, activities, and climate conditions. With this information we could for one side, design coherent online survey and secondly, identify the worthy case studies to be solved by simulations.

The aim of the designed interview is to address the following main research question (RQ0): which are the most valuable aspects of outdoor thermal sensation during summer according to Tal-Tech Campus users? In order to answer RQ0, three research questions have been formulated (Table 2):

Table 2. Definition of question blocks used for the interview.

Block	Questions and Follow-Up Questions
I: Personal background	— Reviewer's introduction — Reminder to the ethic consent given to be recorded — Interviewee's background: gender, nationality, age, education, job position, and city of residence
II: Introduction to the topic	— Presentation of the main concept: outdoor thermal comfort and variables — Presentation of the main research question (RQ0) — Which measurable variable do you appreciate the most? — Which measurable variable do you do not care at all?
III: Outdoor habits during summer	— Which months do you consider as part of summer? Why? — Do you like outdoor activities? Why? — Which outdoor activities do you like to practice during summer? — Tell me the most valuable aspect of these outdoor activities. — How often do you practice them? — When and where do you practice them? Do you practice them alone or in company? — Which level of clothing do you have during these outdoor activities?
IV: Preferred climate conditions during summer	— Which are the most comfortable weather conditions for these preferred outdoor activities? Why? — Which would be a bad day to practice these outdoor activities? Why?
V: Thermal comfort during summer in Taltech Campus	— How often do you go to Taltech Campus? — How do you normally go to the Taltech Campus? — Which spots of the campus do you visit? How often? Why? — Could you point in this map which is the best spot to read/working with the laptop/sunbath/cycling/dancing/picnic. — Which level of clothing would you have for these different situations? Why? — In overall, could you tell me which parts of the campus during summer is the most/less attractive to you? Why?

RQ1: How frequently do Taltech Campus users actually use outdoor areas and why?

RQ2: Which areas of the campus are preferred for different outdoor activities and why?

RQ3: Which are the optimal weather conditions according to Taltech Campus users for different outdoor activities in the Campus?

The interview has been structured in five thematic blocks (Table 2). Block I contains questions that aim to know specific details of the interviewee's personal background such as nationality, age, and city of residence. The objective of Block II is to introduce the re-search topic as well as to discover the valuable climate conditions for the interviewee and to discover how familiar the interviewee is with the topic. Blocks III contains questions that aims to know the interviewee's outdoor habits during summertime in Estonia. The objective of Block IV is to know about the preferred climate conditions during summer in Estonia. Finally, Block V aims to discover the relationship between the interviewee and the Taltech Campus area: visit frequency, outdoor activities, preferred spots, etc.

Ten interviewees with different nationalities and backgrounds were selected. They could give insights to answer RQ0, since they are whether Master/doctoral students or re-search staff in Taltech. The interviews were conducted from 2 July 2021 until 31 September 2021 with the sign of the ethics consent form to protect their privacy [43]. The communication ways were e-mail, posters, and social media such as Facebook posts. The qualitative data collection consisted in the production of transcriptions from each interview [44] and its study with the well-known thematic analysis [45].

2.4. Quantitative Study

2.4.1. Survey Study

We used survey technique to obtain meaningful subjective information from a diverse profiles of Taltech Campus users to set up realistic people cases. The online surveys (the structure of the survey is in Appendix A) were answered by 40 people from September to November 2021. The structure of the survey was built based on the results and outcomes from in depth interview and relevant factors of the metric PET to evaluate outdoor thermal comfort as Figure 1 in the methodology section shows. For the design of the survey, we considered the information explained from the Section 6. Qualitative study based on semi-structured interviews. The online survey had valid 26 answers. There were three thematic blocks as the interview: personal background, preferred weather conditions during summer, and Taltech Campus (Table 3).

Table 3. Definition of question blocks used for the interview.

Thematic Block	Requested Information
Personal background	Place of birth, age, gender, level of education, current occupation, height, weight, physical disability, chronic disease, number of summer periods in Estonia, summer months.
Preferred summer weather conditions	Outdoor activities in summer, preferred level of: solar radiation, temperature, relative humidity, precipitation, wind speed, and sky type.
Taltech Campus	Use months, distance from home, used outdoor areas, less attractive outdoor areas, weekly hours spent in outdoor areas, practised outdoor activities in the campus, main limitation of the actual outdoor areas.

Although results from survey analysis cannot be generalized because of the small sample size, we used these results to set up diverse and realistic campus users' profiles in order to study how to improve their outdoor thermal comfort during summer in Taltech Campus. In this way, we demonstrated how to apply this mixed-method approach to any case study. Statistical analyses of the survey are detailed in Section 3.2.

2.4.2. Modeling and Simulation Process

This part of the study consisted of geometric modeling and CFD simulation. First, all the data of the landscape, buildings, elements, and furniture of the Taltech campus are collected based on on-site observations. Then, the details of the studied area were used for geometric modeling in the ENVI-met software environment, a three-dimensional microscale numerical model that calculates the distribution of heat, momentum, and humidity in the urban environment according to the equations of thermo/fluid dynamics and thermodynamics [46]. We conducted CFD simulations to define the PET-based comfort range for each representative case (i.e., clothing insulation, metabolic rate, physical attributes, etc.) study obtained from statistical analyses from online surveys (built from interviews analyses). The ENVI-met environmental simulation software uses a microclimate model based on the fundamental laws of fluid dynamics and thermodynamics [37]. The simulation process was conducted with ENVI-met to prepare CFD simulations for each day according to the defined scenarios based on the results of the weather preferences of the people in the survey to define the thermal comfort range of the participants based on PET indices.

The thermal comfort assessment was calculated using ENVI-met. In the final stage, using the results of different thermal comfort based on the results of the survey about personal characteristics, activity level, clothing level and preferred weather conditions, different parts of the campus were evaluated to indicate the comfort level and characteristics of each part. Clothing insulation has a major impact on thermal comfort as it affects heat loss and thus thermal balance [47]. In this study, the mean the values for metabolic rate are

65 (class resting), 100 (class low), 165 (class moderate), 230 (class high), and 290 (class very high) W/m^2 for resting, sitting at ease/standing, sustained hand/arm work, and intense work activities, respectively [47].

3. Results

3.1. Interviews Analyses

The 60% of the interviews were conducted virtually due to the pandemic situation as well as availability issues. The interviewees between 23–56 years old were from nine different countries. The 80% of the interviewees were males. There was one senior researcher, one master student and the rest were PhD students from Taltech. The main conclusions after thematic analyses of the interviews are the following:

- June and July were the commonly recognised month as part of the summer.
- Among the most common features of the outdoor activities during summer in Estonia, the interviewees used the following terms: sun, nature, fresh air, and daylight. The preference for a specific spot in the campus depends mainly on the level of nature, accessibility, protection against the sun, the presence of shades, and presence of places comfortable for sitting and working with laptops, laptop battery, climate conditions, and level of privacy. The level of clothing depends mainly on the climate conditions and the type of outdoor activity, but generally they prefer to wear short-sleeves/light clothes in summer (answer to RQ0).
- According to them a perfect summer day would have a temperature between 20 °C and 25 °C with a gentle breeze and almost clear sky (answer to RQ3).
- They consider very hot and humid/stormy days as the less desired ones during summer in Estonia.
- In terms of level of clothing, the 60% of the interviewees like wearing short clothes during summer. The 40% of the interviewees find normal to wear light clothes during summer.
- The half of the interviewees live near the campus so they normally go to the campus by walk. The rest of them would go by car, bus, or even e-scooter. The frequency of visit depends on each person and where they live and work.
- Among 70% of the interviews use the outdoor areas of the campus.

Their favorite areas of the campus are colored in blue in Figure 5, and the less attractive areas are colored in red in Figure 6. The most attractive areas of the campus are concentrated around the inner courtyard of the Taltech campus (answer to RQ2). The vegetation present in these areas and connectivity with majority of main buildings of the university and student houses make this inner courtyard very attractive because of its multi-functionality: several type of activities can be done in this area. In addition, the wooden area next to the U06 building is valued for some interviewees to go cycling, picnicking, or dancing because of the contact with the rough nature and privacy. The less attractive areas of the Taltech Campus are mainly areas with lack of green areas such as the parking lots, dusty pedestrian zones next to Mektory buildings, or noisy areas due to the heavy traffic of the avenue Akadeemia tee. In addition, some interviewees highlighted the thermal discomfort of the garden between student dormitories due to the poor circulation of fresh air and the reflected radiation during summer.

The main areas to develop different activities in the campus considered by the interviewees are concentrated in the main inner courtyard of the Campus, at the same time, parking lots have big potential of improvement. Therefore, the zone of the campus that we aim to study the outdoor thermal comfort because of its potential use and improvement is shown in Figure 5.
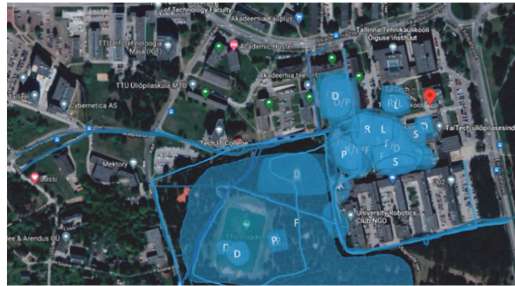
Figure 5. Taltech campus map with colored preferred areas for different outdoor activities. R = reading a book, L = working with the laptop, D = dancing, S = sunbathing, P = picnic, and F = favorite spot. Blue lines represent preferred cycling routes [48].

Figure 6. Taltech campus map with colored less attractive areas by the interviewees [48].

The less attractive areas of the Taltech Campus are mainly areas that lack of green space such as the parking lots, dusty pedestrian zones next to Mektory building or noisy areas due to the heavy traffic of the avenue Akadeemia tee. In addition, some interviewees highlighted the thermal discomfort of the garden between student dormitories due to the poor circulation of fresh air and the reflected radiation during summer.

The main areas to develop different activities in the campus considered by the interviewees are concentrated in the main inner courtyard of the Campus, at the same time, parking lots have big potential of improvement. Therefore, the zone of the campus that we aim to study the outdoor thermal comfort because of its potential use and improvement is shown in Figure 7.
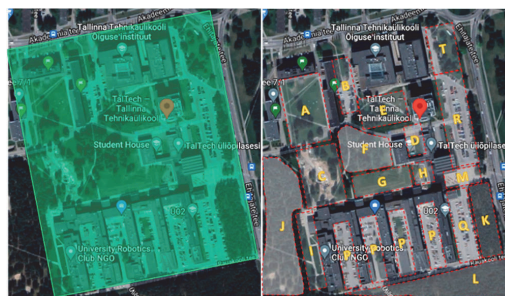
Figure 7. Taltech campus area of study according to the interview's analyses. In the left subfigure, the green zone represents the most critical area of the campus because includes preferred areas and less attractive areas which can be improved in the future. The right subfigure shows the main area divisions referred by the interviewees [48].

The findings obtained from the interviews might contain bias due to the uneven gender balance. Moreover, answers collected virtually from the same interviewee could be different from those collected from face-to-face interviews. The main reasons of these bias sources could be the pandemic situation, interviewees' health, vacation periods, and will to improve the campus in summer, level of applicability of the findings agreed by the practitioners, etc.

3.2. Survey Analyses

The aim of this section is to analyze the survey responses and define the relevant case studies for calculating outdoor thermal comfort. To define these samples, the following analyses are conducted:

First, we highlight the most important features to define campus use samples that have been optimized using the information from the interview analyses. To this end, we developed a simple statistical analysis to define representative profiles of campus users by considering the correlation between variables. In addition, we define the relevant weather conditions and a summer month based on the survey responses. We classify the preferred outdoor activities on campus based on metabolic rates.

Therefore, in order to determine a more accurate PET that is close to a person's thermal sensation, it is important to consider all the important parameters.

Based on the data obtained from the survey, an initial assessment of PET was made by evaluating the CFD model of 4 June in the ENVI-met environment. This step aims to find out the main characteristics and variables that are meaningfully related to PET as thermal comfort indices in the study and should be considered in defining the samples of the study with different definitions. Then, an initial sample of 30 people who use the campus in summer was drawn based on various characteristics—e.g., gender, age, clothing level, activity level (influence on metabolism), height and weight (influence on body surface area). The analysis of BIO-met benefits from the microclimatic data of the campus at 16:00 on the CFD simulation results of 4 June, the first day according to the study scenario.

In this part, Pearson correlation was applied, which shows the correlated variables and PET. According to the definition of Pearson correlation, the values close to 1 show perfect correlation and the low correlated ones close to 0. On one hand, according to Figure 8, the most significant variables correlated with PET are metabolic rate, M (the results of activity level and weight and height) and I_{cl} (the level of clothing). On the other hand, gender and age are not such strongly correlated with PET, hence, we do not analyze survey data separately according to these two variables. Therefore, during the simulation process, we selected the following personal characteristics: clothing insulation (I_{cl}), metabolic rate, weight, height, and age and gender—where we assumed a 30-year-old male in all samples.

We analyzed answers from participants who used somehow the outdoor areas of the campus and facilitated relevant personal information (28 participants): age, gender, weight, and height. There were 46% females and 54% males. Moreover, 46% of participants are Estonian, 14% from Iran, and the rest of participants are from different countries such as Philippines, Spain, Pakistan, Indonesia, Russia, and Mexico. Participants' occupation was very diverse: from assistant professors to bachelor students. Most of participants (95%) do not have chronic disease and none of the respondents had a physical disability. The minimum, mean, and maximum age are 20, 30, and 44 years old, respectively.

Firstly, we calculated the body mass index (BMI) (weight in kg)/(height in m^2) from height and weight data (Figure 8). Secondly, we calculated representative values of the studied sample: minimum, Q1, Q2, Q3, and maximum values (Table 4) of relevant factors of the PET metric. In addition, we selected the closest weights related to each of the representative values of BMI values from survey answers. Finally, we determined the height from these representative BMI and weight values (Table 4). Furthermore, we defined 5 BMI-driven subject profiles (Table 4). As we mentioned, we fixed age and gender variables. According to our survey answers, the gender and age of these five profiles could be equivalent to a 30-years old male, respectively.
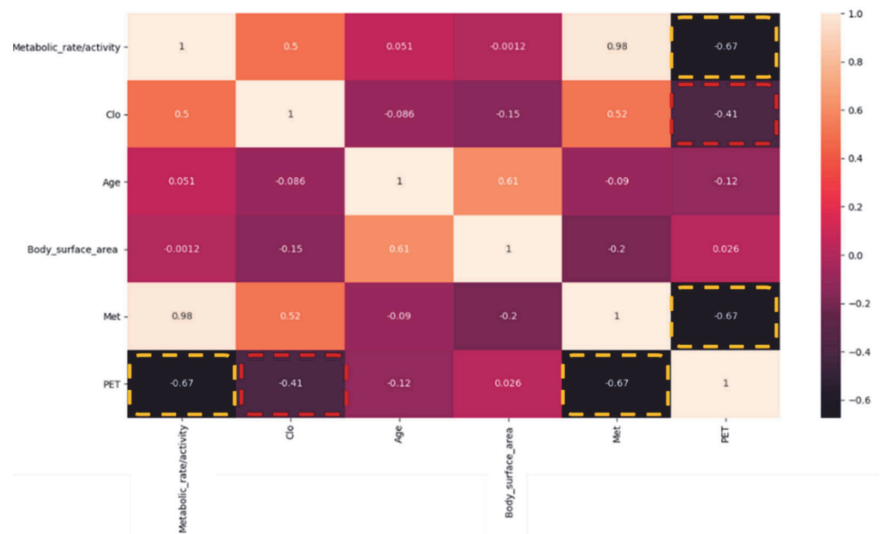
Figure 8. Dependency values between PET metric and factors in outdoor thermal comfort.

Table 4. Classification of metabolic rates by activity.

Value	BMI (kg/m^2)	Weight (kg)	Height (m)	Profile ID
Minimum	17.99	52	1.70	1
Q1	20.10	67	1.83	2
Q2	23.52	68	1.70	3
Q3	25.78	83	1.79	4
Maximum	31.55	101	1.79	5

The months in which the participants used more the Taltech Campus are May (92.9%) and September (71.4%) (Figure 9). However, May and September are considered as part of the summer period in Estonia by only a 7.1% and 3.6% of the participants, respectively (Figure 10). Although July is considered the summer month by most participants (96.4%), only a 53.6% of the participants used the outdoor areas of the campus against 71.4% related to June. Furthermore, we select June as the most relevant summer month by our studied sample because of its consideration of summer month and frequency of use of the outdoor areas of the campus.
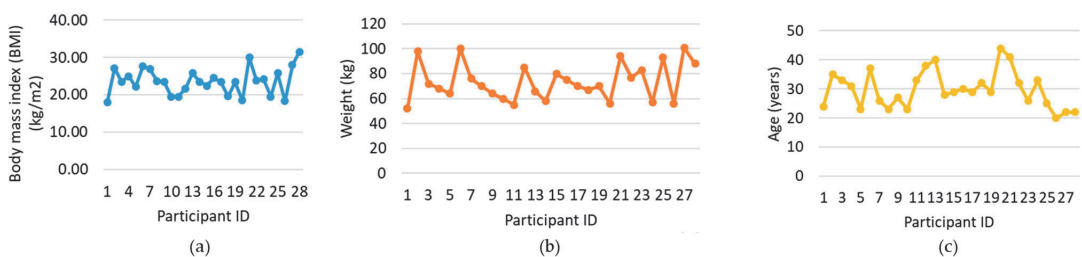
Figure 9. Participants' body mass index (BMI) (**a**), weight (**b**), and age (**c**).
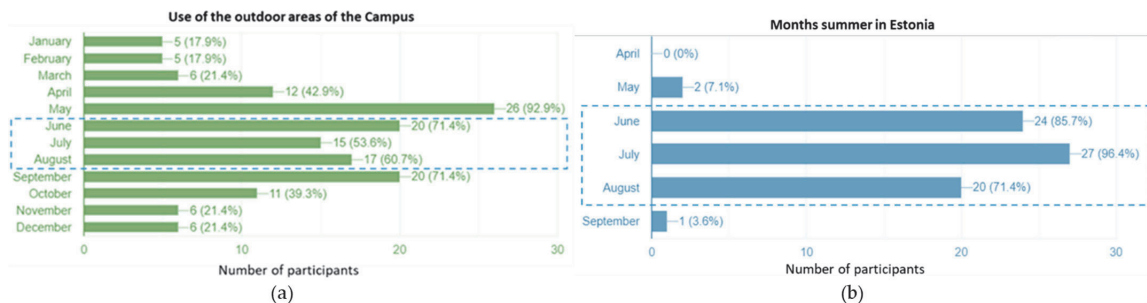
Figure 10. Months in which participants use the Taltech Campus (**a**) and considered summer months (**b**).

In terms of preferred weather conditions (Figure 10), the overall perfect summer day according to most participants include the following weather conditions: medium solar radiation (75%) (Uvi = 3–5) (Figure 11a), medium temperature (60.7%) (20–25.9 °C) (Figure 11b), medium relative humidity (57.1%) (40–60%) (Figure 11c), no rain (82.1%) (Figure 11d), gentle breeze (wind speed < 4 m/s) (Figure 11e), and full clear sky (50%) (Figure 11f). Furthermore, we consider these weather conditions in the next simulation-based CFD analysis as the input weather data based on measured campus microclimatic data and Tallinn climatic data. Different weather conditions than these ones might not encourage participants to spend time in the outdoor areas of the Campus, and therefore the analysis of the outdoor thermal comfort is not as relevant as under preferred weather conditions.
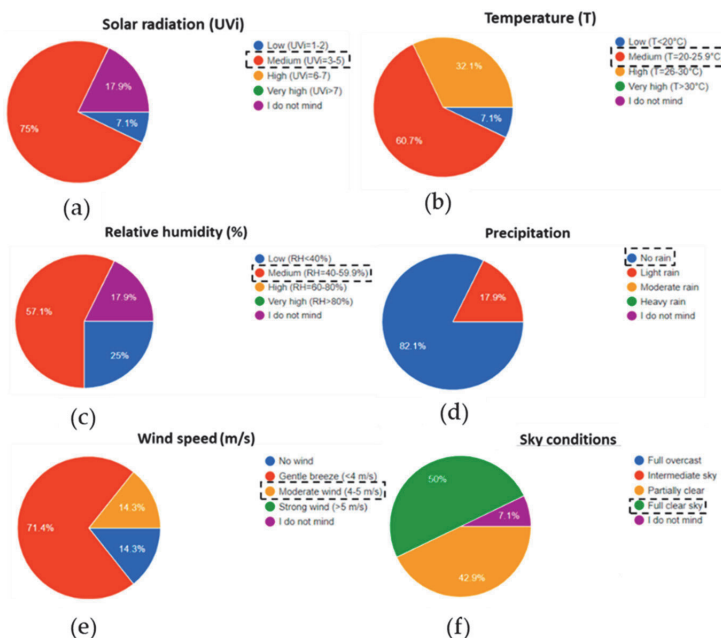
Figure 11. Preferred levels of weather conditions according to survey participants: solar radiation in terms of UV-index (UVi) (**a**), outdoor dry bulb temperature (T) (**b**), relative humidity (%) (**c**), precipitation (**d**), wind speed (m/s) (**e**), and sky condition (**f**). The classes used in these survey questions were designed according to interviews' answers (Section 3.1).

The scenarios that define the days of the CFD simulation process and the climate data are based on two sources. First, on the preferences of the respondents in the survey and the

days they described as most thermally comfortable conditions in the Tallinn climate data (Table 5), and second, on the same preferred conditions in days based on the Taltech campus microclimatic data temperature 20 °C, wind speed 0.7 m/s and relative humidity 67%.

Table 5. Selected days for CFD simulation, based on the survey. Microclimatic data of Taltech campus in June 2021.

Date	Air Dry Bulb Temperature (°C)	Solar Radiation (UV-Index)	Wind Speed (m/s)	Wind Direction (°)	RH (%)	Precipitation (Yes/No)	Sky Type
4 June	20–22	5	3.3	90	46–75	No	Passing clouds
7 June	21–23	5.2	4.2	90	48–77	No	Passing clouds
10 June	22–23	5.5	4. 2	90	46–82	No	Passing clouds
11 June	24–25	5.7	3.9	90	35–85	No	Passing clouds

Once the profiles of the individuals and the weather conditions are established, the most common metabolic rates associated with common outdoor activities on campus need to be defined. There are different classes of activities depending on the metabolic rate (Table 6).

Table 6. Classification of metabolic rates by activity developed in the Taltech campus [19].

Class	Metabolic Rate (W/m^2)	Example	Outdoor Activities (Number of Times Selected)	Selected Times
Resting	65	Resting	Chilling out in outdoor areas (14), read a book (6)	20
Low	100	Sitting at ease/standing	Go picnic (9), e-scooter (3)	12
Moderate	165	Sustained hand/arm work	Hang out (18), walking (18), go camping (1)	37
High	230	Intense work	Jogging (8), hiking (3), table tennis (2), cycling (1)	14
Very high	290	Very intense to maximum activity	Running (5), beach-volleyball (1), badminton (1) street workout (3), basketball (3), orienteering (1)	14

According to the outdoor activities developed in the outdoor areas of the campus by the survey participants, activities with very low metabolic rate (65 W/m^2) such as "chilling out" or "read a book" were selected by the participants 20 times. Moreover, activities with moderate metabolic rate (165 W/m^2) such as "hang out", "walking", or "go camping" were selected 37 times in total. On the other hand, "low", "high", and "very high" outdoor activities were not such popular for the participants. This could be due to the limited space and ground conditions in the part of the campus analyzed (Figure 11) for certain activities such as e-scooter, hiking, cycling, running, beach-volleyball, badminton, street workout, basketball, etc.

Therefore, we consider "very low" and "moderate" level of activities as the most common ones practiced by the participants in the outdoor areas of the campus during summer.

Input Elements to Assess Thermal Comfort Based on the Survey

Some of the parameters that lead to a person feel thermally comfortable are inherent in the characteristics of a space, while others are more personal to the environment. The complexity of these relationships means that none of them follow a simple cause-and-effect situation [33]. The data that defined the sample of thermal comfort assessment studies are based on survey information and literature reviews [19,49–51]. According to participants' answers the preferable level of clothing during summer could vary from swim trunks/bikini with sandals and a cap to short sleeved clothes. Despite of the influence of the body movements and air action on clothing thermophysical properties [52], we used a simplified approach based on static I_{cl} values. Specifically, the static I_{cl} values, meaning the I_{cl} value for a person wearing a bikini (upper and lower part) or swim trunks (0.1 and 0.3 clo, respectively) [18,19,49]. Moreover, we considered a I_{cl} of 0.5 clo as an average value to represent light summer short-sleeve clothes according to Blazejczyk [49].

The spatial distribution of outdoor thermal comfort in terms of the metric PET is calculated via simulation for 20 defined cases (5 BMIs × 2 Icls × 2 Ms) study under preferred summer conditions (Figure 11) in Section 3.2.

Finally, we define 20 cases (5 BMIs × 2 I_{cl}s × 2 Ms) of interest that represent our analyzed sample. As can be seen in Table 7, we can consider a 30-year-old man with different BMI conducting activities with very low (65 W/m^2) and moderate (165 W/m^2) metabolic rates while wearing different summer outfits (I_{cl} of 0.1 and 0.5 clo) under different climate conditions. The average PET assessed for each case study experience under the supposed microclimatic data in the campus.

Table 7. Definition of the cases study considered for the calculation of outdoor thermal comfort using the PET metric under preferred weather conditions, the level of activity, and I_{cl}.

	Personal Data of Samples				Average PET (°C) in the Campus Area				
					Campus Microclimatic Data				Tallinn Climatic Data
Case ID	Weight (kg)	Height (m)	Metabolic Rate (W/m^2)	I_{cl} (clo)	4 June	7 June	10 June	11 June	11 June
p1m1c1	52	1.70	65	0.1	43	44	30.2	47.2	43.9
p2m1c1	67	1.83	65	0.1	36	38.2	49.8	40.1	43.6
p3m1c1	68	1.70	65	0.1	42.3	42.7	40.3	48.2	44.1
p4m1c1	83	1.79	65	0.1	31.3	41.7	41.3	44.8	44.2
p5m1c1	101	1.79	65	0.1	35.7	37.1	29.3	37.2	43.8
p1m2c1	52	1.70	165	0.1	39.3	38.7	38.7	39.9	28.6
p2m2c1	67	1.83	165	0.1	40.1	37	38.5	37.6	29.2
p3m2c1	68	1.70	165	0.1	29.5	35.6	39.4	39.7	40.3
p4m2c1	83	1.79	165	0.1	45.6	38.4	39.1	36.3	40.3
p5m2c1	101	1.79	165	0.1	38.8	39.1	29.7	39.6	39.5
p1m1c2	52	1.70	65	0.5	39.4	33.8	39.1	46.6	42.2
p2m1c2	67	1.83	65	0.5	42.6	40.1	44.5	45.9	37.9
p3m1c2	68	1.70	65	0.5	33	37	27	46.4	29.8
p4m1c2	83	1.79	65	0.5	43.6	37.9	25.4	45.8	42.9
p5m1c2	101	1.79	65	0.5	33.7	30.4	39.7	47.7	42.9
p1m2c2	52	1.70	165	0.5	38.8	38.3	38.3	39.2	40.1
p2m2c2	67	1.83	165	0.5	36.4	36.4	30.5	37.5	38.4
p3m2c2	68	1.70	165	0.5	30.1	37.6	29.6	30.6	39.9
p4m2c2	83	1.79	165	0.5	38.3	31.7	39	39	39.4
p5m2c2	101	1.79	165	0.5	38.4	37.8	35.2	38.6	41.3

3.3. CFD Simulation Analysis and Results

Since the study focuses on comparing the objective personal and climatic data with people's subjective thermal perception data, needs, and preferences, this section considers all the collected and analyzed survey data from Table 8 to define the thermal comfort of people with different activity levels, metabolic rates, clothing levels, and body surfaces in PET indices in the studied area.

In the next step, the data of each person considered to evaluate PET on four days based on the microclimatic data of the studied area on 4, 7, 10, and 11 June 2021 and the weather data of Tallinn on 11 June 2021 using the BIO-met function of ENVI-met software as a post-processor tool to calculate the human thermal comfort indices from the output files of ENVI-met model.

The simulation of CFD was conducted for each date at 16:00 (the hottest time of the day based on people's weather preferences) and the atmospheric data were used as input data to evaluate PET for each sample in the study.

Table 8. Evaluation of PET range in different samples, based on the microclimatic data.

The Average PET Range	Type of Activity	I_{cl} (clo)	Height (cm)	Weight (kg)	Dry Bulb Temperature (°C)	Wind Speed (m/s)	Relative Humidity (%)
Slightly warm (23–29 °C)	Resting	0.5	170	68			
	Resting	0.5	179	83	22–23	4.17	46–82
Warm (29–35 °C)	Resting	0.1	179	83			
	Resting, sustained arm or hand work	0.1, 0.5	170	68	20–22	3.33	46–75
	Resting	0.5	179	101			
	Resting	0.5	170	52			
	Resting	0.5	179	101	21–23	4.16	48–77
	Sustained arm or hand work	0.5	179	83			
	Resting, sustained arm or hand work	0.1, 0.5	170	52, 68			
	Resting, sustained arm or hand work	0.1, 0.5	179	101	22–23	4.17	46–82
	Sustained arm or hand work	0.5	183	67			
	Sustained arm or hand work	0.5	170	68	24–25	3.88	35–85
Hot (35–41 °C)	Resting, sustained arm or hand work	0.1, 0.5	183	67			
	Resting, sustained arm or hand work	0.1, 0.5	179	101	20–22	3.33	46–75
	Resting, sustained arm or hand work	0.1, 0.5	170	52			
	Sustained arm or hand work	0.5	179	83			
	Resting, sustained arm or hand work	0.1, 0.5	183	67			
	Resting, sustained arm or hand work	0.1, 0.5	179	83, 101	21–23	4.16	48–77
	Sustained arm or hand work	0.1, 0.5	170	52			
	Sustained arm or hand work	0.1, 0.5	170	68			
	Resting, sustained arm or hand work	0.1	170	68			
	Sustained arm or hand work	0.1	183	67			
	Sustained arm or hand work	0.1, 0.5	179	83	22–23	4.17	46–82
	Resting, sustained arm or hand work	0.1, 0.5	170	52			
	Resting, sustained arm or hand work	0.5	179	101			
	Resting, sustained arm or hand work	0.1, 0.5	183	67			
	Resting, sustained arm or hand work	0.1, 0.5	179	101			
	Sustained arm or hand work	0.1, 0.5	170	52	24–25	3.88	35–85
	Sustained arm or hand work	0.1	170	68			
	Sustained arm or hand work	0.1, 0.5	179	83			
Very hot (>41 °C)	Resting	0.1	170	52, 68			
	Resting, sustained arm or hand work	0.1, 0.5	179	83	20–22	3.33	46–75
	Resting	0.5	183	67			
	Resting	0.1	170	52, 68			
	Resting	0.1	179	83	21–23	4.16	48–77
	Resting	0.1, 0.5	183	67			
	Resting	0.1, 0.5	170	52, 68			
	Resting	0.1, 0.5	179	83			
	Resting	0.5	179	101	24–25	3.88	35–85
	Resting	0.5	183	67			

4. Analyzing the Results of PET Evaluation

The aims of the final section consist of analyzing results of the CFD simulation to reach a better understanding of the thermal comfort condition of the campus, the causes and effects on comfort and non-comfort areas, and give suggestions to improve outdoor thermal comfort.

The first analysis of the results of PET relates to how the range of PET depends on the personal and weather data. Table 8 shows the ranges of PET in different samples with the range of activities (metabolic rate: 65 W/m^2 at rest and 165 W/m^2 at high activity), I_{cl} (0.1 and 0.5 clo) and microclimatic conditions on campus on 4, 7, 10, and 11 June. For example, according to Table 8, on a day with a dry bulb temperature of 22–23 °C, wind speed of 4.17 m/s and relative humidity of 46–82%, only people with a height of 170 and

179 cm and weight of 68 and 83 kg, respectively, can experience a slightly warm thermal comfort feeling, while their clothing level is 0.5 and they are doing the least activity.

Thus, the results of Table 8 give enough data to understand the relation between thermal comfort value, human body surface, the activity level, I_{cl}, and weather data. For example, in the same weather condition a person of 170 cm height and 68 kg weight can experience different ranges of PET that represent a normal body area, depending on the level of activity and preferred clothing level

Figure 12 shows a simple example based on the microclimatic data of 10 June 2021. For example, if a person has only a baseline activity level (65 W/m^2) and wears the lowest clothing (0.1), it is hot PET (i.e., 35–41 °C), while the person with more clothing (I_{cl} = 0.5) has a lower value of PET, i.e., slightly warm in the range of 23–29 °C. If the same person has a higher activity level (metabolic rate of 165 W/m^2) with the same microclimatic data, the thermal comfort is hot or warm, at I_{cl} 0.1 and 0.5 respectively.
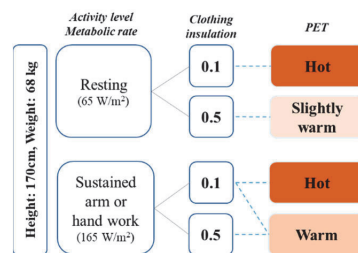
Figure 12. Different PET ranges, based on the microclimatic data on 10 June 2021.

5. Application of the Study Results

5.1. Visualizing the Results

The visual map calibrates the output data of the PET results to facilitate the application of thermal comfort while showing the diversity of the thermal environment.

Figure 13a,b shows how activity level affects the thermal comfort of PET for two identical individuals engaged in different activities. For the case p3m2c1 with high metabolic rate and low clothing, thermal comfort in different parts of the campus ranges 24.5–55.5 °C (Figure 13a). According to thermal comfort evaluation, in this condition, most parts of the campus are comfort with slightly warm and warm thermal comfort values. However, in Figure 13b evaluated thermal comfort value for the person (p4m1c2), with lower metabolic rate and higher level of clothing is in the higher ranges 27.70–54.10 °C. Thus, most areas of the campus are considered non-comfort areas, meaning hot or very hot.

Figure 13. Visual results of PET of evaluation sample p3m2c1 on 4 June (high level of activity) (**a**), Visual results of PET of evaluation sample p4m1c2 on 4 June (low level of activity) (**b**).

5.2. Mitigation Strategies for the Improvement Thermal Comfort

This paper aims to assess outdoor thermal comfort by focusing on bridging the disciplines by examining the relationships between objectively measured w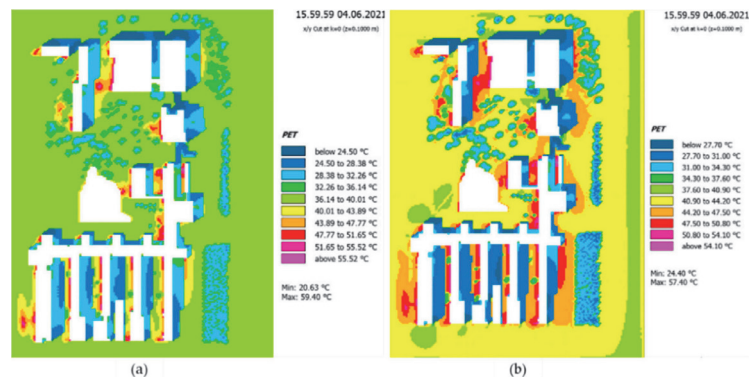eather conditions and subjectively perceived thermal comfort, as well as the spatial characteristics of the area studied, as explained in Section 1.2. Therefore, matching the results is necessary to find the optimal areas for thermal comfort with the areas that need to be improved to meet people's needs. Similarly, we determine which features and characteristics cause this zone to be highlighted with high or low levels of thermal comfort perception according to PET ranges shown in Table 8. Firstly, we searched for areas in the high range of thermal comfort in each date (Figures 14 and 15). Secondly, we looked for the reason why some areas are thermally comfortable and others are not. In this step, the areas that have PET in the range of 35–41 °C and more than 41 °C are classified as "hot" and "very hot", respectively. We consider "very hot" areas needed of thermal comfort improvements. The final part of this assessment is to find out which areas of the campus are more suitable for different activities.

Figure 14. Visual results of PET and areas with PET values higher than need improvement, sample p4m2c1, date 4 June, based on the microclimatic data of the campus (**a**), sample p1m1c1, date 7 June, based on the microclimatic data of the campus (**b**).
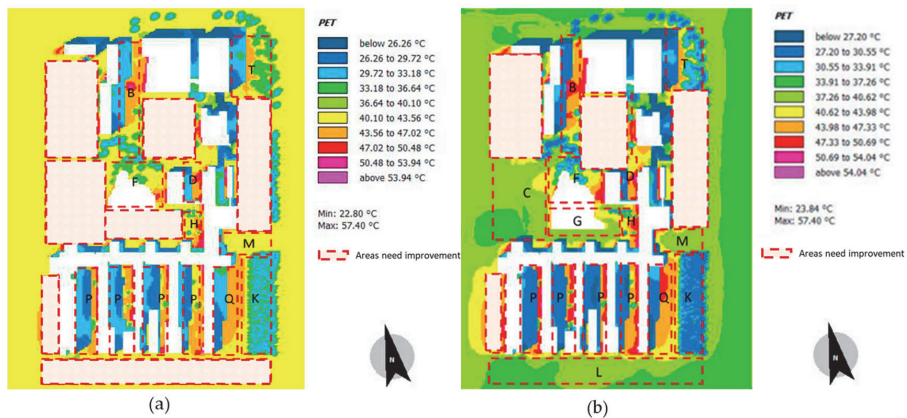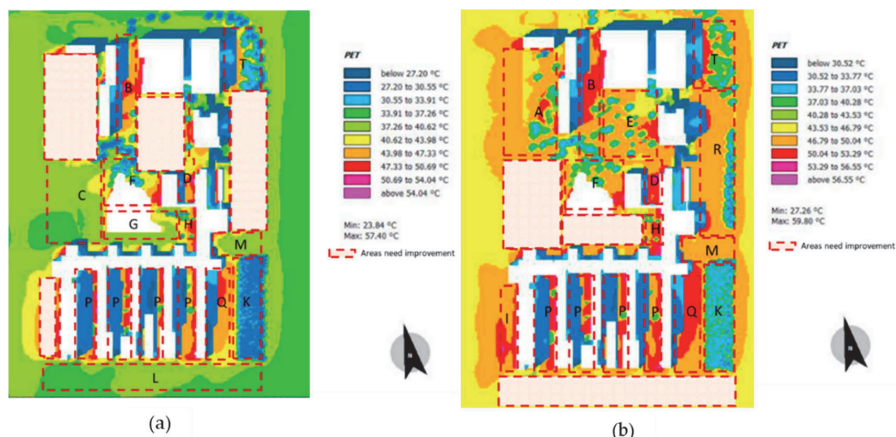

Figure 15. Visual results of PET and areas with PET values higher than need improvement sample p2m1c1, date 10 June, based on the microclimatic data of the campus (**a**), sample p3m1c1, date 11 June, based on the microclimatic data of the campus (**b**).

In summary, based on the analysis and according to Figure 16, outdoor areas: A, C, E, G, L, I, and R needs improvements to offer the better thermal comfort.
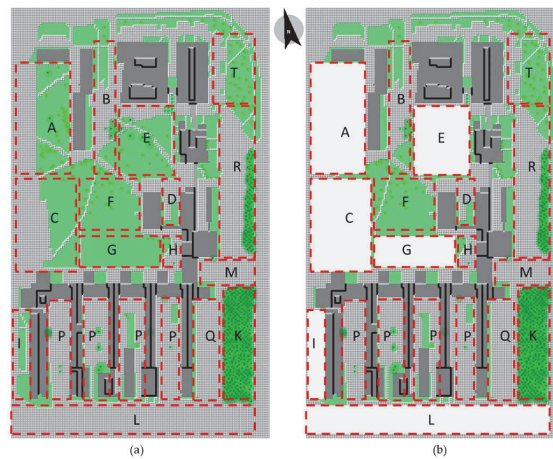

Figure 16. Different areas in the campus, generated in ENVI-met (**a**), areas needing improvement (**b**).

Table 9 uses the information areas in the campus in Figure 16 to define the major elements and the possible activities (the common activities based on the survey, Table 9). In this part of the study, we use the label of comfort or non-comfort to categorize areas in the campus based on PET assessment in the study.

Table 9. Specification of different part of the campus and the comfort level based on PET analysis.

Name	The Major Elements	Possible Activities (Based on PET Evaluation)	Main Plants	Comfort Lable
A	Green area	Chilling out, reading	Grass, trees	Non-comfort
B	pavement	Hang out, walking, camping	Grass, bushes	Comfort
C	Green area, pavement	Chilling out, reading	Grass, bushes	Non-comfort
D	Green area, pavement	Hang out, walking, camping	Grass, pavement	Non-comfort
E	Green area	Chilling out, reading	Grass, bushes, trees	Comfort
F	Green area	Hang out, walking, camping	Grass, bushes, trees	Non-comfort
G	Green area	Chilling out, reading	Grass	Comfort
H	Green area, pavement	Hang out, walking, camping	Grass	Comfort
I	Green area, pavement	Chilling out, reading	Grass	Non-comfort
K	Green area	Hang out, walking, camping	Grass, trees	
L	Pavement, street	Chilling out, reading	-	Non-comfort
M	Pavement, parking lot	Hang out, walking, camping	-	Comfort
Q	Pavement, parking lot, green area	Hang out, walking, camping	-	Comfort
P	Pavement, parking lot, green area	Hang out, walking, camping	Grass, tress	Comfort

According to the survey data and the results of Table 9, Figure 17 shows different areas of the campus under the labels to define which areas are comfortable and popular, which are not comfortable but preferred by people, and likewise which areas are comfortable but not attractive to people based on the analysis, and finally which areas are neither comfortable nor attractive to people.

Figure 17 also shows that people prefer to spend their time and various activities such as computer work, dancing, picnicking, sunbathing, etc. in the main inner courtyard and mostly green areas of the campus such as A, C, E, and G. The area labeled L is also not

pleasant but is popularly used as a bicycle path. Although park areas such as P, Q, R, and K offer thermally pleasant conditions and benefit from the shade of surrounding buildings, they are not attractive to people. In addition, some interviewees consider these areas to be places with great potential for improvement. There are also some areas in the middle of the campus, (e.g., B, F, D, and H), which are both comfortable and used by people for various activities such as gathering, reading, and sunbathing. Finally, the area marked T in the northern part of the map provides optimal thermal comfort and is also a green space, but not very popular due to the high acoustic pollution related to the traffic flow
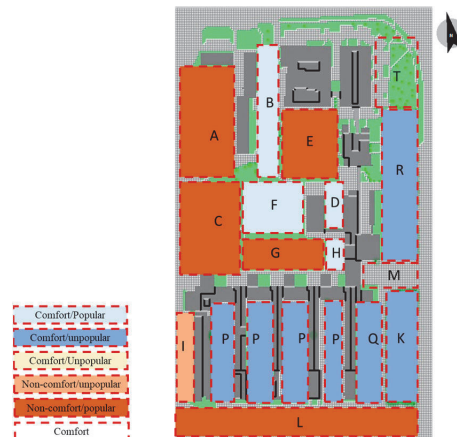

Figure 17. Comparison between the areas that are more popular for different activities and the comfort or non-comfort areas based on PET analysis.

6. Conclusions

The aim of the study is to use objective and subjective elements to assess outdoor thermal comfort at Tallinn University of Technology, Tallinn, Estonia, using qualitative and quantitative approaches. In this study, we mainly wanted to show how to use this mixed-methods approach to identify which areas of an urban area would be needed improvement of the outdoor thermal comfort during summer. The main outcomes and practical implications of this study are as follows:

1. The analysis of the relationship between personal characteristics and perception of thermal comfort shows that height, weight, degree of clothing, and activity level have an impact on thermal comfort values. The result shows that when the activity level is high, the level of clothing has no significant influence on the thermal comfort value. Furthermore, comparing people with high and low BMI under the same microclimatic conditions, with the same activity and clothing level, the maximum values of PET are higher in people with low BMI than in people with high BMI. Therefore, people with a lower BMI can move more under the same weather conditions without feeling uncomfortable compared to those with higher BMIs.

2. The analysis of non-environmental variables that depend on people has shown that the choice of place where people are active sometimes depends not only on pleasant thermal conditions, but also on other characteristics of the place, such as visual enjoyment, openness, or enclosure and green spaces. It is like some sun-exposed areas that are not pleasant according to the thermal comfort analysis, but which people choose for their activities. It is the same reason with areas between buildings or car parks that are shaded by the surrounding buildings and provide optimal thermal conditions, but are not chosen by people as places to stay because the lack of green/semi-open areas, and urban furniture. Based on the mentioned findings, it seems that spatial planning for site renovation and improvement should focus on the

physical aspects of the areas, taking into account people's preferences to provide the enjoyable environment for visitors, while considering the thermal comfort of outdoor residents, e.g., by providing some shaded places. This concept covers part of Taltech's strategic plan in the project called Green Transition.

The sample size is a limitation of our study. Furthermore, we could not generalize our findings to all Taltech Campus users. Moreover, in this study assessment of thermal comfort is based on the PET metrics, but it can be done and evaluated in other indices such as the UTCI among others. The other limitation is related to the limited area and duration that thermal comfort assessed. Our assessment criterion of thermal properties of clothing might be conservative because we did not consider the wind effect on static clothing insulation. Thus, in summer conditions, the thermal comfort levels could be underestimated. Nevertheless, this approach could lead into design strategies which are on the safe side in future microclimate conditions, which can be further studied in future investigations.

Other future studies to complete this research and cover the limitations are planned in two parts. the study in the field of urban thermal comfort analysis in the campus will first explore possible solutions and strategies based on people's preferences and the results of the present study to improve the quality of thermal comfort on campus. Secondly, applying the solutions and reflecting them in the software environment of the urban simulator to analyses the CFD simulation of the campus to find the optimal one to improve areas with low thermal and non-thermal comfort. A future study will take into account a survey done with more participants in other zones of the city. The study will follow the results of thermal comfort assessment in different seasons in Tallinn through different comfort indices.

Author Contributions: Conceptualization, N.E., A.S. and F.D.L.; Applying qualitative method, A.S.; Applying quantitative method, modeling, simulation N.E.; Analysis, N.E. and A.S.; Methodology, N.E. and A.S.; Supervision, F.D.L. and K.S.L.; Results, N.E. and A.S.; Writing—original draft, N.E.; Writing—review and editing, N.E., A.S., F.D.L. and K.S.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Tallinn University of Technology (Council Resolution No. 44 of 17 October 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

CFD	Computational Fluid Dynamics
PET	Physiological Effective Temperature
PMV	Predicted Mean Vote Index
PPD	Predicted Percentage Dissatisfied Index
ISO	International Organization for Standardization
MEMI	Munich Energy-Balance Model for Individuals
UTCI	Universal Thermal Climate Index
TSV	Thermal Sensation Votes
BMI	Body Mass Index
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
OUT_SET	Outdoor Standard Effective Temperature

Appendix A

The structure of the online survey

1. How old are you?
2. Which is your gender?

   [Female, Male, Transgender female, Transgender female, Gender variant/non-coforming, Not listed, Prefer not to answer]

3. Which is your level of education?
4. Which is your current occupation?
5. Which months do you consider as part of the summer in Estonia?

   [June, July, August]

6. How many summer periods have you been in Estonia?
7. How much is your height?
8. How much is your weight?
9. Do you have any physical disability? In affirmative case, which one?
10. Do you have any chronic disease? In affirmative case, which one?
11. Which outdoor activities do you usually practice during summer in Estonia?

   Swimming at lakes, Swimming at the sea, Go surfing, Go paddle, Go surfing, Go kitesurfing, Tennis, Table tennis, Hang-out, Read a book, Dancing, Cricket, Baseball, Go picnic, Go Camping, Running, Jogging, Walking, Hiking, Football, Basketball, Beach-volleyball, Combat sports, Badminton, Street workout, Skateboarding, Longboarding, (e-)scooter, Chilling out, Other(s)

12. Please select the level of these variables to create a perfect summer day in Estonia to practice any of your preferred outdoor activities mentioned in question 11 (UVi = UV index):

Solar radiation	Low (UVi = 1–2)	Medium (UVi = 3–5)	High (UVi = 6–7)	Very high (UVi > 7)	I do not mind
Temperature (T)	Low T < 20 °C	Medium T = 20–25.9 °C	High T = 26–30 °C	Very high T > 30 °C	I do not mind
Relative humidity (RH)	Low (RH < 40%)	Medium (RH = 40–59.9%)	High (RH = 60–80%)	Very high (RH > 80%)	I do not mind
Precipitation	No rain	Light rain	Moderate rain	Heavy rain	I do not mind
Wind speed	No wind	Gentle breeze (<4 m/s)	Moderate wind (4–5 m/s)	Strong wind (>5 m/s)	I do not mind
Type of sky	Full overcast	Intermediate sky	Partially clear	Full clear sky	I do not mind

13. In which months do you use the outdoor areas of the Taltech Campus?

    [June, July, August]

14. How far (in km) do you live from Taltech Campus?
15. In which areas of the Taltech campus do you often spend time? (Figure 7)
16. According to your opinion, which spot(s) of the TalTech Campus are the less attractive to spend time in? (Figure 7)
17. How many hours per week do you enjoy the outdoor areas of the Campus during summer?
18. Which outdoor activities do you practice within the Taltech Campus during summer in Estonia?

    Swimming at lakes, Swimming at the sea, Go surfing, Go paddle surfing, Go kitesurfing, Tennis, Table tennis, Hang-out, Read a book, Dancing, Cricket, Baseball, Go picnic, Go Camping, Running, Jogging, Walking, Hiking, Football, Basketball, Beach-volleyball, Combat sports, Badminton, Street workout, Skateboarding, Longboarding, (e-)scooter, Chilling out, Other(s)

19. Please, select the main limitations of the actual outdoor areas of the Taltech Campus:

    Green areas, Lack of tables and chairs, Places with privacy, Conditions of the ground, Conditions of the ground, Actual laces to sit are uncomfortable, Accessibility with different ways of transport, Lack of plugs/sockets, Dirty spots, Lack of enough shadow in certain spots, Lack of benches in certain spots, Other(s)

20. Please, leave in the following box any comment you would like to express:

References

1. Chen, L.; Ng, E.Y.Y. Outdoor thermal comfort and outdoor activities: A review of research in the past decade. *Cities* **2012**, *29*, 118–125. [CrossRef]
2. Eslamirad, N.; Mahdavinejad, M. Multi Objective Computing and Applying Expert System in Double Skin Façade system. In Proceedings of the e-Energy '18: The Ninth International Conference on Future Energy Systems, Karlsruhe, Germany, 12–15 June 2018; pp. 459–461. [CrossRef]
3. Lai, D.; Zhou, C.; Huang, J.; Jiang, Y.; Long, Z.; Chen, Q. Outdoor space quality: A field study in an urban residential community in central China. *Energy Build.* **2014**, *68*, 713–720. [CrossRef]
4. Gehl, J. Life between the Buildings: Using Public Space. Available online: http://www.krost-concern.ru/upload/life-between-buildings-1chapter.pdf (accessed on 15 February 2022).
5. Lenzholzer, S.; Klemm, W.; Vasilikou, C. Qualitative methods to explore thermo-spatial perception in outdoor urban spaces. *Urban Clim.* **2018**, *23*, 231–249. [CrossRef]
6. Ahmed, K.S. Comfort in urban spaces: Defining the boundaries of outdoor thermal comfort for the tropical urban environments. *Energy Build.* **2003**, *35*, 103–110. [CrossRef]
7. Nikolopoulou, M.; Lykoudis, S. Thermal comfort in outdoor urban spaces: Analysis across different European countries. *Build. Environ.* **2006**, *41*, 1455–1470. [CrossRef]
8. Wang, Z.; Wang, J.; He, Y.; Liu, Y.; Lin, B.; Hong, T. Dimension analysis of subjective thermal comfort metrics based on ASHRAE Global Thermal Comfort Database using machine learning. *J. Build. Eng.* **2019**, *29*, 101120. [CrossRef]
9. Alur, R.; Deb, C. The significance of Physiological Equivalent Temperature (PET) in outdoor thermal comfort studies. *Int. J. Eng. Sci. Technol.* **2010**, *2*, 2825–2828.
10. Luo, Y.; He, J.; Ni, Y. Analysis of urban ventilation potential using rule-based modeling. *Comput. Environ. Urban Syst.* **2017**, *66*, 13–22. [CrossRef]
11. Eslamirad, N.; De Luca, F.; Lylykangas, K.S. The role of building morphology on pedestrian level comfort in Northern climate. *J. Phys. Conf. Ser.* **2021**, *2042*, 012053. [CrossRef]
12. Li, K.; Zhang, Y.; Zhao, L. Outdoor thermal comfort and activities in the urban residential community in a humid subtropical area of China. *Energy Build.* **2016**, *133*, 498–511. [CrossRef]
13. Taleghani, M.; Kleerekoper, L.; Tenpierik, M.; van den Dobbelsteen, A. Outdoor thermal comfort within five different urban forms in the Netherlands. *Build. Environ.* **2015**, *83*, 65–78. [CrossRef]
14. Lai, D.; Guo, D.; Hou, Y.; Lin, C.; Chen, Q. Studies of outdoor thermal comfort in northern China. *Build. Environ.* **2014**, *77*, 110–118. [CrossRef]
15. Shaw, E. Thermal Comfort: Analysis and applications in environmental engineering, by P. O. Fanger. 244 pp. DANISH TECHNICAL PRESS. Copenhagen, Denmark, 1970. Danish Kr. 76, 50. *R. Soc. Health J.* **1972**, *92*, 164. [CrossRef]
16. d'Ambrosio Alfano, F.R.; Olesen, B.W.; Palella, B.I.; Povl, O.F. Impact Ten Years Later. *Energy Build.* **2017**, *152*, 243–249. [CrossRef]
17. Taleghani, M.; Tenpierik, M.; Kurvers, S.; van den Dobbelsteen, A. A review into thermal comfort in buildings. *Renew. Sustain. Energy Rev.* **2013**, *26*, 201–215. [CrossRef]
18. ASHRAE Standard. *Thermal Environmental Conditions for Human Occupancy*; ANSI/ASHRAE Standard 55-2004 (Supersedes ANSI/ASHRAE Standard 55-1992); American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.: Atlanta, GA, USA, 1791. Available online: http://www.ashrae.org (accessed on 15 February 2022).
19. Havenith, G.; Holmér, I.; Parsons, K. Personal factors in thermal comfort assessment: Clothing properties and metabolic heat production. *Energy Build.* **2002**, *34*, 581–591. [CrossRef]
20. Lin, T.-P. Thermal perception, adaptation and attendance in a public square in hot and humid regions. *Build. Environ.* **2009**, *44*, 2017–2026. [CrossRef]
21. Ng, E.Y.Y.; Cheng, V. Urban human thermal comfort in hot and humid Hong Kong. *Energy Build.* **2012**, *55*, 51–65. [CrossRef]
22. Kántor, N.; Égerházi, L.A.; Unger, J. Subjective estimation of thermal environment in recreational urban spaces—Part 1: Investigations in Szeged, Hungary. *Int. J. Biometeorol.* **2012**, *56*, 1075–1088. [CrossRef]
23. Krüger, E.; Drach, P.; Emmanuel, R.; Corbella, O. Assessment of daytime outdoor comfort levels in and outside the urban area of Glasgow, UK. *Int. J. Biometeorol.* **2012**, *57*, 521–533. [CrossRef]
24. Höppe, P. Different aspects of assessing indoor and outdoor thermal comfort. *Energy Build.* **2002**, *34*, 661–665. [CrossRef]
25. Fiala, D.; Havenith, G.; Bröde, P.; Kampmann, B.; Jendritzky, G. UTCI-Fiala multi-node model of human heat transfer and temperature regulation. *Int. J. Biometeorol.* **2011**, *56*, 429–441. [CrossRef] [PubMed]

26. Nikolopoulou, M.; Baker, N.; Steemers, K. Thermal comfort in outdoor urban spaces: Understanding the human parameter. *Sol. Energy* **2001**, *70*, 227–235. [CrossRef]

27. Knez, I.; Thorsson, S.; Eliasson, I.; Lindberg, F. Psychological mechanisms in outdoor place and weather assessment: Towards a conceptual model. *Int. J. Biometeorol.* **2008**, *53*, 101–111. [CrossRef] [PubMed]

28. Böcker, L.; Dijst, M.; Faber, J. Weather, transport mode choices and emotional travel experiences. *Transp. Res. Part A Policy Pract.* **2016**, *94*, 360–373. [CrossRef]

29. Eliasson, I.; Knez, I.; Westerberg, U.; Thorsson, S.; Lindberg, F. Climate and behaviour in a Nordic city. *Landsc. Urban Plan.* **2007**, *82*, 72–84. [CrossRef]

30. Klemm, W.; Heusinkveld, B.G.; Lenzholzer, S.; van Hove, B. Street greenery and its physical and psychological impact on thermal comfort. *Landsc. Urban Plan.* **2015**, *138*, 87–98. [CrossRef]

31. Mahmoud, A. Analysis of the microclimatic and human comfort conditions in an urban park in hot and arid regions. *Build. Environ.* **2011**, *46*, 2641–2656. [CrossRef]

32. Zachariáš, J.; Stathopoulos, T.; Wu, H. Microclimate and Downtown Open Space Activity. *Environ. Behav.* **2001**, *33*, 296–315. [CrossRef]

33. Nikolopoulou, M.; Steemers, K. Thermal comfort and psychological adaptation as a guide for designing urban spaces. *Energy Build.* **2003**, *35*, 95–101. [CrossRef]

34. De Luca, F.; Naboni, E.; Lobaccaro, G. Tall buildings cluster form rationalization in a Nordic climate by factoring in indoor-outdoor comfort and energy. *Energy Build.* **2021**, *238*, 110831. [CrossRef]

35. Attaianese, E.; Alfano, F.R.D.; Palella, B.I.; Pepe, D.; Vanacore, R. An Integrated Methodology of Subjective Investigation for a Sustainable Indoor Built Environment. The Case Study of a University Campus in Italy. *Atmosphere* **2021**, *12*, 1272. [CrossRef]

36. Nazarian, N.; Sin, T.; Norford, L. Numerical modeling of outdoor thermal comfort in 3D. *Urban Clim.* **2018**, *26*, 212–230. [CrossRef]

37. Wania, A.; Bruse, M.; Blond, N.; Weber, C. Analysing the influence of different street vegetation on traffic-induced particle dispersion using microscale simulations. *J. Environ. Manag.* **2012**, *94*, 91–101. [CrossRef] [PubMed]

38. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [CrossRef]

39. De Luca, F. Outdoor Comfort Analysis in a University Campus during the Warm Season and Parametric Design of Mitigation Strategies for Resilient Urban Environments. Computer-Aided Architectural Design, "Design Imperatives: The Future is Now". In Proceedings of the 19th International Conference CAAD Futures 2021, University of Southern California, Los Angeles, CA, USA, 16–18 July 2021.

40. Available online: https://www.wunderground.com/history/monthly/ee/tallinn/EETN/date/2021-6 (accessed on 7 June 2021).

41. Taltech, Nzeb. Available online: https://taltech.ee/en/nzeb-technological-test-facility (accessed on 15 February 2022).

42. Chen, Y.-C.; Lin, T.-P.; Matzarakis, A. Comparison of mean radiant temperature from field experiment and modelling: A case study in Freiburg, Germany. *Arch. Meteorol. Geophys. Bioclimatol. Ser. B* **2014**, *118*, 535–551. [CrossRef]

43. Mills, J.; Birks, M. *Qualitative Methodology: A Practical Guide*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2014; ISBN 9781446248980. [CrossRef]

44. Available online: https://www.wunderground.com/history/monthly/ee/tallinn/EETN/date/2021-6-7 (accessed on 15 February 2022).

45. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101, ISSN 1478-0887. [CrossRef]

46. Berkovic, S.; Yezioro, A.; Bitan, A. Study of thermal comfort in courtyards in a hot arid climate. *Sol. Energy* **2012**, *86*, 1173–1186. [CrossRef]

47. Gao, N.; Shao, W.; Rahaman, M.S.; Zhai, J.; David, K.; Salim, F.D. Transfer learning for thermal comfort prediction in multiple cities. *Build. Environ.* **2021**, *195*, 107725. [CrossRef]

48. Mehta, H.; Kanani, P.; Lande, P. Google Maps. *Int. J. Comput. Appl.* **2019**, *178*, 41–46. [CrossRef]

49. Blazejczyk, K. New climatological-and-physiological model of the human heat balance outdoor (MENEX) and its applications in bioclimatological studies in different scales. *Zesz. IgiPZ PAN* **1994**, *28*, 27–58.

50. Available online: https://www.engineeringtoolbox.com/clo-clothing-thermal-insulation-d_732.html (accessed on 16 November 2021).

51. Rawal, R.; Manu, S.; Shukla, Y.; Thomas, L.E.; de Dear, R. Clothing insulation as a behavioural adaptation for thermal comfort in Indian office buildings. In Proceedings of the 9th Windsor Conference: Making Comfort Relevant, Windsor, UK, 7–10 April 2016; pp. 403–415.

52. Alfano, F.R.D.; Palella, B.I.; Riccio, G.; Malchaire, J. On the Effect of Thermophysical Properties of Clothing on the Heat Strain Predicted by PHS Model. *Ann. Occup. Hyg.* **2015**, *60*, 231–251. [CrossRef] [PubMed]

Curriculum vitae

Personal data

Name:	Nasim Eslamirad
Date of birth:	29.06.1980
Place of birth:	Tabriz, Iran
Citizenship:	Iranian

Contact data

E-mail:	nasim.eslamirad@taltech.ee, naesla@ttu.ee

Education

2020–2023	PhD, Tallinn University of Technology
2014–2016	MSC, Imam Khomeini University, Qazvin, Iran
2011–2014	BSC, Azad University, Tehran, Iran
1998–2002	BSC, University of Tabriz, Tabriz, Iran
1993–1997	High school, Iran

Language competence

English	Fluent
Persian	Mother tongue, Fluent
Azerbaijanian	Mother tongue, Fluent
Turkish	Fluent
Arabic	Intermediate
German	Intermediate

Professional employment

2020–2024	Employment Researcher, Smart City Center of Excellence (Finest Twins), Tallinn University of Technology, Tallinn, Estonia
2017–2019	Membership Researcher, High-Performance Architecture Lab (HAL), Tarbiat Modares University, Tehran, Iran
2017–2018	Employment Designer in Rochi Design Idea Co, Tehran, Iran
2013–2017	Employment Interior designer, Supervisor of construction, Tehran, Iran
2003–2009	Employment Food Science Engineer, Iran

Peer-review journal articles

1. N. Eslamirad, A. Sepu´lveda, F. De Luca, K. Sakari Lylykangas, and S. Ben Yahia "Outdoor Thermal Comfort Optimization in a Cold Climate to Mitigate the Level of Urban Heat Island in an Urban Area", Energies, 16(12), 4546, (2023).
2. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, S. Ben Yahia, M. Rasoulinezhad "Geoprocess of Geospatial Urban Data in Tallinn, Estonia", Data in Brief, 48, 109172, (2023).
3. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, S. Ben Yahia, "Data Generative Machine Learning Model for the Assessment of Outdoor Thermal and Wind Comfort in a Northern Urban Environment", Frontiers of Architectural Research, 12 (3), 541-555, (2023).
4. N. Eslamirad, A. Sepu´lveda, F. De Luca, K. Sakari Lylykangas, "Evaluating Outdoor Thermal Comfort Using a Mixed-Method to Improve the Environmental Quality of a University Campus", Energies. 15, 4, 1577, (2021).

5. N. Eslamirad, S. Malekpour Kolbadinejad, M. Mahadavinejad, M. Mehranrad, "Thermal Comfort Prediction by Applying Supervised Machine Learning in Green Sidewalks of Tehran", Smart and Sustainable Built Environment, 9 (4), (2020).

6. M. Mahadavinejad, M. Mehranrad, N. Eslamirad, "Evaluation and Comparison of LEED, BREEAM, and the 19th Issue of National Building Regulations of Iran (NBRI) in Perspective of Sustainability by MCDM TOPSIS Method", Journal of the Structural Engineering and Geotechnics (JSEG), 9 (2019).

Conference papers

1. N. Eslamirad, F. De Luca, A. Sepu´lveda, K. Sakari Lylykangas, "Methodology for Improving Wind Comfort in a Cold Region Through Modular Urban elements", eCAADIA 2023 Conference, Tallinn, Estonia, (2023).

2. N. Eslamirad, F. De Luca, S. Ben Yahia, K. Sakari Lylykangas, "From Near Real-time Urban Data to Explainable City-scale Model to Help Reducing the Urban heat island (UHI) Effect", CISBAT 2023 Conference, (2023).

3. A. Sepu´lveda, N. Eslamirad, F. De Luca, "Multi-performance Method for Urban Densifications", Building Simulation, BS2023 conference, Shanghai, China, (2023).

4. A. Sepu´lveda, N. Eslamirad, F. De Luca, "Machine Learning-based Optimization Design Workflow Based on Obstruction Angels for Building Facades", eCAADIA 2023 Conference, Tallinn, Estonia, (2023).

5. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, "Building Morphology Influence on Outdoor Comfort in Urban Environments. An Analysis through Machine Learning", CISBAT Conference, (2021).

6. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, "Applying Machine Learning in Urban Scale to Predict Comfort Quality, Case Study Tallinn, Estonia", AI in AEC Conference, (2021).

7. N. Eslamirad, M. Mahadavinejad, S. Dehghani Tafti, "The Impact of Green Sidewalks in Thermal Comfort and Air Pollution", Symposium on Simulation for Architecture and Urban Design (SimAUD), (2021).

8. H. Zarrinkafsh, N. Eslamirad, F. De Luca, "Concentrated Solar Power (CSP) for Sustainable Architecture to Supply Domestic Hot Water and Heating Loads of buildings", CISBAT Conference, (2021).

9. H. Zarrinkafsh, N. Eslamirad, F. De Luca, "Classification of the Energy Production Potential of Water Lens Solar Concentrators Using Machine Learning", CISBAT Conference, (2021).

10. N. Eslamirad, M. Mahadavinejad "External Shadings Effect on Operating Energy Based on LCEA, Case Study: A Residential Building in Tehran", SimBuild, by ASHRAE and IBPSA-USA, Chicago, USA. ASHRAE, (2018).

11. N. Eslamirad, M. Mahadavinejad "Multi Objective Computing and Applying Expert System in Double Skin Facade system", Energy '18: The Ninth International Conference ACM, Karlsruhe, Germany, (2018).

12. N. Eslamirad, Y. Gorji Mahlabani "Shading Effects on Reducing Annual Cooling Loads in Tehran Climate", 15th International Conference in Science and Technology, Lon don, UK, (2016).

13. N. Eslamirad, A. Sanei, "Double Skin Facades in Use, A Study of Configuration and Performance of Double Skin Fa¸cade, Case Studies Some Office Buildings", 2th International Conference on Research in Science and Technology, Istanbul, Turkey, (2015).

Elulookirjeldus

Isikuandmed

Nimi:	Nasim Eslamirad
Sünniaeg:	29.06.1980
Sünnikoht:	Tabriz, Iraan
Kodakondsus:	Iraani

Kontaktandmed

E-post:	nasim.eslamirad@taltech.ee, naesla@ttu.ee

Hariduskäik

2020–2023	Tallinna Tehnikaülikool, doktoriõpe
2014–2016	MSC, Imam Khomeini Ülikool, Qazvin, Iraan
2011–2014	BSC, Azadi Ülikool, Teheran, Iraan
1998–2002	BSC, Tabrizi Ülikool, Tabriz, Iraan
1993–1997	Keskkool, Iraan

Keelteoskus

Inglise keel	valdamine
Pärsia keel	emakeel, valdamine
Aserbaidžaani keel	emakeel, valdamine
Türgi keel	valdamine
Araabia keel	kesktase
Saksa keel	kesktase

Teenistuskäik

2020–2024	Töötaja-uurija, Targa Linna Tippkeskus (Finest Twins), Tallinna Tehnikaülikool, Tallinn, Eesti
2017–2019	Liikmesuurija, Kõrge Jõudlusega Arhitektuurilabor (HAL), Tarbiat Modaresi Ülikool, Teheran, Iraan
2017–2018	Töötaja Disainer Roçhi Disainiidee OÜ, Teheran, Iraan
2013–2017	Töötaja Sisekujundaja, Ehituse Järelevalve, Teheran, Iraan
2003–2009	Töötaja Toitumisteaduse insener, Iraan

Teadusajakirjade artiklid

1. N. Eslamirad, A. Sepu´lveda, F. De Luca, K. Sakari Lylykangas, and S. Ben Yahia "Outdoor Thermal Comfort Optimization in a Cold Climate to Mitigate the Level of Urban Heat Island in an Urban Area", Energies, 16(12), 4546, (2023).
2. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, S. Ben Yahia, M. Rasoulinezhad "Geoprocess of Geospatial Urban Data in Tallinn, Estonia", Data in Brief, 48, 109172, (2023).
3. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, S. Ben Yahia, "Data Generative Machine Learning Model for the Assessment of Outdoor Thermal and Wind Comfort in a Northern Urban Environment", Frontiers of Architectural Research, 12 (3), 541-555, (2023).
4. N. Eslamirad, A. Sepu´lveda, F. De Luca, K. Sakari Lylykangas, "Evaluating Outdoor Thermal Comfort Using a Mixed-Method to Improve the Environmental Quality of a University Campus", Energies. 15, 4, 1577, (2021).

5. N. Eslamirad, S. Malekpour Kolbadinejad, M. Mahadavinejad, M. Mehranrad, "Thermal Comfort Prediction by Applying Supervised Machine Learning in Green Sidewalks of Tehran", Smart and Sustainable Built Environment, 9 (4), (2020).

6. M. Mahadavinejad, M. Mehranrad, N. Eslamirad, "Evaluation and Comparison of LEED, BREEAM, and the 19th Issue of National Building Regulations of Iran (NBRI) in Perspective of Sustainability by MCDM TOPSIS Method", Journal of the Structural Engineering and Geotechnics (JSEG), 9 (2019).

Konverentsiartiklid

1. N. Eslamirad, F. De Luca, A. Sepu´lveda, K. Sakari Lylykangas, "Methodology for Improving Wind Comfort in a Cold Region Through Modular Urban elements", eCAADIA 2023 Conference, Tallinn, Estonia, (2023).

2. N. Eslamirad, F. De Luca, S. Ben Yahia, K. Sakari Lylykangas, "From Near Real-time Urban Data to Explainable City-scale Model to Help Reducing the Urban heat island (UHI) Effect", CISBAT 2023 Conference, (2023).

3. A. Sepu´lveda, N. Eslamirad, F. De Luca, "Multi-performance Method for Urban Densifications", Building Simulation, BS2023 conference, Shanghai, China, (2023).

4. A. Sepu´lveda, N. Eslamirad, F. De Luca, "Machine Learning-based Optimization Design Workflow Based on Obstruction Angels for Building Facades", eCAADIA 2023 Conference, Tallinn, Estonia, (2023).

5. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, "Building Morphology Influence on Outdoor Comfort in Urban Environments. An Analysis through Machine Learning", CISBAT Conference, (2021).

6. N. Eslamirad, F. De Luca, K. Sakari Lylykangas, "Applying Machine Learning in Urban Scale to Predict Comfort Quality, Case Study Tallinn, Estonia", AI in AEC Conference, (2021).

7. N. Eslamirad, M. Mahadavinejad, S. Dehghani Tafti, "The Impact of Green Sidewalks in Thermal Comfort and Air Pollution", Symposium on Simulation for Architecture and Urban Design (SimAUD), (2021).

8. H. Zarrinkafsh, N. Eslamirad, F. De Luca, "Concentrated Solar Power (CSP) for Sustainable Architecture to Supply Domestic Hot Water and Heating Loads of buildings", CISBAT Conference, (2021).

9. H. Zarrinkafsh, N. Eslamirad, F. De Luca, "Classification of the Energy Production Potential of Water Lens Solar Concentrators Using Machine Learning", CISBAT Conference, (2021).

10. N. Eslamirad, M. Mahadavinejad "External Shadings Effect on Operating Energy Based on LCEA, Case Study: A Residential Building in Tehran", SimBuild, by ASHRAE and IBPSA-USA, Chicago, USA. ASHRAE, (2018).

11. N. Eslamirad, M. Mahadavinejad "Multi Objective Computing and Applying Expert System in Double Skin Facade system", Energy '18: The Ninth International Conference ACM, Karlsruhe, Germany, (2018).

12. N. Eslamirad, Y. Gorji Mahlabani "Shading Effects on Reducing Annual Cooling Loads in Tehran Climate", 15th International Conference in Science and Technology, Lon don, UK, (2016).

13. N. Eslamirad, A. Sanei, "Double Skin Facades in Use, A Study of Configuration and Performance of Double Skin Fa¸cade, Case Studies Some Office Buildings", 2th International Conference on Research in Science and Technology, Istanbul, Turkey, (2015).