AN URBAN MORPHOLOGY BASED METHOD FOR AIR TEMPERATURE ESTIMATION IN INFORMAL SETTLEMENTS.

SALLY AMA SAMPSON

JULY 2024

SUPERVISORS:
Dr. M.KUFFER
Dr. J WANG

AN URBAN MORPHOLOGY BASED METHOD FOR AIR TEMPERATURE ESTIMATION IN INFORMAL SETTLEMENTS.

SALLY AMA SAMPSON
Enschede, The Netherlands, July 2024.

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:
Dr. M. Kuffer
Dr. J. Wang

THESIS ASSESSMENT BOARD:
Dr. M. Koeva (Chair)
D. S. Vanhyusse (External Examiner, Universite Libre de Brussels)

ABSTRACT

The increasing impacts of climate change as well have reinforced concerns about heat exposure, especially in urban areas where temperatures are exaggerated due to the urban heat island effect. Specifically, informal settlements, within urban areas are attributed with these higher temperatures as compared with surrounding formal areas. Several studies have attempted to model and characterise temperature variations using both Land Surface Temperature and Air temperature and have attributed variations to differences in the urban form and fabric such as morphology and green infrastructure. Other studies have established that informal settlements can and often exhibit different urban characters including morphology. They have been characterised by dense, compact structures and limited vegetation. Despite these gaps and the understanding of the relevance of urban morphology to temperature, none of these studies have extensively included detailed morphological parameters in modelling air temperature.

This study addresses the limitations of current methods by leveraging urban morphometrics and random forest to estimate air temperature within informal settlements, with a specific focus on Nairobi, Kenya. It explores the relationship between urban morphology and air temperature, hypothesizing that distinct morphological characteristics of informal settlements significantly influence local temperature variations. Through a detailed morphological analysis and advanced machine learning techniques, the study models air temperature with high spatial resolution, providing insights into the patterns and drivers of thermal variations.

From this study, it was established that there are strong morphological differences between formal and informal areas in the study area. The random forest model successfully predicted air temperature with an R^2 of 0.73, revealing significant warmer temperatures within informal settlements. The model also allowed the assessment of importance of detailed morphology to air temperature. This analysis revealed urban morphology to only play a supplementary role in air temperature prediction. Urban morphometrics only influence temperature in a confounding manner, thus only at a contextual level or neighbourhood level, does it impact these variations.

Keywords: urban morphology, Informal settlements, air temperature estimation, random forest regression

ACKNOWLEDGEMENTS

This thesis phase has been full of both personal and academic enrichment and in as much as it was a tough journey, it was also a fulfilling and enjoyable one. I would like to first and foremost thank God, for seeing me through my master's programme. For the overwhelming peace only, He could provide during the tough nights when my codes were not running, and it felt like my life was over.

I am grateful for the support of my family. I hadn't expected to miss them this dearly when I left Ghana to pursue this master. Through it all, they have tried their best to be by me, calling to check in on me almost every day. To father particularly, thank you for always pulling through when I needed more funds. Even though you would always say 'You are a big girl now, figure it out', you always came through. To my mum; your prayers and encouragement always felt like hug, when I needed it the most.

My sincere appreciation goes to my supervisors Dr Monika Kuffer and Jon Wang for your mentorship. I don't think it would have been possible for me without your guidance. Monika, thank you for your efforts in helping my progress not only with my thesis but with my internship and future career. Jon, I truly am appreciative of the constructive criticism and technical support. Thank you, for your belief in me. There were moments I did not think I had the capacity to employ some technique. But your surety in believing I could do it, gave me the confidence to.

To the entire Onekana team, Angela Abascal, Sabine Vanhuysse, and Stefanos Georganos. I thank you for welcoming me on board and treating me as an equal. Even going beyond and above to help with insights not only to internship but my thesis. I am eternally grateful.

Finally, to the family and support system I have built here in the Netherlands, specifically Ama Serwah, Babu and Abdul, special thanks to you for constantly showing me that family is not just blood.

TABLE OF CONTENTS

Contents

List of figures v
List of tables ....
vii v
1. Introduction 1
1.1. Background and Justification 1
1.2. Problem Statement 2
1.3. Research Objectives and Questions 3
1.4. Conceptual Framework 3
2. Literature Review 5
2.1. Heat and Urbanisation 5
2.2. Heat Exposure and Vulnerability in Informal Settlements 5
2.3. Conceptualizing Informal Settlements in Temperature Research 6
2.4. Influence of Urban Morphology on Local Temperature 6
2.5. Current Advances in Air Temperature Estimation 7
3. Methodology 9
3.1. Study Area 10
3.2. Research Methodology 11
3.3. Morphometric Computation 12
3.3.1. Data Preprocessing for Morphometric Computation 13
3.3.2. Morphometrics Computation Hyperparameters 13
3.4. Clustering Analysis 14
3.5. Insitu Air Temperature Measurement 15
3.5.1. Data Collection Methods 15
3.5.2. Insitu Air Temperature Preprocessing 16
3.6. Auxiliary Variables Computation 17
3.6.1. Sentinel 2 Image Preprocessing and Land cover computation 17
3.6.2. Land Surface Temperature Preprocessing 18
3.6.3. Characterising Urban Greenery 18
3.7. Scale of Analysis and Data Aggregation 19
3.8. Modelling Air Temperature-Morphology Association Using Random Forest 19
3.8.1. Random Forest Model Implementation and Optimisation 20
3.8.2. Model Accuracy Measures 22
3.8.3. Citizen Science Model Validation 22
4. RESULTS 23
4.1. Urban Morphometrics Results 23
4.1.1. Assessing Urban morphological Clusters in The Study Area 23
4.1.2. Characterising Emerged Clusters 25
4.2. Analysing Data Descriptives 27
4.2.1. Correlation Between Covariates and Air temperature 27
4.2.2. Determining Feature Selection using Principal Component Analysis 28
4.3. Temperature - Morphology Relationship Modelling 28

4.3.1. Determining best hyperparameters for tuning the model 28
4.3.2. Important predictors for temperature across all models 30
4.3.3. Sensitivity Analysis: Analysing feature influences 31
4.3.4. Analysing relationships between Morphology and predicted temperature 32
4.4. Spatio-Thermal Dynamics of Predicted Temperature 33
4.5. Citizen science Validation of Predicted Temperature Maps 36
5. Discussions 37
5.1. Applications of Proposed Model 38
5.2. Air Temperature Variability at The Local Scale 38
5.3. Important Influences to Temperature Variation 41
5.4. Limitations and recommendations 42
5.5. Uncertainties from citizen science data 43
5.6. Random Forest Model Limitations 44
6. Conclusion 44
6.1. Ethical Considerations 45
List of references ....
49 46
7. APPENDIX 51

LIST OF FIGURES

Figure 1: Conceptual Framework vi
Figure 2:Study Area Map 10
Figure 3:Flowchart 12
Figure 4: K-means algorithm process; adapted from Han et al., (2012) 14
Figure 5: Community Engagement and Data Collection 16
Figure 6:Workflow of field data preprocessing 16
Figure 7: Spatial unit of analysis 19
Figure 8: Random forest algorithm;. 20
Figure 9:Data splitting: 80% training and hyperparameter testing and 30 for model validation 21
Figure 10:Validation and training grids 21
Figure 11:Model Implementation 22
Figure 12:Morphological clusters, k=2 24
Figure 13:Kmeans evaluation with reference data 25
Figure 14: K-means Important features for cluster separation; k=2 26
Figure 15: Statistically significant covariates correlation to temperature. 27
Figure 16: Selecting best set of hyperparameters 28
Figure 17: Training and Testing Accuracies Model 1 29
Figure 18: Testing Accuracies Models 2 and 3 30
Figure 19: Feature importance of global model 31
Figure 20:Feature relationship with predicted temperature Model 1. 33
Figure 21: Predicted Air temperature 34
Figure 22: Boxplot of Predicted Air temperature 35
Figure 23:Gated residential NonDUA in Waruku 39
Figure 24:Industrial NonDUA in Mukuru 39
Figure 25: Less deprived within in informal area Kibera 39
Figure 26: Derived cluster description 40

LIST OF TABLES

Table 1:Literature Review on Existing Methods 8
Table 2: Morphometric hyper parameter tuning 13
Table 3: Acquisition dates per settlement 15
Table 4: Sensor Specifications 15
Table 5: Land cover metrics computation 18
Table 6:Selected set of hyperparameters for model training 29
Table 7:Sensitivity Analysis 32
Table 8: Important features across morphological distinction and Temperature prediction 35

1. INTRODUCTION

Climate change and associated global temperature rises have resulted in growing concerns about heat exposure. This concern is especially essential in urban areas due to the urban heat island effect, which reveal significant temperature variations between and within urban and rural areas. Different types of urban areas show very contrasting temperatures. In particular informal areas tend to have higher temperatures compared to neighbouring formal areas (Mehrotra et al., 2018). Current methods exploring these thermal variations focus both on air temperature and Land Surface Temperature. The latter, measures the temperature of the earth's surface and may not necessarily be what people perceive. Traditional air temperature, on the hand, measured 2m above the earth surface, is measured by meteorological stations and may better reflect perceived temperatures. These stations are however, often sparsely distributed at irregular intervals thus making it difficult to explore local scale variabilities of air temperature.

Meanwhile, recent efforts on machine learning propose new ways to address current data and methodological gaps in exploring these thermal variations. This research aims to provide insights into air temperature variabilities in informal settlements by leveraging machine learning.

1.1. Background and Justification

The latest IPCC report explains with certainty that extremely hot temperatures have increased both in frequency and intensity since the 1950s (IPCC, 2023). The report further describes extreme temperature as one of the most fatal climate change hazards that adversely impacts human health and livelihoods. Liu et al., (2017) account tens of thousands of heat related deaths globally within the last decade, including the death toll of over 30,000 people during the 2003 heatwave in Europe as reported by the United Nations Environment Programme, (2003). Kjellstrom et al., (2018), on the other hand, explain the economic impacts of heat exposure by revealing that increasing heat exposure results in heat stress of workers and thus reduced productivity, accounting for substantial losses in work hours annually.

Urban areas are particularly exposed to these temperature hazards due to the urban heat island (UHI) phenomenon which refers to higher temperatures in urban areas compared to surrounding rural areas. These hazards are however not experienced similarly within urban areas (Kisters et al., 2022), as several studies reveal significant temperature variabilities, indicating that informal areas, are exposed to higher heat exposure (Scott et al., 2017; Egondi et al., (2012).

Informal settlements, also referred to as deprived urban areas or slums (Abascal et al., 2022), thought varied across different contexts, are often predominantly inhabited by the urban poor, with lower socio-economic conditions and therefore lack the adaptive capacities to deal with the impacts of heat exposure and heat stress. With higher recorded incidence of heat-related health risks and mortality, these areas are more vulnerable as they bear a higher compounding burden to the impacts of this climate stressor (Nag et al., 2009; Rathi et al., 2017).

These temperature differences have been associated with the densification and reduced vegetative cover as urban surfaces, such as concrete and asphalts etc, typically exhibit low albedos and thus absorb more heat than they reflect resulting in the warming (Ochola et al., 2020). Kotharkar et al., (2023); and Peng et al., (2022), share similar results while investigated this notion and found morphology to be an influencing factor of temperature differences in the urban space. This intuitively presupposes that different configuration of the urban fabric experience temperature, and its impacts differently. Informal settlements are often characterised by distinct morphologies, described by Abascal et al., (2022) as compact dense settlements with limited vegetative cover. Given that these are same urban characters associated with heat absorption, this could explain the significant temperature differences between formal and informal areas.

Much of the studies on thermal variations have been explored using Land Surface Temperature (LST) measurements from various satellite imagery. Though LST is useful in understanding surface heating and may provide foundational insights on temperature, it does not directly reflect the thermal comforts that people experience. To provide a more accurate understanding of heat exposure, air temperature offers a more precise representation of perceived heat (Gholamnia et al., n.d.). However, the sparse distribution of meteorological stations makes it difficult to assess local-scale temperature variations within urban areas, particularly in informal settlements.

Recent advancements in machine learning offers methods to address these data gaps. By integrating several data sources, machine learning can be used to predict air temperature at finer spatial resolution, allowing for detailed analysis of local temperature. Despite the opportunities such models provide in exploring thermal variations within the informal context, unfortunately, none of such studies have been employed in estimating air temperature in informal settlements. Moreover, urban morphology, which as previously established, contributes to temperature, especially at the local level, have not been extensively considered in such studies. This research, therefore, aims to leverage machine learning techniques to provide deeper insights into air temperature variabilities within informal settlements. By utilizing advanced modelling approaches, the study seeks to fill the gaps of traditional air temperature assessments to offer a more comprehensive understanding of thermal variations. The findings aim to contribute to the development of targeted interventions to mitigate heat exposure and improve living conditions for residents of informal settlements.

1.2. Problem Statement

Whilst previous studies have only focused on mapping out detailed patterns of Land Surface Temperature in informal settlements, e.g., Cao et al., (2022), Mehrotra et al., (2018), Ochola et al., (2020) and Scott et al., (2017), this study chooses air temperature over land surface temperature because it is directly related to human comfort as it represents the temperature that people perceive and experience (Zhang et al., 2016a). It is the temperature that influences daily activities and thermal comfort considerations. Increased air temperature is associated with heat stress and heat stress-related mortality and as such pose great risks to public health (Nascetti et al., 2022). Therefore, as this study concerns exposure to climate stressors in deprived communities, air temperature is more appropriate. In addition to its importance to exposure and public health, air temperature is one of the most used climatic variables in climate change studies as it plays an important role in most biological and physical processes, making it a significant predictor of terrestrial environmental conditions (Benali et al., 2012; Nascetti et al., 2022; Y. Z. Yang et al., 2017). Accurate spatiotemporal measurement of air temperature is tremendously beneficial and in high demand for environmental studies (Y. Z. Yang et al., 2017).

Traditional air temperature studies however do not extensively include morphology in modelling air temperature, and none are designed for the context of informal settlements, which often exhibit distinct morphologies. Thus, no studies exist that models air temperature in informal settlements to assess how different configurations and morphologies of informal settlements contribute to local urban heat, as such, the essence of this study. The study, therefore, fills two gaps;

- Measuring and exploring local scale patterns of air temperature within informal settlements.
- Extensively incorporating urban morphological characteristics in modelling air temperature.

This study thus proposes a morphologically based approach to modelling air temperature by employing morphometrics to leverage the detailed characterization of informal morphology.

Estimating air temperature by extensively including morphology of informal settlements is of great relevance in recent times of UN-Habitat and World Bank slum upgrading agenda. These initiatives aim to improve the living conditions of slum dwellers with an emphasis on participatory climate resilience and adaptation strategies. This intersection with SDG 11.1 target (ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums) and target 13.1 (Strengthen resilience and adaptive capacity to climate-related hazards, especially in vulnerable communities) makes this research of great relevance as it is imperative to assess the root causes of temperature variability to successfully develop targeted and effective interventions.

1.3. Research Objectives and Questions

The main objective of this study is to explore local scale temperature variations in informal settlements to assess the relationship that exist between daily air temperature (in peak summer season) and the urban morphology by leveraging machine learning methods. The hypothesis underlying this study is that there are air temperature variations within and across formal and informal settlements and these variations are related to and can be explained my morphological differences. In line with this, specifically, the sub-objectives and research questions are as follows.

1. To measure the urban morphological characteristics within and around informal settlements in Nairobi
    - Are there distinguishing morphological patterns observed within and between formal area?
    - What morphological characteristics are key in distinguish informal settlements from each other and formal areas and why?
2. To measure and analyse the patterns of air temperature within and between informal and formal areas.
    - What spatial patterns of air temperature are observed within informal settlements?
    - What spatial patterns of air temperature are observed across informal and formal settlements?

3. Modelling Air temperature from slum morphological characteristics, and auxiliary variables, employing random forest.
    - Based on literature, what covariates are significant to model air temperature?
    - What is the accuracy of the Air temperature model?
    - What are the key morphological characteristics that contribute significantly to the temperature variations?

1.4. Conceptual Framework

The underlying principle to local scale air temperature estimation in this study, is in imbedding extensively morphological aspects of the urban fabric of informal settlements. As morphology plays a crucial role in local microclimatic conditions (Stewart & Oke, 2012), and may provide great justifications to temperature variations. With the hypothesis that the physical morphology of an area plays a role in its air temperature, slum morphometrics is central to the air temperature model in explaining higher temperatures in informal settlements.

Though not a risk analysis, this study touches on both hazard and vulnerability. It provides a framework for estimating air temperature hazard intensity taking into consideration the physical conditions of informal settlements, often, a physical manifestation of their vulnerabilities (Isunju et al., 2016; Ren et al., n.d.) and thus beneficial to risk assessment.

The figure below provides a graphical representation of the framework, highlighting the focus study areas, data, and methods, that have been used in previous studies and the proposed approach (in green) employed in this study to address gaps identified. While current studies do not focus on informal settlements and do not extensively incorporate detailed description of morphology in air temperature estimation, this study seeks to account for these gaps.

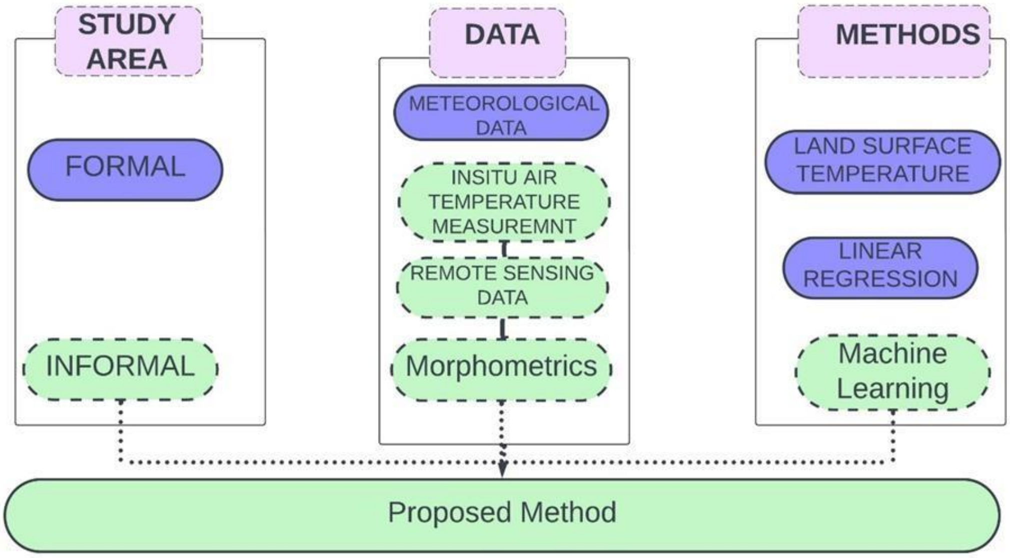


Figure 1: Conceptual Framework

2. LITERATURE REVIEW

2.1. Heat and Urbanisation

Much of the impacts of heat exposure are recorded in cities due to the urban heat island effect (Koppe et al., 2004). Urbanization has caused increasing impervious surfaces, which retain heat and exacerbate temperature burdens causing higher temperatures in urban areas as compared to their surrounding semiurban and rural areas. This is referred to as an UHI, which has been described by Ochola et al., (2020) as being one of the most extensive and significant signs of climate modification in the urban space.

The world's urban population is expected to increase from 55% to 70% by 2050 with most of this projected urbanization estimated to happen in cities in the global South, specifically in Sub-Saharan Africa (World Cities Report 2022, n.d.; United Nations, 2018). Again, heat extremes, are projected to increase both in intensity and frequency as the climate warms. Specifically in Africa, these extremes are projected to have increases in magnitudes and durations twice more relative to Europe (IPCC, 2023; Z. Liu et al., 2017; Scott et al., 2017). Considering these combined adverse projections, heat exposure and its effects will get worse in the coming years, as such, urban heat islands and its effects are a great cause of concern not only to urban planners but to policy makers, public health officials and urban disaster risk management community. This climate risk in Africa is thought more alarming especially because countries in the region tend to have inadequate resources to properly deal with and manage climate related hazards and risks (Scott et al., 2017).

2.2. Heat Exposure and Vulnerability in Informal Settlements

Exposure to heat is therefore not distributed proportionately globally. Equally, even within regions and nations, these hazard events do not occur uniformly across space as Morrow (2008) and Yoon (2012) explain hazards and impacts are shared disproportionally within nations, often with vulnerable groups experiencing higher exposure levels as they tend to reside in harmful conditions in high-risk locations and bearing a larger brunt of the economic and health impacts because they lack access to adaptive resources. In relation to extreme heat and heat exposure, several studies have been conducted that localizes heat related risks and imply that informal settlements are more at risk to heat. For example, Nag et al., (2009); Rathi et al., (2017) revealed higher temperatures, heat related illness and mortality in informal and slum areas in Indian cities. Similarly in the African context, Egondi et al., (2012; Scott et al., (2017) analysed temperature patterns in Nairobi, Kenya and observed higher temperatures and heat related illness and mortality in slums relative to neighbouring formal settlements. These studies, focused on characterizing risks by addressing heat vulnerability using demographic, and other socio-economic attributes.

However, since risk is defined as a function of both vulnerability and exposed population to hazard (Reisinger et al., n.d.), assessing hazard intensity and exposure cannot be neglected in risk assessment. Unfortunately, only few research analyses the heat exposure component of risk in relation to informal settlements. Wang et al., (2019) examines local scale temperature patterns in slum areas in Ahmedabad and revealed the exposure of slum areas to intense locally high temperatures. This analysis also exposed the size of slums as a contributing factor to temperature intensity, associating larger slums with higher local temperatures in comparison to smaller slum settlements. Similarly, Mehrotra et al., (2018); Ochola et al (2020); Ramsay et al., (2023); Scott et al., (2017); Wu et al., (2018) assessed the intrinsic patterns and variability of temperatures in informal settlements and slums within the urban landscape.

The methods and techniques used by these studies varied substantially with Wang et al., (2019) focusing on morphologically based exploration of thermal patterns using land surface temperature (hereon LST) and Scott et al., (2017) examining temperature variations between informal settlements with air temperature measurements obtained from installing 50 ground temperature sensors across strategic locations within the slums and comparing them to data from meteorological weather stations in Nairobi. Ochola et al., (2020) studied the variability of temperatures in Nairobi by characterizing the entire urban landscape (including informal settlements) into Local Climate zones (LCZ) and explored the variability of LST amongst these zones.

2.3. Conceptualizing Informal Settlements in Temperature Research

Several variations of informal settlements have been considered in thermal studies. This is largely due to the polysemic nature of the term, as what is 'informal' is widely contextual and very much dependent on existing formal spatial-governing structures. Thus, the definition of informal settlements is location specific and varies across countries. For example, whilst Mehrotra et al., (2018) studied all urban informal housing clusters within the urban landscape, Scott et al., (2017) considered informal settlements to be areas that have a majority share of buildings with low, dense housing types built from poor construction materials such as galvanized iron sheets, mud, clay, and wood and also lack accessibility to basic human services such as healthcare and clean water. Wu et al., (2018) on the other hand considered urban villages in China characterized by overcrowded areas with small compact substandard storey buildings situated in areas with little to no vegetation cover and improper sanitary conditions as a form of informal settlement.

Despite their varied definitions, all these informal settlements share commonalities, particularly in terms of lower socio-economic status, poverty, and deprivation. Building on this foundation of deprivation, Abascal et al. (2022) acknowledges the intersections among these areas and propose expressing their degrees of deprivation to ensure standardization across differing contexts. This expression, "deprived urban area" (DUA), encompasses all these characterizations to define poor areas and their residents. As such, from hereon; this term is used interchangeably with informal settlements.

This deprivation often manifests in the physical structures and is expressed in characterisation of the fundamental elements of the urban form such as streets, buildings, and plots termed morphology (Abascal et al., 2022; Kuffer et al., 2022; Oliveira et al., 2023).

2.4. Influence of Urban Morphology on Local Temperature

It is important to note that DUA's are not homogenous nor monolithic (Scott et al., 2017). Again, as seen in the works of Nag et al., (2009); Ochola et al., (2020); Rathi et al., (2017); Scott et al., (2017); Wang et al., (2019) and Wu et al., (2018), they have varied, heterogenous surfaces and characteristics and thus have different configurations that may influence local temperature variations. Again, as the morphological characteristics of informal settlements tend to be very distinct from formal surroundings, it suggests that, that morphology may be the main contributing factor to temperature variations between informal settlements and surrounding formal areas.

Scott et al., (2017) reiterates this by explaining that physical environment of a local area influences its microclimate and thus temperature and suggests that the internal characteristic of urban morphology is highly correlated to urban temperatures. In the same vein, numerous studies have explored the relationship between urban morphology and both land surface and air temperature. For example, Peng et al., (2022) found building density to be the most important variable explaining much of the variation in local thermal conditions which is in line with findings by Li et al., (2023a) that indicates an increase in air temperatures with increase in building coverage ratio up to 40 percent (similar to building density). This finding is explained by larger heat loss through radiation as concrete surface areas increase with corresponding increase in buildings. The findings also reveal that, when building coverage ratio increases above 40% up to 70%, the results is an increase in air temperature as a result of the shadow effect of denser buildings resulting in cooler local thermal conditions. Puche et al., (2023), shared the similar findings by associating higher building density with higher air temperatures locally, though its extent was not explored as was done in Li et al., (2023a). This study also explored building height factor as a much higher significant contributor to lower air temperatures than building density. Though these studies uncover the significance of building density to air temperature, Puche et al., (2023), finds that it does not influence LST distribution significantly however building height and the presence of vegetation are the two main factors contributing to LST decrease in the urban environment.

The study of local climate zones as proposed by Stewart & Oke, (2012) provide a link between urbanization and its impacts on local thermal characteristics and supports the claim of a strong relationship between the urban configuration and temperature. The studies classify several urban morphology characteristics and relate them to temperature values, finding significant variations in thermal conditions between different urban classes.

It is seen, therefore, from these numerous studies that morphology plays a significant role in micro-climatic conditions, especially at the local level, as such are important predictors to estimation of both LST and air temperature.

2.5. Current Advances in Air Temperature Estimation

Traditionally, air temperature is measured by meteorological stations that provide highly accurate estimates with high temporal resolution but limited by sparse distribution in space resulting in course spatial resolution. This point-scale measurement makes it difficult to analyse spatial variations of temperature continuously (Wu et al., 2018). To compensate for this shortcoming various attempts are made at interpolating station measurements using various geostatistical approaches such as kriging in order to spatially map out air temperature. These approaches may be effective in temperature estimates near the stations but have substantial errors with increase in distance from stations. This is especially the case in instances where there is sparse or inconsistent distribution of stations (Janatian et al., 2017; Y. Z. Yang et al., 2017). In the context of developing countries, these stations are substantially inadequate and in instances where available, are highly variable in spatial distribution and usually outside the scope of informal settlements (Janatian Nasime et al., 2017; Scott et al., 2017) therefore, such techniques will not provide accurate spatial estimates and are inappropriate for this case study.

Air temperature is an intricate phenomenon to model due to the complex relationships that exist between environmental factors that influence it (Gholamnia et al., n.d.). Surface heating and cooling processes are the main modulators of daily Tair cycle determined by the energy balance of the earth-atmosphere system through an interplay of absorption of incoming solar radiation, emission of infrared radiation and sensible and latent heat exchange between the earth surface and the atmosphere. This results in a high variability of spatio-temporal patterns of TAIR resulting from the complex variability of these environmental factors such as latitude, cloud cover, particulate matter, wind speed, time, thermal inertia, soil moisture, albedo and emissivity etc that control the energy balance of earth-atmosphere relationship (Benali et al., 2012; Janatian Nasime et al., 2017; Y. Z. Yang et al., 2017).

Numerous studies have been undertaken in an attempt to address these complexities associated to air temperature estimation. Several have leveraged the power of earth observation satellites to provide data on the reflected, absorbed, and transmitted solar energy from the sun's interactions with the earth surface. In this vein, remotely sensed land surface temperature (LST) has been investigated as a possible measure to improve the spatio-temporal accuracies of air temperature estimation as it is available on a regular basis and is spatially contiguous making it especially beneficial to areas with very sparse stations (Y. Z. Yang et al., 2017). This author defines LST as the 'skin' temperature of the earth surface and is a key indicator of the net surface energy balance driven by surface emissions. Benali et al., (2012) explains that though LST and air temperature are highly correlated, they mean different things physically, have different magnitudes and respond very differently to atmospheric conditions, thus, have a complex relationship. Several other factors influence the interplays and interactions between LST and air temperature. For example, vegetation plays a crucial role in the dynamics between LST and air temperature. Highly vegetated areas provide shade and can significantly reduce the amount of direct sunlight reaching the ground, thus a reduced LST in that area. Again, the large surface roughness and low albedo of vegetation promotes sensible heat distribution. Through evapotranspiration, vegetation contributes to larger latent heat fluxes influencing microclimatic conditions and thus air temperature (Benali et al., 2012; Gholamnia et al., n.d.).

Despite the complex relationship between LST and air temperature, many studies have estimated air temperature using LST successfully. Below is a table that shows an overview of such studies.

Table 1:Literature Review on Existing Methods

Author	Data	Source	Method	Result
(Yang et al., 2017)	LST, View Zenith Angle Clear Days ,View Times NDVI	MODIS AQUA 8-day average LST 2002–2016	multiple linear regression	T_{Max}: R_2 0.90 T_{Min}: R_2 0.94 T_{Mean}: R_2 0.94
(Nascetti et al., 2022)	LST NDVI	Landsat-8	Neural Network (NN) regression model	T_{mean}: R_2 0.87
(Benali et al., 2012)	LST, Latitude, Meteorological data Distance to coast and clearwater	MODIS TERRA	Linear Regression	MEF: 0.93
(Janatian Nasime et al., 2017)	LST, Altitude, Julien Day Latitude, NDVI, EVI	MODIS TERRA	Multiple Linear Regression	R^2 :0.91
(Li et al., 2023b)	Meteorological data ENVI-MET urban canopy simulated air temperature. Urban3D morphology	China meteorological data center Airborne laser scanning	Random Forest Regression	
(Gholamnia et al., n.d.)	LST, Land cover maps Meteorological data Elevation	INSAT3-D	Linear Regression	MAEs between [0.1, 2.9] ◦C
(Keung Tsin et al., 2020)	NDVI, Distance to water bodies, Sky view factor Distance to roads, Elevation,	Landsat 8 Transect walks	Linear Regression model	R^2: 0.69
(Zhang et al., 2016b)	Metrological data, LST, NDVI, Longitude, Latitude, Julien day, Solar Zenith Angle	MODIS	Linear Regression, Support Vector Machine, Neural Network	

Table 1 above provides and overview of studies that estimate air temperature using LST. For example, Benali et al., (2012) estimated average air temperature of Portugal over a 10-year period based on MODIS LST data utilizing several simple linear models. Based on the results of the study, daytime LST explained 83.3% of variability while nighttime LST explained 92% with an RMSE of 1.38 °C. Similarly, Zhang et al., (2016) recorded an average RMSDs of all models ranging from 1.81°C to 2.64°C with nighttime LST contributing to higher model accuracies. Though all the models performed relatively well, the methods and products (models) have some limitations which can further be improved. The majority of the models focused only on using liner regression models in Tair estimations, though as explained in sections above, there exists nonlinear relationships between LST and Tair. This sentiment is shared by both Yang et al., (2017) and Benali et al., (2012) by explaining the need for advanced non-linear algorithms in improving the performance of Tair especially given the variability and complexities of the variables. Furthermore, though many studies have stressed the importance of morphology in local air temperature, most models fail to include these characteristics as features. To the best of the authors knowledge, there only a handful of studies that include urban morphology in Tair modelling and none are extensive, and none focuses on informal settlements. Though Li et al., (2023c) introduces some urban morphology characteristics, the method itself; highly dependent on data from meteorological stations and the use of Envi-met simulation which (a paid software) raises questions on transferability of method. Concerning generalizability, however, Gholamnia et al., n.d.; Nascetti et al., (2022); and Yang et al., (2017) express concerns on the variability of results when applied on different settings indicating the influential role of spatial context and land cover classes in Tair modelling. This implies, since no studies exist that models air temperatures in informal settlements despite their distinct morphological characteristics, this research is imperative.

According to Yu et al., (2020) by evaluating the complex relationships between the input and the target variables, feature importance machine learning algorithms such as random forest evaluate the significance of each feature to the target variable. By estimating the total decrease in impurity of each feature across all trees, random forest algorithm is able evaluate the most contributing features to model prediction. Thus, features with higher 'importance scores' were frequently used in splits contributing to large reductions in impurity and thus more significant to model prediction.

Given that this study seeks to understand the relationship between temperature and morphometric characters, Random Forest was used to tease out the most significant influencers of thermal variability. The RF algorithm has become widely used in recent times in the study of air temperature prediction due to its interpretability in relation to input variables (Li et al., 2023a; D. Yang et al., 2023). Besides its interpretability, this model was selected for the study, as, according to He et al., (2022) this algorithm is relatively fast to train, has a high precision, offers strong generalization abilities and is best-known application for evaluating the impact of each feature on model performance as such revealing important predictors within a dataset.

3. METHODOLOGY

This chapter provides a comprehensive overview of the methodology employed to achieve the results of the study. Given the case study approach of this research, the chapter begins with a general description of the city and then narrows down to the specific informal settlements selected for the study. Following this, a general overview of the research methodology is presented, accompanied by a flowchart illustrating the processes. After which specific details of the research methodology and methods ensues and concludes the chapter.

3.1. Study Area

This study area is Nairobi, the capital city of Kenya located in the southwestern part of the country at approximately 1.2921° S, 36.8219° E, and covering a total land mass of 696.1km2. Sitting at an altitude of approximately 1,795 metres (5,889 ft) above sea level, the city has a subtropical highland climate under the Koppen climate classification with two major rainfall seasons in mid-March – May and October – December. Nairobi experiences its warmest months between January and March with mean max temperatures between approximately 25.8°C and 26.7°C and coolest months between June and July with mean min temperatures of approximately 12 °C (NCCG, 2020). According to the recent climate action report of Nairobi County however, studies of historical meteorological data reveals that the city has been experiencing significant increasing trends in temperatures. Figure 1 below presents the study area map as well as some physical characteristics of Nairobi in Appendix 1.

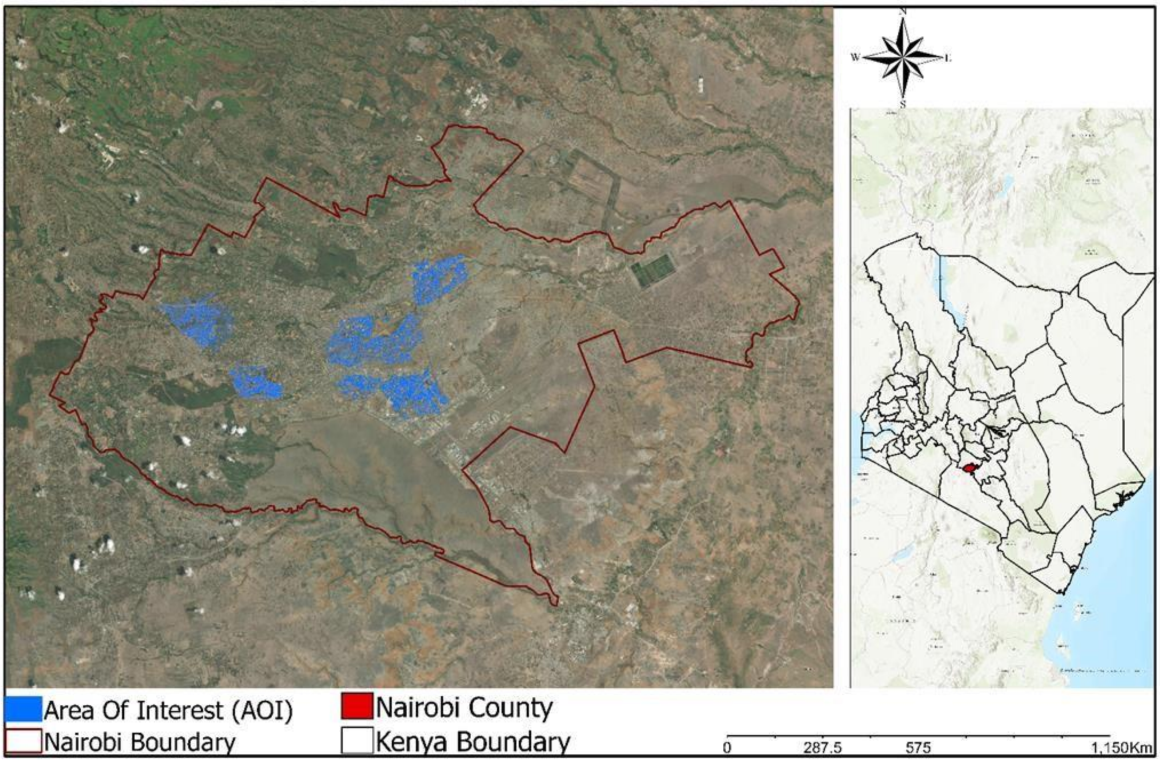


Figure 2:Study Area Map

As one of the fastest growing cities in Africa and a major regional economic hub in East Africa, Nairobi is home to about 4,397,073 people according to the country's population census of 2019 and is projected to be increase to 6 million people by 2030 (NCCG, 2020). Due to this rapid population growth however, the city is not able to meet the high demand for land which has resulted in an incessant expansion of informal settlements in the city. It is estimated that though informal settlements occupy less than 10% of the total land mass of Nairobi, roughly 60% of the population resides in informal settlements (NCCG, 2020; UNHABITAT, 2005). Hosting numerous informal settlements including Kibera; the largest informal settlement in Africa, the Nairobi County government acknowledges the vulnerabilities these settlements face especially in the face of climate change. Specifically referencing multiple studies that indicate that these informal settlements suffer from multiple climate stressors, including higher LST and air temperatures and heat stress related deaths compared to formal settlements. As a result, the city's government is interested in reducing the impacts of climate stressors on these settlements.

This study scope is limited to 5 informal settlements in Nairobi; Kibera, Mukuru, Koriogocho, Waruku, and Pumwani with boundaries defined by taking convex hulls of all insitu air temperature routes(transects). These settlements together formed the area of interest pictured in figure 2 above. See appendix 14 for aerial views.

3.2. Research Methodology

The general flow of this research was structured in four phases, reflecting the objectives of the study. Descriptions of all datasets used in the study are summarised in Appendix 2

The first phase, (shown in yellow frame from figure 6 below), address's objective1 and constitutes the morphometric computation. This phase starts with the preprocessing of building footprint dataset, which is then used in the computation of morphometrics. Following this, K-means clustering is employed to morphologically distinguish between DUA and non DUA's. Statistical tests are finally employed to tease out the important features. Phase two, (shown in blue frame from figure 6 below) incorporates the measurement and assessment of air temperature data. It involves the preprocessing and exploratory spatial data analysis of collected insitu air temperature measurements. The third phase (shown in purple frame from figure 6 below), involves the preprocessing and computation of all covariates; NDVI, NDWI, NDBI, LST and elevation. The final phase, (shown in purple frame from figure 6 below), employ the main outputs of all previous phases; insitu measurements, computed morphometrics, and covariates in building a random forest model of air temperature. A graphical representation of these phases is shown in the flowchart below.

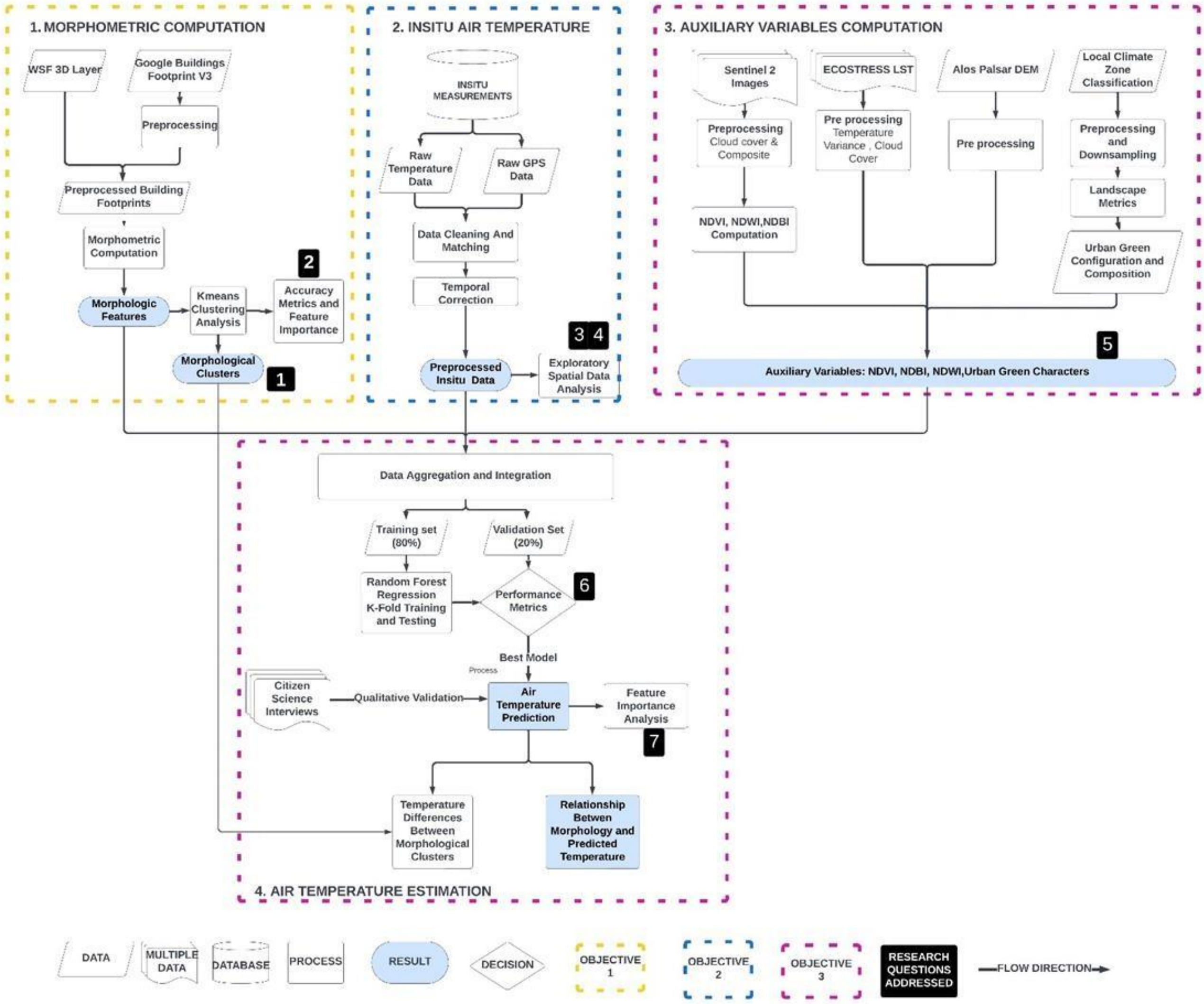


Figure 3:Flowchart

3.3. Morphometric Computation

The method of measuring morphology as applied in this study adopts morphometrics approach by Fleischmann et al., (2022) which provides an extensive method of detailing out numerical characteristics of the urban form. The method describes three fundamental elements of urban form: buildings, streets, and plots. However, since informal settlements tend to lack clearly defined streets, and thus lack of street configuration datasets, only buildings and plots are considered morphological elements in this study. For buildings, the height is also considered to capture the full extent of the third dimensional morphology. Given the polysemic nature of plots, morphological tessellations; a Voronoi tessellation based spatial unit derived from building footprints is adopted, following the work of (Fleischmann et al., 2022; J. Wang et al., 2023). The tessellation provides a means to describe the morphological influence each building exerts in its immediate spatial context.

Based on these elements, a selected group of 31 morphometric characters are implemented from the comprehensive list of 71 characters in the appendix 3 provides a defined list of morphometric characters used in the study as well as the theoretical foundations for computation.

A key principle of morphometrics supposes an identifiable relationship between the morphometric characters. As such, each character is measured at different scales characterized by the topological relations that exist between the elements. Thus, each character is measured at two scales: the individual element itself (small), and the element in the context of neighbours within a specified topological step (Large). These primary characters, measure the individual morphometric elements but fail to detail spatial patterns that exist in the study area. To account for this, each primary character is defined by its tendency in its context which is an aggregation of morphological cells within three topological steps of the element. This is done by employing four spatially lagged contextual characters; Interquartile mean to capture the local tendency, Interquartile-range, Interdecile Theil index (IDT), and Simpsons diversity index(SDI) to capture the distribution of values within the specified context. Specific details on the calculation of these characters are provided in appendix 4.

For calculating each contextual character, each primary character becomes an input and thus the full set of morphometric characters employed is 26 primary characters and 104 contextual characters totalling 130 characters which then account for detailed characterization of urban morphology in the study area. The methods applied in executing this phase is explained the sections below.

3.3.1. Data Preprocessing for Morphometric Computation

Building footprint data covering the full extent of Nairobi with confidence level above 70% was obtained from the Google building footprint dataset Version 3. This resulted in approximately 800,000 buildings. All multipolygons were exploded to ensure that each building had a unique ID's, after which buildings with faulty or no geometry information were removed. Polygons with an area smaller than 10m² were removed as such smaller objects were assumed to be ancillary structures. Height data sets from World settlement footprint 3d were obtained and pre-processed.

Using a spatial join, the height information was appended to the centroids of intersecting buildings, while the mean height value was computed to replace all missing height data.

3.3.2. Morphometrics Computation Hyperparameters

As explained previously, the urban form characters used in the study are buildings (building footprints) and plots represented by morphological tessellations. Building tessellation geometries from building footprint requires tuning two hyperparameters.

- Inward offset distance: This distance is used in shrinking polygons, to generate gaps between adjacent geometries, to be used in defining the boundaries of each tessellation around the building footprint. Selecting too large values may lead to the collapse of building polygons, while too little results in incorrect saw-like geometries. Thus, this hyperparameter requires priori knowledge of the study area characteristics. The default setting for this hyperparameter is 0.4m based on Fleischmann et al., (2020)'s experimentation on Zurich.
- Discretisation interval: The tessellation generation process involves approximating the polygonal shape of buildings into points, to be used in generating Voronoi tessellations (a subsidiary step to morphological tessellation). This hyperparameter value defined the distance between these points, thus a large value will result in saw-like polygons and loss of shape details whilst the opposite increases computational burden. The default value of this hyperparameter is 0.5m

With the understanding that Nairobi's informal settlements are unlike Zurich and characterised by dense, small buildings (Scott et al., 2017), using the default settings was not optimal as it resulted in massive data loss to building collapse and faulty geometries. Following heuristics, a trial-and-error approach was used in selecting the bet set of hyperparameters that reduced polygon collapse, while reducing overlapping features. Following this experiment, an inward offsetting of 0.6 and discretisation value of 0.4 were adopted as it was optimum in reducing data loss as seen from table 2 below. The derived tessellation along with the pre-processed building footprints were then used as inputs in computing morphometrics.

Table 2: Morphometric hyper parameter tuning

	Inward offset distance	Discretisation interval	Polygon collapse	Polygon overlap
Default	0.4m	0.5m	1	9178

V1	0.5m	0.4	19	705
V2	0.55	0.4	26	425
V3	0.6	0.4	32	342
V4	0.65	0.4	71	303
V5	0.7	0.4	109	283

3.4. Clustering Analysis

To determine distinct morphological areas within the study area, clustering was employed on the computed urban morphometric characters. Clustering is a method of partitioning data points in a feature space into individual clusters, such that members within a cluster, are more similar than members of other clusters. Han et al., (2012) explains that the choice of clustering algorithm should depend on the input requirements, data dimensionality as well as model interpretability. Considering these requirements, in relation to the morphometric data, K-means algorithm was selected. This algorithm is suited for numeric continuous data, is highly scalable, thus can handle high feature dimensionality as well as large volume of samples, and is easily interpretable Han et al., (2012).

K-means assigns clusters by first choosing an arbitrary k object as initial cluster centres, and iteratively assigns and reassigned each sample to the cluster centres of which the sample is most similar, while updating cluster means at each step. Cluster means and reassignment are updated constantly until each sample is closest to its cluster centre than the any other cluster. An illustration of this process is seen below.

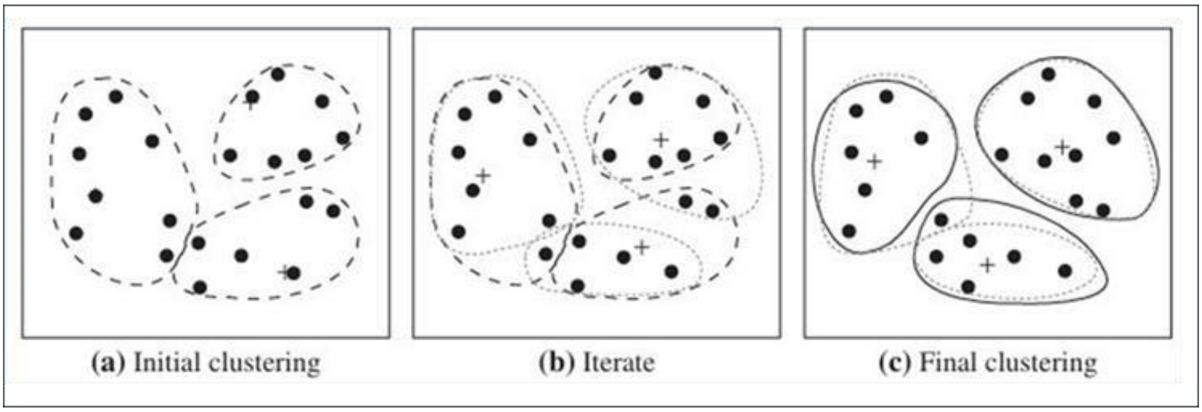


Figure 4: K-means algorithm process; adapted from Han et al., (2012)

The k-means model was necessary to assess if there were significant morphological differences between formal and informal areas. This algorithm requires the selection of a hyperparameter; number of clusters or 'K'. Using all the contextual morphometric characters as features(n=112), this model was run using k = 2 to cluster the study area in two distinct areas.

K-means algorithm has been known to be sensitive to initialisation seeds, (Qi et al., 2016). To buffer for this, the study, applied the k-means++, a derivative of k-means which selects initial cluster centroids based on the empirical probability distribution of sample point's contribution to WCSS is used. The model is repeated 10,000 times across different initialisation, after which the model with the best results in terms of WCSS is selected. To assess the quality of the clustering, the silhouette score, which is a measure of the separation distance between clusters was used.

The k-means algorithm does not provide and inherent feature importance metric. As such, to better understand the specific morphological features most important in discriminating the classes, the mean values of all features were determined per cluster. After which the variance of these feature means is calculated across the clusters. An ANOVA statistic was then performed on these means to evaluate significant variations between the cluster.

3.5. Insitu Air Temperature Measurement

This section details out the data collection, preprocessing and analysis of the insitu air temperature measurements used in this study.

3.5.1. Data Collection Methods

Air temperature measurements were obtained from the study area following a collaborative citizen science approach to leverage local knowledge in design and implementation. Due to this, settlements were selected based on the availability and willingness of community activist groups to collaborate in the data collection process.

Workshops were organised for each settlement to train the participants on the guidelines and the use of equipment for the data collection exercise. During these workshops, participants were engaged to design mobile routes through and around each informal settlement. The routes were designed to cover diverse urban characters based on a qualitative assessment of participants local knowledge of the area stated below;

- Land cover and land uses.
- Building densities
- Road types (major, minor, footpaths).
- Building morphology (high rise, low rise, size, shape).

Participants (citizens) walked the paths of the designed route, while holding an already configurated Kestrel Drop 2 air temperature sensor and Garmin GNSS receivers attached to sticks at approximately 2m above the ground. These sensors logged temperature data at a 10 second interval, thus allowing detailed recordings of temperature. For each settlement, the data collection period spanned 15.00 to 17.00 at which time peak temperatures are expected due to radiative heat loss. During the mobile transect walk period, the Kunak air temperature sensor was simultaneously employed as a reference station to control for daily weather conditions and for temporal corrections during data preprocessing. This sensor was stationed at a fixed location in each settlement, positioned about 2 meters above the ground, away from any artificial influences of temperature.

This method was repeated for each of the 6 settlements, resulting in approximately 18,250 temperature datapoints, acquired between 27th January to 16th February 2024. The table and figures and tables below provide further details on the specification of fieldwork.

Table 3: Acquisition dates per settlement

Settlement	Date	Time
Kibera	27-01-2024	15.00 To 17.00
Mukuru	30-01-2024	
Kariobangi/Koriogocho	02-02-2024	
Pumwani	04-02-2024	
Waruku	08-02-2024	

Table 4: Sensor Specifications

Sensor	Resolution	Accuracy	Settings	Purpose
Kestrel Drop 2	0.1 °C	± 0.5 °C	10 seconds log	Fixed / Mobile Temperature Measurements
Garmin eTrex 10	3metres	3metres	20 seconds log	Mobile Location logger
Kunak Air station	0.01 ºC	± 0.9 ºC	10 seconds log	Stationary Temperature Measurement

Figure 5: Community Engagement and Data Collection

3.5.2. Insitu Air Temperature Preprocessing

Field data was pre-processed following three distinct phases highlighted in the flowchart in figure 9 below. From the figure, Box 1 highlights the methods used in attaining the geolocated temperature. Box 2 explains the temperature gradient modelling, necessary to understand the temperature temporal patterns in the study area. Box three finally applies the outputs of the previous steps for temporal correction, thus ensure synchronous times between the temperature measurement.

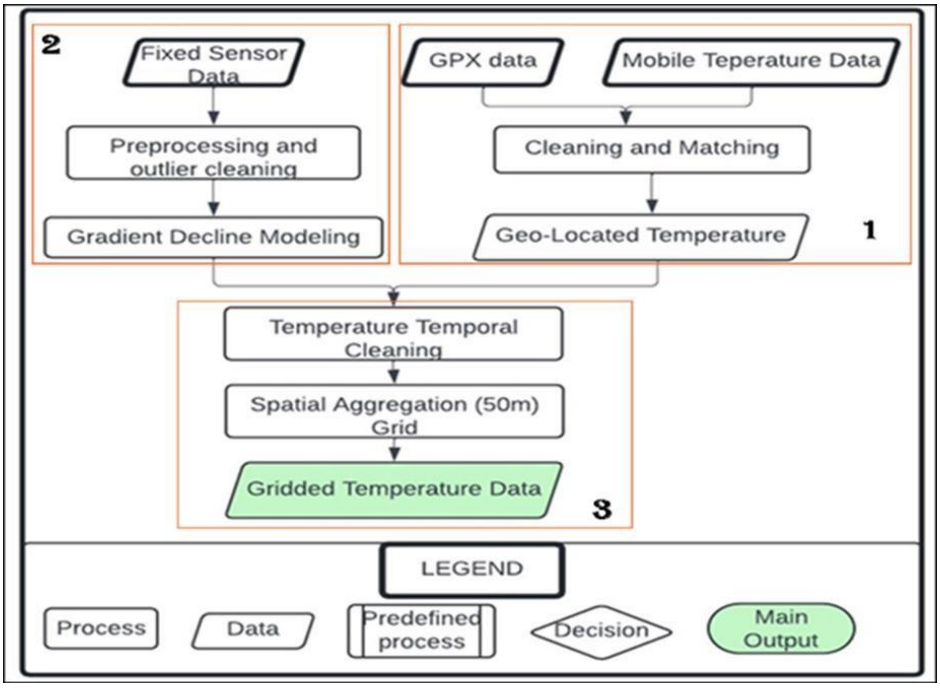


Figure 6:Workflow of field data preprocessing

The numbered boxes in the workflow above guide the flow of progression in preprocessing the insitu measurements. The two datasets obtained from the insitu measurements were cleaned and merged using the time parameter. The data was then cleaned of outliers by, averaging out all datapoints with a rate of change higher than the mean rate of change within the route data. This was done to eradicate sharp spikes and drops within the data. Following this preliminary cleaning, temporal correction was employed, as seen in box 2. According to the work of Fung et al., (2009) and L. Liu et al., (2017), temporal corrections must be applied to mobile temperature to guarantee the comparability of temperature readings. Temporal adjustments should be applied to bring values to a single, unified time point because the temperature records are taken at non-synchronous times. Typically, reference temperature data is obtained concurrently from stationary weather stations, from which temporal patterns are drawn, by assessing the trend that is created between the start and finish times.

Temporal correction was performed using the Kunak stationary sensor as a reference, in accordance with the single temporal correction modelling approach suggested by (Fung et al., 2009). By assuming perfect consistency of temperature variability spatially, this method proposes modelling temperature trends using a single stationary sensor.

Thus, the temporal gradient modelling at each settlement, used the Kunak as baseline dataset in highlighting the trend of temperature. Different degrees of polynomial regressions were then employed to model the temperature pattern over time after which mobile temperature readings were transposed to the same time reference time point (15.15). This results in geo-located air temperature data at the same time point and as such suitable for comparability. Finally, the data was then aggregated into 50X50metre grids, using the mean.

3.6. Auxiliary Variables Computation

Per the objectives of this study, several auxiliary covariates were derived and used in the modelling process. Amongst them included remotely sensed indices such as NDVI, NDWI, NDBI as well as LST. These land cover indices have been widely used in estimating air temperature and have been established to generally have some relationship to air temperature (Benali et al., 2012; Fung et al., 2009; Gholamnia et al., n.d.). This section of the methodology provides the theoretical background of the use of such indices in the study as well as methods used in their preprocessing and analysis.

3.6.1. Sentinel 2 Image Preprocessing and Land cover computation

Sentinel 2 images were collected and used in the computation of land cover indices. This sensor provides publicly available global multispectral imagery with a revisit time of 5 days. It acquires images in the visible and near infrared at a 10m cell size and 20m resolution in the short-wave infrared (SWIR) and thus makes it possible to capture the urban land cover complexity at a fine resolution with a single acquisition. Specifically, Sentinel-2 MSI Level-2A multispectral images were acquired as they are atmospherically corrected surface reflectance values in cartographic geometry.

Using the Google Earth engine API, 5 atmospherically cleaned sentinel 2 images spanning the extent of the study area were acquired within the period of the insitu measurements. Selection criteria was based on prioritising (1) Images within the same season as the data collection period (peak summer season) and (2) Images with the least amount of cloud cover. By following these criteria, heuristics was used in determining the optimum number cloud percentage allowance. 17% cloud cover was selected because it was realised that going below the 17% threshold reduced resulting images from 5 to only 1 image. Thus, ultimately 5 images (each with band 3, 4, 8,11 representing Green, Red and Near Infrared respectively and SCL band) with varying acquisition dates within the summer season were acquired.

The scene classification band (SCL) provided by Sentinel 2 uses a scene classification algorithm to identify cloud pixels of varying probabilities. This was used to identify and mask out all cloud pixels, after which the mean was applied to all remaining pixels across each band, resulting in 4 mean composite bands of band 3, 4, 8, and 11. By visually inspecting the composites, it was realised that there existed some gaps, where there was no pixel information. Thought these areas were located outside the scope of the area of interest, gap filling was applied by replacing gaps with mean values of nearest neighbour pixels within a 3x3 window.

Following the works of the foundational works establishing these indices, as well as their recent uses urban studies, the table below provides a breakdown on the formulas used in computing the NDVI, NDBI and NDWI of the images. The results of these are land cover metrics are seen in appendix1.

Table 5: Land cover metrics computation

Description	Acronym	Formulation	Adapted from Sentinel 2 bands	Reference
Normalized Difference Vegetation Index	NDVI	(NIR − Red)/ (NIR + Red)	(B8 − B4) / (B8 + B4)	(Purevdorj et al., 1998; Yu et al., 2020)
Normalized Difference Builtup Index (NDBI)	NDBI	(SWIR − NIR)/ (SWIR + NIR)	(B11 – B8) / (B11 + B8)	(McFeeters, 1996; Onačillová et al., 2022)
Normalized Difference Water Index (NDWI)	NDWI	(Green − NIR)/ (Green + NIR)	(B3 – B8) / (B3 + B8)	(Zha et al., 2010)

3.6.2. Land Surface Temperature Preprocessing

For a detailed mapping of LST in the study area, The ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) was used as it provides global coverage of the highest resolution (70m) freely available data thus supporting replicability. This satellite enables calculation and mapping of land Surface Temperature every 1-5 days at a 70metre resolution.

Similar to Sentinel 2 data retrieval, ECOSTRESS atmospherically cleaned images spanning the entire extent of Nairobi were obtained. Image selection was limited to the peak summer (01-12-2023 to 30-03-2024) during which data collection period occurred. Following this filtering, four images as well as their corresponding Control (QC) flags were obtained. Using the QC flags, pixels corresponding to bits 1 and 2 (corresponding to best quality data with least uncertainties in retrieval) were kept, whilst all other pixels were set to nan. Following this, only two images; one acquired during the daytime, and the other during nighttime, covered the entire extent of the Area of interest. Minor gaps (nan values), found outside the scope of the study area, were nonetheless filled by computing mean values of nearest neighbour pixels within a 3x3 window.

3.6.3. Characterising Urban Greenery

Lin et al., (2023) and Stewart & Oke, (2012), suggest that different configurations and compositions of greenery potentially influence temperature differently, as such, going beyond the usual NDVI estimation was necessary to capture this more detailed interaction. A local climate zone classification map from WUDAPT with a 70% accuracy at 100m resolution was used to characterize the composition and configuration of greenery within the study area. The data was reclassified such that all classes below class10, that signify different built-up configurations according to the local climate zone classification (see appendix 6) were reclassified to class 0. All other classes (from 11-17) representing different variants of greenery were the focus of this analysis. The resulting raster was then down sampled using the nearest neighbour from 100m to50m resolution (Spatial unit of analysis). Using a buffer of 150meters, around each data point, class area and aggregation index landscape metrics computation is done. Class area was used as a proxy to quantify the composition of greenery whiles aggregation index characterises the aggregation or compactness of greenery within a 150-metre buffer. after which configuration and composition metrics computations were undertaken following the works of (Lin et al., 2023; Peng et al., 2010).

3.7. Scale of Analysis and Data Aggregation

All datasets were unified into a single tabular data frame as this is a requirement for machine learning models. The insitu measurements being the variable of interest, was used as reference data frame, unto which all other variables were attached. However, to ensure consistency, all variables were transformed into the same spatial unit prior to merging. Per the objectives of the study, local analysis of temperature variations considering morphology, two conditions were considered in the selection of the optimal spatial unit of analysis. Given that the insitu measurements were done along routes within the settlement, while morphology, was at a building level, the spatial unit had to be;

- High enough resolution so as limit the aggregation and loss of local details of all variables especially temperature.
- Low enough resolution to maximise information gain of morphological characters.

Previous studies such as (Mahabir et al., 2020) used a trial-and-error approach in deciding the optimal spatial unit. Similarly, this study experimented with varying sizes of regular grid cells of 10x10m, 20x20m, and 50x50m grids. The grids were overlaid with the temperature and morphological datasets, after which the grid size of 50m was chosen following visual inspection as seen in the figure below.

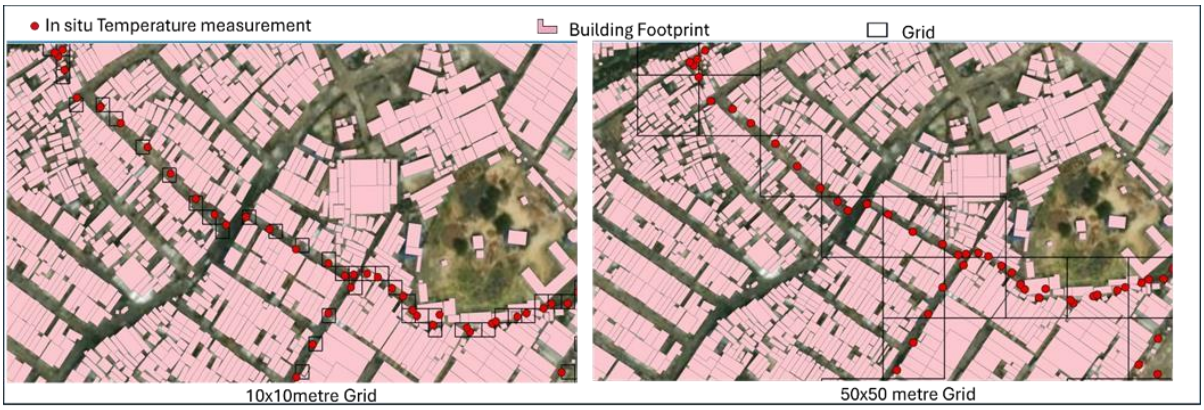


Figure 7: Spatial unit of analysis

By using a spatial join, the insitu temperature measurements are aggregated using the mean and joined to the 50x50 metre grids. Similarly, previously calculated morphometric characters (at the building level) were joined by averaging the morphometric values of building centroids intersecting with the grid. This was replicated for all land cover indices (NDVI, NDWI, NDBI) as well as elevation data obtained from ALOS PALSAR at 12.5m resolution.

In the case of ECOSTRESS LST however, at a 70m resolution, the images had to be resampled to 50m prior to joining. Using the bilinear resampling tool in ArcGis Pro, the 50m grid was snapped to the environment, serving as a reference cell size. This is done to ensure that the resulting image is the same cell size and aligns perfectly with the grid. Finally, resampled raster at the same resolution and alignment as the grid is merged using a spatial join. Thus finally, this step resulted in a tabular format dataset, consisting of 154 independent variables, including morphometrics, auxiliary variables and insitu air temperature measurements as the dependent variable.

3.8. Modelling Air Temperature-Morphology Association Using Random Forest

Proposed by (Breiman, 2001), the random forest algorithm is a non-parametric ensemble machine learning algorithm that leverages several decision trees in predictions. Each tree in the algorithm is built from a randomly selected subset of training data samples and variables after which the final prediction is based on the aggregation of the individual predictions of each tree. The power of Random Forest Regression lies in its randomness and ensemble approach, which mitigates overfitting by introducing two variations of randomness:

1. A bootstrapped sample is used in training each tree.
2. At each node, a randomly select subset of available variables are used.

Bagging, which is essentially a resampling with replacement technique, is used to create several variations of the training data, on which each tree is built, thus each tree is built on a different subset of the training data. By doing this, noises and biases in the data are distributed across the trees resulting in less risk of overfitting.

For each tree, sample data is used in splitting the values of selected variables recursively resulting in tree like structures of several parent and child nodes. Decisions at node points are made by first evaluating all possible splits for all features in the subset and considers all possible threshold values to data splitting based on the optimal split point that ensures the largest decrease impurity or variance for regression. This is repeated severally in each tree until the stopping criteria such as the maximum tree depth or no further impurity is left. All prediction across the trees are then aggregated, as the prediction result.

During this process, due to the random selection of samples and variables, about 1/3rd of the dataset is not used in building the model; the out-of-bag (OOB) which can then be used in calculating the OOB error score for hyperparameter optimisation. Adapted from Carranza et al., (2021), the figure below shows a graphical representation of the random forest algorithm.

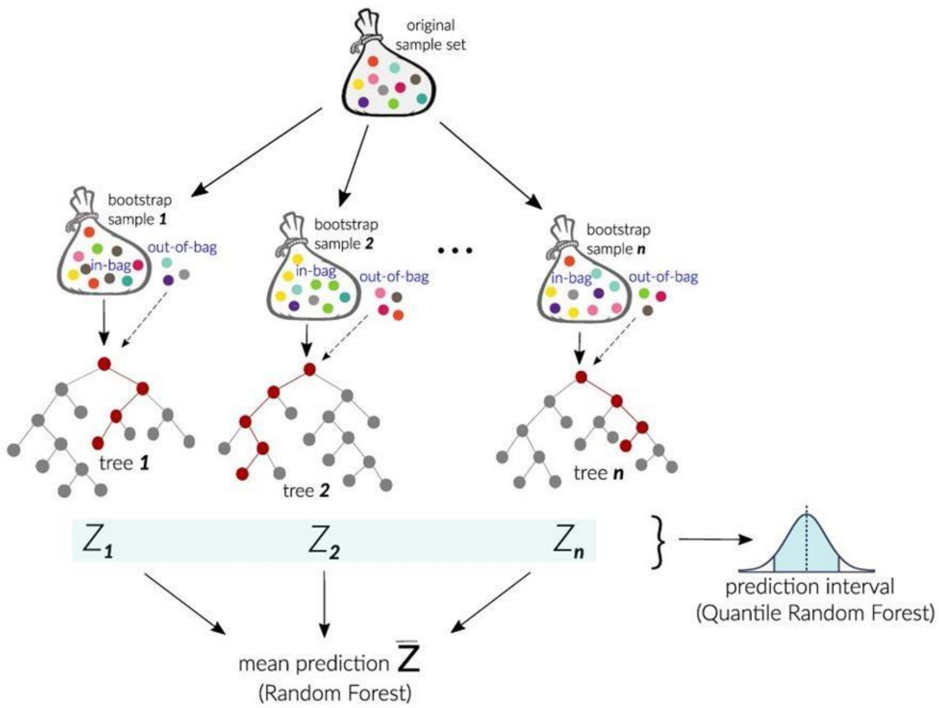


Figure 8: Random forest algorithm; By using bootstrapped samples, each tree is built by splitting feature values at the nodes based on decrease in impurity or variance. The process is repeated until stopping criteria is reached.

3.8.1. Random Forest Model Implementation and Optimisation

Prior to training the random forest model, the dataset was split into 80% for training, testing sets and 20% for validation as is typical for most random forest modelling(Breiman, 2001; He et al., 2022; Yu et al., 2020). This was done by first dividing the entire dataset into 20 equally split grids. 20% of the grids were selected by teasing out grids with the least interaction (minimal boundary intersection) with other grids to ensure spatial disjoints. Thus, validation points should minimally touch boundaries with any training points. This was done to reduce optimistic predictions based on spatial information leakages resulting from spatial autocorrelation.

A two-step hyperparameter tuning was employed to ensure the most optimal set of hyperparameter for training. First an initial model was run for each of the hyper parameters with a varying range of randomly selected values, whilst setting the other hyperparameters constant. The best performing 4 of each hyperparameter were then selected to constitute values for determining the best set of hyperparameter using the grid search cv method. The GridSearchCV method used for decided the best hyperparameter combination, used a 5-fold cross validation to train and test each pairwise combination of hyperparameters, returning the best set that maximised the R^2. Using a 5-fold cross validation method, the training dataset was then employed in training the random forest model. The figures below plot the initial training and test points and 5-fold validation sets. Appendix 7 shows the k-fold cross validation train and testing sets.

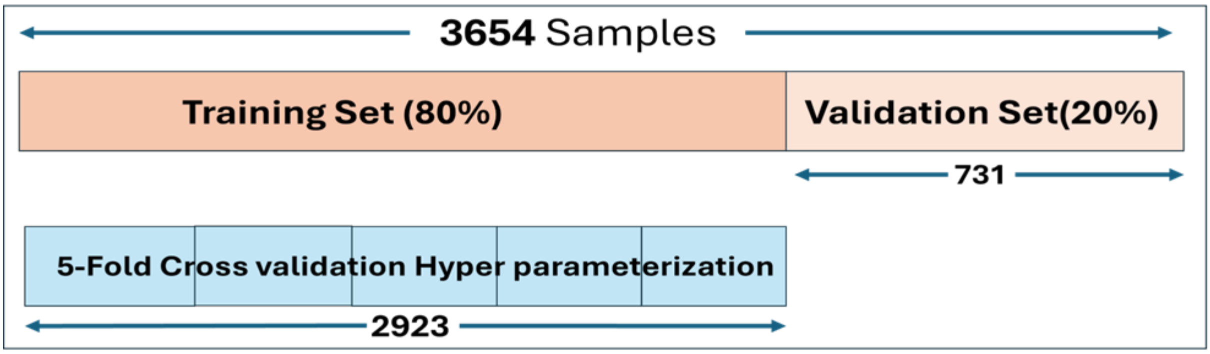


Figure 9:Data splitting: 80% training and hyperparameter testing and 30 for model validation

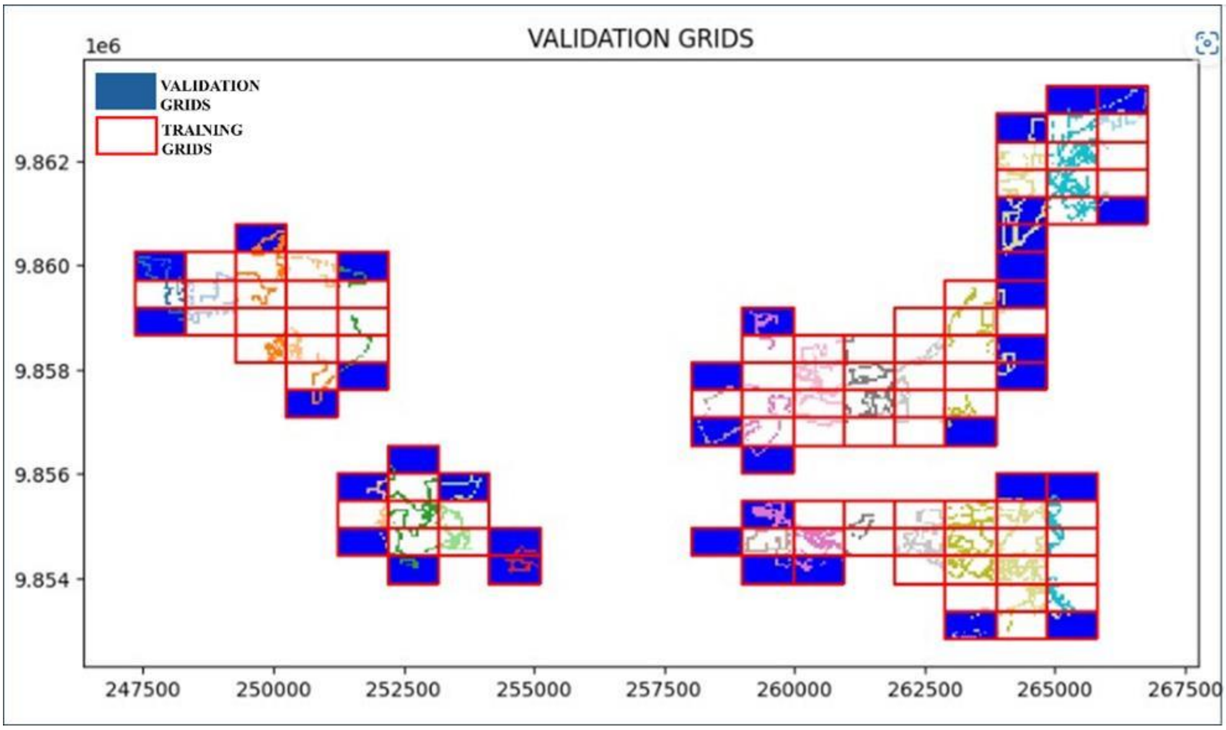


Figure 10:Validation and training grids

Using the same model parameters and configuration, three variants of random forest was built. The first model; a global model, trained across all 6 settlements within the study area provided insights into the global relationships between the features and air temperature. The second and third models were trained and built at regional levels, allowing the exploration of more detailed relationships at a higher scale. This was achieved by exploring the data, revealing significant temperature and elevation differences between the Western and Eastern parts of the study area. Acknowledging the high correlation between elevation and temperature elevation, k-means clustering model was run on the elevation datapoints thus separating the study area into two distinct groups, perfectly separating western and eastern settlements as seen in figure below. Each cluster thereby was used as a region on which the two independent regional models was built, following the same methods used in model1.

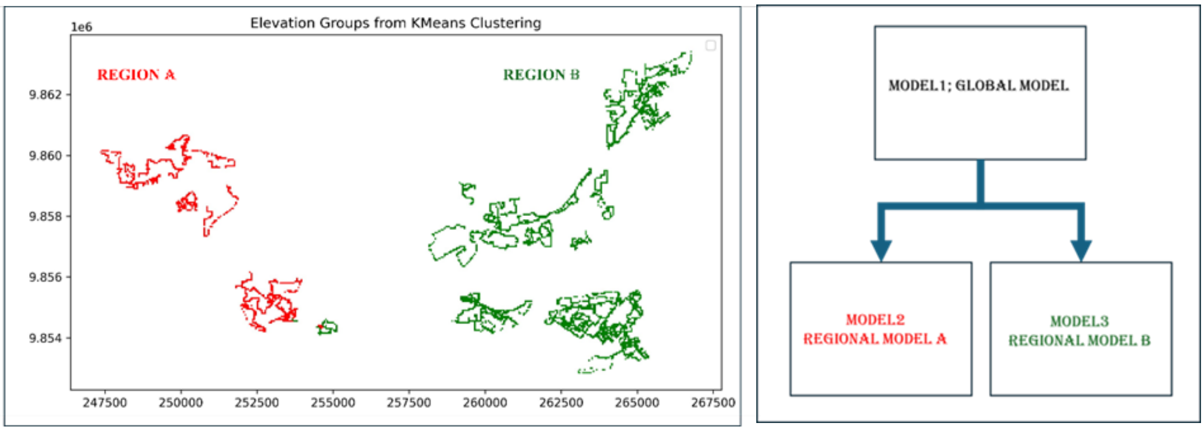


Figure 11:Model Implementation

3.8.2. Model Accuracy Measures

In line with Benali et al., (2012), He et al., (2022), Nascetti et al., (2022) and Peng et al., (2022) coefficient of determination (R^2) as well the correlation score(r) will be used evaluation frameworks on model assessment. These two-accuracy metrics ensured a thorough assessment of the performance and accuracy of the model. The coefficient of determination (R^2) (aka the goodness of fit) is an accuracy measure that measures the proportion of variance of the predicted values, explained by the model whilst correlation coefficient assesses the direction and strength of linear relationships between actual and predicted values, therefore allowing the understanding of underlying patterns.

3.8.3. Citizen Science Model Validation

The predicted air temperature was validated by citizens as this study is related to perceived temperature in these settlements. Three active members and leaders of the community activist groups used during the data collection process were consulted for this validation. These participants have been residents of Kibera, Pumwani and Waruku for not less than 10 years and are knowledgeable about conditions within their settlements.

By engaging in online semi-structured interviews, each participant was asked about the relative thermal conditions of each neighbourhood within the settlement using google earth engine basemap as a reference. Selected neighbourhoods with perceived higher temperatures were pinned after which the map of predicted temperatures is presented. Comparisons of perceived hotspots areas were made with predicted hotspots, after which discussions ensued on prediction deviations.

4. RESULTS

This section of the report presents and interprets the results and findings from the study. The chapter is organised based on the research objectives as such, it begins with an analysis of morphometrics, aiming to explore and decode the various morphological configurations and patterns within the study area, addressing Objective 1. Following this, section 4.2 analyses the covariates through correlation analysis and PCA to provide insights for feature selection as inputs for the random forest model. Section 4.3 highlights the model results and presents important features to temperature variations. The following section analyse local-scale temperature variations across these morphological configurations.

4.1. Urban Morphometrics Results

4.1.1. Assessing Urban morphological Clusters in The Study Area

Following the computation of the urban morphometric characteristics, the K-means unsupervised algorithm was performed to discriminate morphologically homogenous regions. The resulting clusters were mapped as seen in figure12.

The average silhouette coefficient of the clustering was 0.3 which typically indicates a clear distinction or separation between clusters. However, 0.3 is much closer to 0 (decision boundary) than 1 (perfect separation) or -1 (wrong labels) implying that though there is a clear split, data points are close to the decision boundary. Given the proximity between the formal/ informal areas, in this study, it is not expected to have such sharp vast differences in physical conditions.

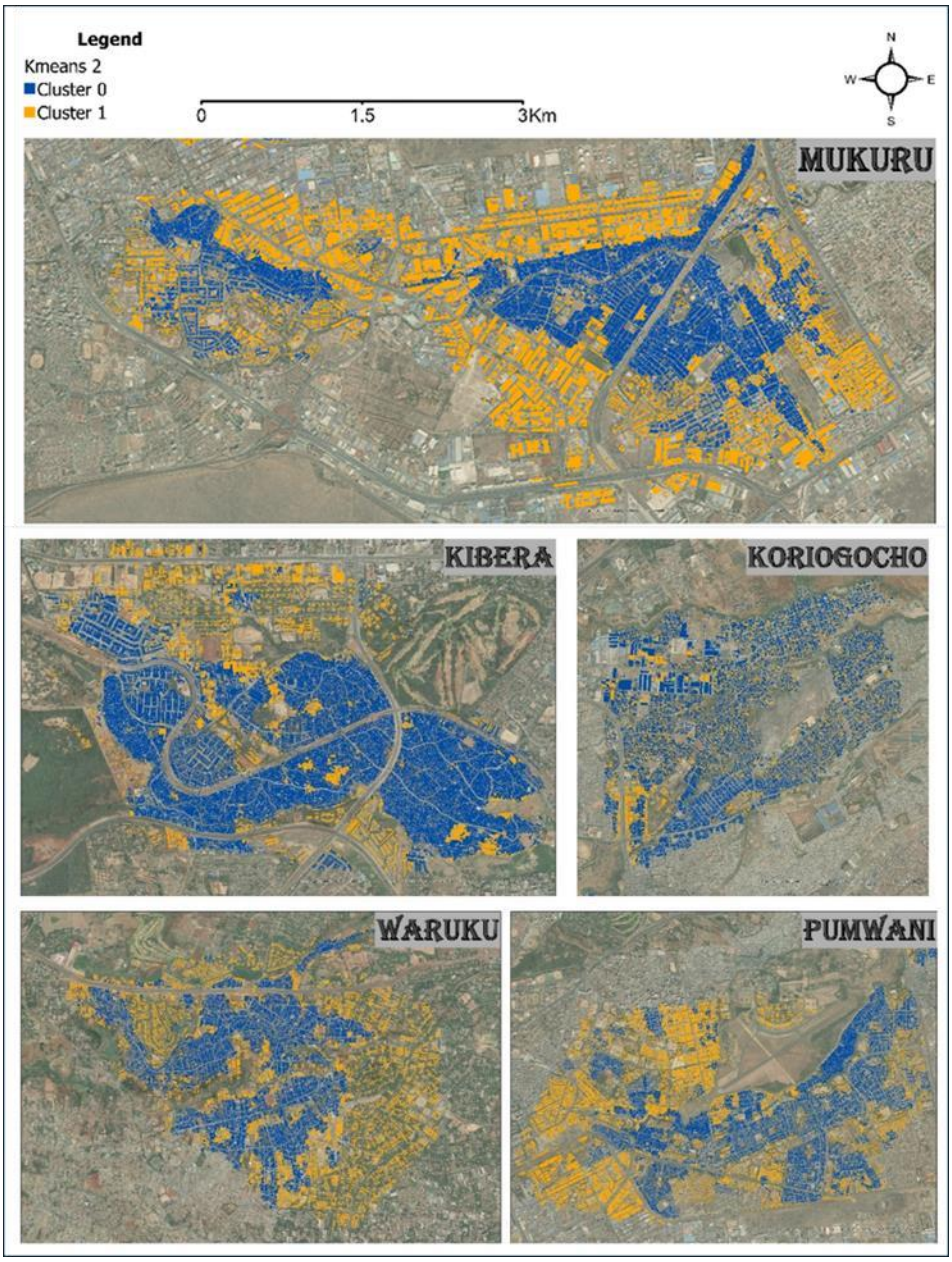


Figure 12:Morphological clusters, k=2

By visual inspection, it is seen that for most of the settlements, there is evident clustering. Much of the central areas of each settlement fall within the same cluster 0, with cluster 1 prevailing more towards the outskirts of the settlement. This suggests that central parts of these settlements tend to exhibit similar morphological characteristics and vice versa. It is interesting to note, however, that all the settlements, but Koriogocho exhibited this pattern. Much of this settlement was classified in cluster 0, with a few random classes of 0 resulting in salt and pepper-like pattern, suggesting no apparent spatial clustering.

The resulting clusters were then superimposed with the deprived urban area buildings to evaluate if the clusters were aligned to the formal/informal categories. By calculating the number of buildings in each cluster that intersected with the formal and informal polygons, it was realised that, globally, 77% of all buildings from cluster 0, belonged in informal areas while 66% of all cluster 1 buildings were in formal areas. This suggests that though with some margin of error, the model was able to separate between DUA and non-DUA areas, indicating that indeed there are significant morphological differences between these areas. This is seen in the figure below, visualising a bar plot and snapshot of the analysis on Waruku.

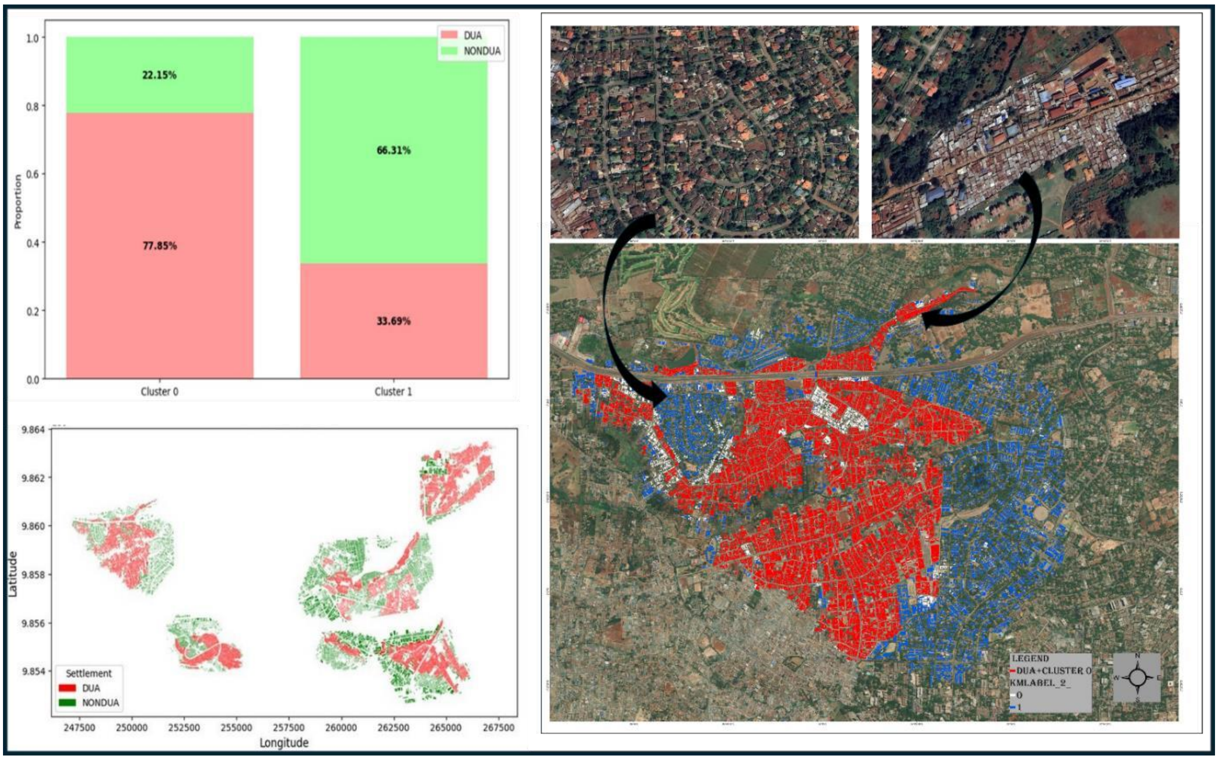


Figure 13:Kmeans evaluation with reference data

4.1.2. Characterising Emerged Clusters

From the feature importance analysis, four metrics; building Perimeter (bPeri_rang), tessellation weighted neighbour (tWNeigh_me), tessellation longest axis length (Tlal_meanl), tessellation covered area (tCA_mea) and building volume façade (BvolFaca_2) had the highest f-statistic score > 90000 with the lowest pvalues ≤ 0.005. This suggests that there were significant differences in mean values of these features between the clusters relative to within cluster variance and thus could better discriminate the clusters. Evidence of this is shown in Figure 18 below that presents a boxplot of the range of the values for the features per cluster. As seen in the plots there are substantial difference in the value ranges of all features between the two clusters, highlighting the importance of these features in separating the clusters.

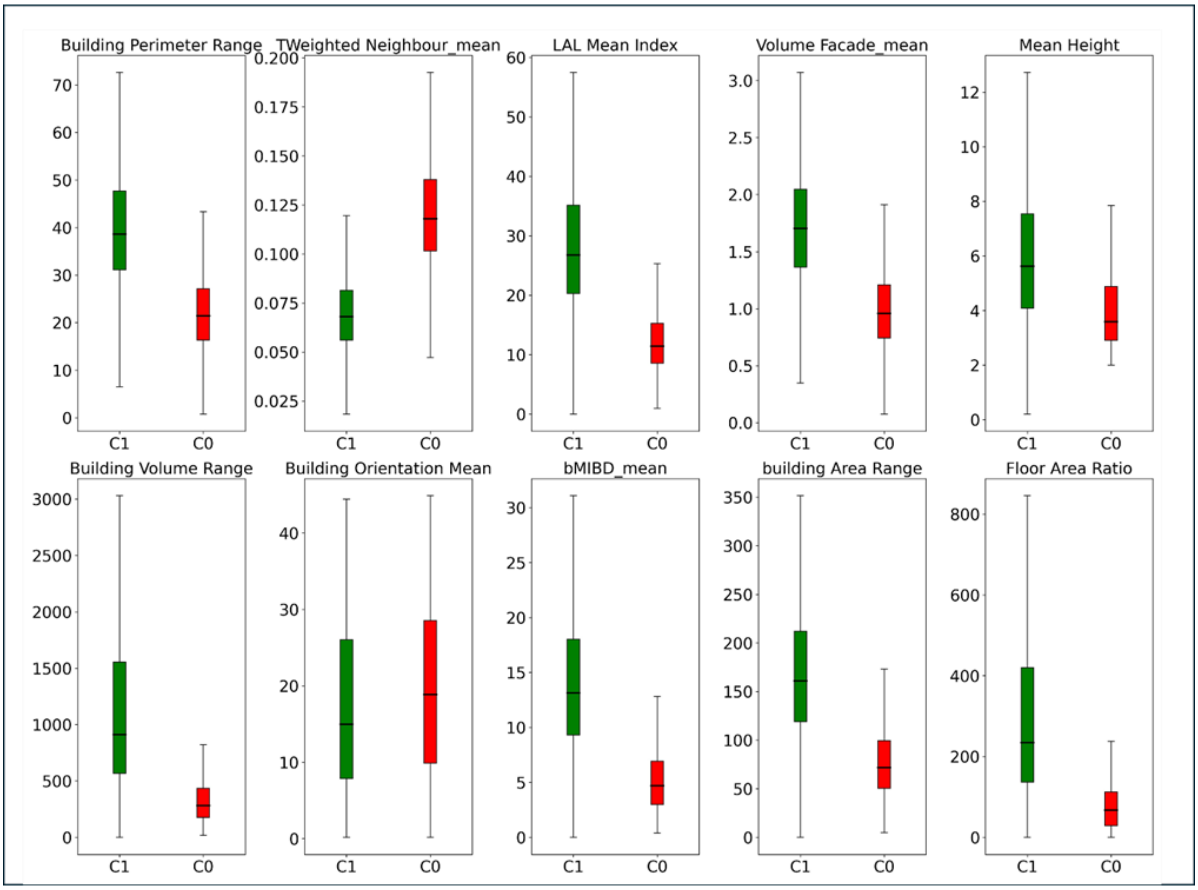


Figure 14: K-means Important features for cluster separation; k=2

Analysing the features most informative in separating the clusters, provides a means to understand the differences that exist between these morphologically diverse areas. From the plots, its intuitive that building volume, has the most difference amongst the two clusters with range of values of cluster 1 about twice that of cluster 0. Thus, buildings within cluster 1 have significantly larger building volumes, this can be associated with the higher building heights and building areas (seen in plots 5 and 9). The opposite is true for buildings within cluster 0, which generally exhibits smaller values regarding building size thus volume, height, area, and perimeter. This is true, for most informal settlements tend to be populated with temporary structures, which are usually smaller in size and height. In line with findings from Abascal et al., (2022), there are substantial differences in size between formal and informal areas.

With regards to compactness metrics, such as mean interbuilding distance and tessellation weighted neighbour, cluster 1 exhibits lower less compactness. Mean interbuilding distance, quantifies the mean distance between a building and its neighbours, providing some insights into how close buildings are to one another. Tessellation weighted neighbour on the other is a weighted number of neighbours within a specified neighbourhood (3 topological steps). Higher values indicate more neighbours within a neighbourhood whilst lower values indicate higher isolation levels and possible open spaces. These metrics provide insights as to how dense and compact neighbourhoods are. With cluster 1 showing wider range of values with a median score much higher than the maximum score of clusters 0, it can be presumed that there tends to be larger intervals and spaces between formal areas as compared to informal areas. This could be associated with formal areas building according to planning building regulations and codes. Combined with low values of weighted neighbours, areas within cluster 1, are likely to have more open spaces at the neighbourhood level, in comparison to cluster1.

Following the evaluation of the K-means clustering and the subsequent analysis of morphological variables within the clusters, it became evident that the study area is divided into two primary morphologically distinct areas corresponding to formal and informal classifications. This initial classification highlights significant differences in the built environment, spatial arrangement, and overall structure between formal and informal settlements.

This diversity is expressed through variations in specific metrics such as size, and compactness, as discussed in section 4.1.2. These findings provide a detailed understanding of the urban fabric and its variations within the study area. Given these established morphological differences, the next critical step is to explore and understand the measured temperature patterns across these clusters.

4.2. Analysing Data Descriptives

Prior to estimating air temperature, exploratory spatial data analysis (ESDA) was conducted to gain insights into the data structure, identify spatial patterns and relationships, and ensure data quality. ESDA is critical for model configurations by providing insights to address spatial dependencies and validating model assumptions, as well as feature selection and engineering. ESDA thus enhances the reliability and accuracy of the modelling process, leading to more robust and spatially informed predictions.

Details of these results on the insitu measurements are highlighted in appendix 8 for further reference.

4.2.1. Correlation Between Covariates and Air temperature

As explained in the methodology several covariates, were calculated to complement morphometrics in the modelling of air temperature. All the covariates were then evaluated in relation to temperature to understand correlations that exist between them in an attempt to understand covariates relationships pre modelling. In total, there were 155 covariates to 1 dependent variable: temperature. Out of these covariates, 142 had statistically significant Pearson correlation with temperature as seen in figure below.

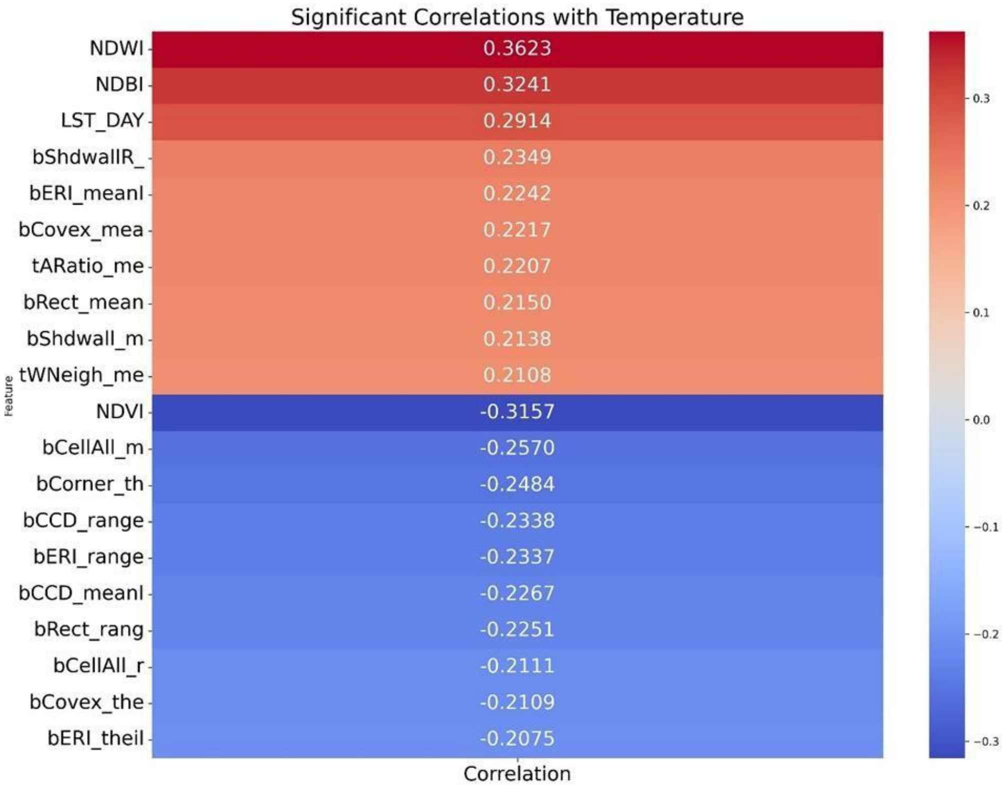


Figure 15: Statistically significant covariates correlation to temperature. From left to right, NDWI, NDBI,LST_DAY, building shared wall ratio range, mean building equivalent rectangular index, mean building convexity, mean tessellation area ratio, mean building mean building rectangular index, mean tessellation weighted neighbour, mean shared wall ratio, elevation, NDVI, mean building cell alignment, building corner theil, building circular compactness range, building rectangular index range, building cell alignment range, building convexity theil.

With a Pearson correlation score of -0.69, elevation had the strongest correlation with temperature, followed by NDWI, NDBI, NDVI and LST_DAY. All morphology related features, though significant had correlation scores below 0.3 with shared wall ratio, being the most correlated at a score of 0.29. Out of all the 136 morphometric features with a significant relationship to temperature, 118 were contextual features with only 18 primary features. This underscores the importance of considering broader neighbourhood characteristics rather than individual building attributes in understanding temperature variability.

Overall, outside elevation, none of these covariates had a particular strong relationship with temperature. This low correlation does not however only signify weak or no relationship to temperature but could also imply that relationships are nonlinear and too complex to be captured by simple bivariate analysis necessitating more advanced analyses.

4.2.2. Determining Feature Selection using Principal Component Analysis

Principal component analyses was attempted to reduce the feature dimensionality, ensuring that only important were considered for modelling. Results of this analysis (seen in appendix 9) showed that no single component could explain more than 20% of the variance individually. Instead, it took the combined effort of the first 10 components to account for 70% of the total variance. This indicates that the dataset lacks a dominant single source of variability that can be captured in a single linear combination of features. Rather, the dataset's variability is spread across multiple dimensions. This underlines the complexity of the relationships among the variables, suggesting potential non-linearities. This finding aligns with the correlation analysis, affirming that while individual variables may not show strong linear relationships with the target, their combined effects across multiple dimensions significantly influence the dataset's variability.

4.3. Temperature - Morphology Relationship Modelling

As previously explained, the Random Forest model was used in modelling air temperature in an attempt to understand the relationships that exist between urban morphometrics and air temperature. Following the results and insights from both the principal component and correlation analysis, all variables were employed in the random forest modelling process.

4.3.1. Determining best hyperparameters for tuning the model.

During the training stage, the range of values for the cross-validation training were selected based on their efficacy in reducing the out of bag error. The results of this step are seen in the figures below.

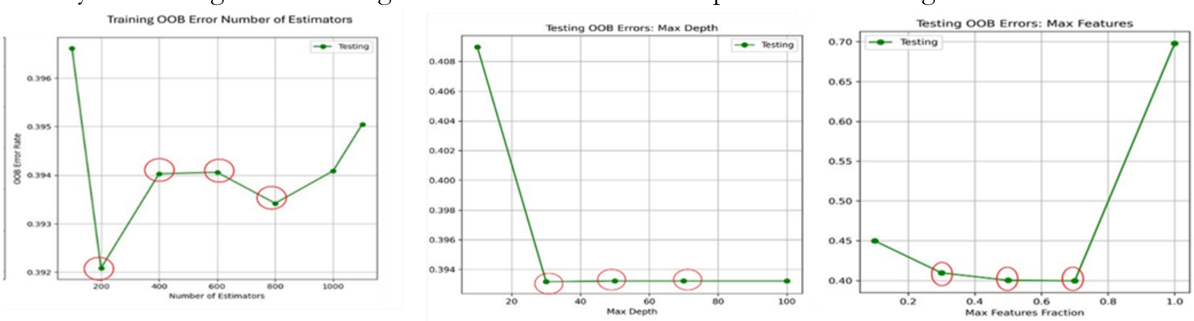


Figure 16: Selecting best set of hyperparameters

From the figure, the best three hyperparameters each that contributed to the least OOB error was selected, to be constitute the range of hyperparameters to use in the Grid search Cross validation method in the next steps. It is important to note that the first initialisation seeds for this experimentation is always influenced by heuristics, as such, it is best to assume local optimal hyperparameters and not global. As expected in machine learning training, increasing the number of estimators as well as max features, initially resulted in reduced OOB errors as the model learns the patterns within the data until a point of overfitting, at which errors increase substantially on unseen data resulting in high OOB scores. The following range of hyperparameters ('selected range' in the table below), thus were considered. And finally an optimum set of hyperparameters; 800 trees, maximum tree depth of 30, and maximum feature selection of 0.7.

Table 6:Selected set of hyperparameters for model training

Hyper Parameter	**Selected Range**	**GridSearchCV best**
Number of estimators	200, 400, 600, 800,1000	800
Max depth	30, 50, 70	30
Max features	0.3,0.5,0.7	0.7

Model1 had an accuracy (R^2) of 0.89, Root Means Square Error of 0.36 and out of bag error rate of 0.46 on the training set. Figure17 A represents the training accuracy, that plots the relationship between the predicted and actual temperature values. From the plot, there is an almost perfect alignment of fit of the predicted and actual values seen by comparing the line of perfect fit. This indicates that the model performed very well in predicting the already seen (trained) datapoints. With a testing accuracy (R^2) of 0.73, RMSE 0.98, and OOB error of 0.6, the test set performed slightly below the training, as expected. From the right plot in the same figure 17there are slight deviations from the line of perfect fit, especially towards the tail ends of the temperature distribution, suggesting that the model errors increase as it predicts increasing temperature values. Thus, for the higher temperature ranges, the model tends to predict lower temperatures than the observed. This is a typical shortfall of the Random Forest regression algorithm, in that it takes an average prediction of all trees, resulting in smoothed predictions. This means that outliers are penalised as they have to significantly affect the split points to influence predictions. This occurrence is seen in the kernel density functions comparisons between the training dataset and predicted values in appendix 10. This notwithstanding, model1 results indicates a good performance as the features were able to explain about 78% of the variance in temperature.

Figure 18 B on the other hand plots the testing accuracies for Model 2 and 3 regional models (Trained on the data subset). With an R^2 of 0.54 and 0.5 respectively, these models performed significantly poorer than the global model. This can be associated with the relatively small training data points used in building these models, resulting in possible model overfitting and thus low testing accuracies. Despite this fact however, both models exhibit similar characteristics in prediction as seen in the residual errors in the tail ends of both distribution plots.

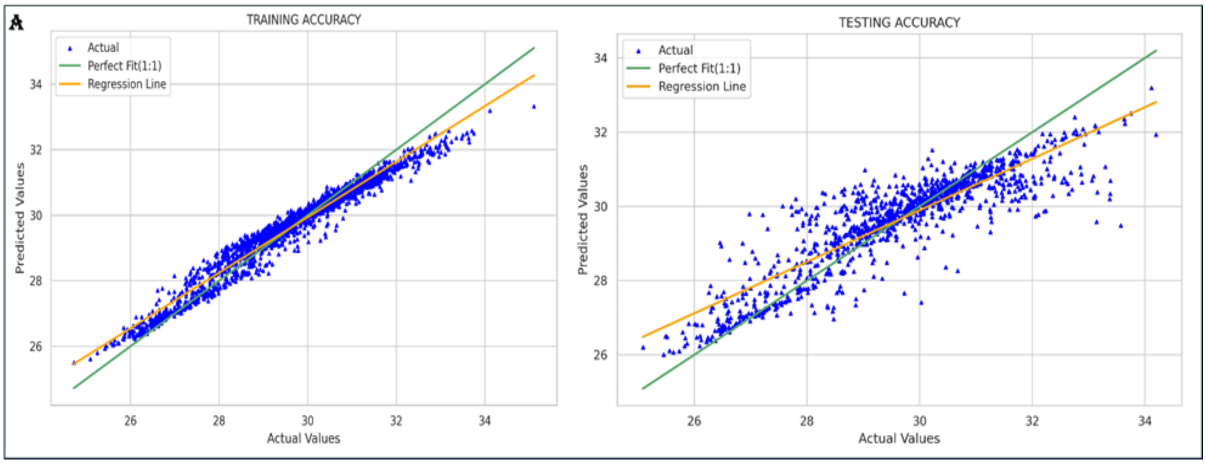


Figure 17: Training and Testing Accuracies Model 1

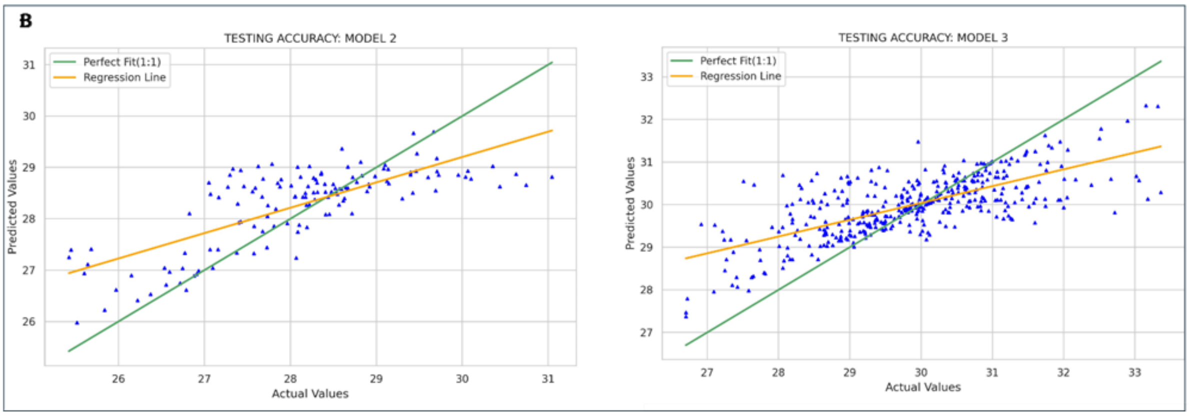


Figure 18: Testing Accuracies Models 2 and 3

4.3.2. Important predictors for temperature across all models

Using both the mean decrease in impurity and the permutation importance feature of the random forest algorithm(global) the independent variables were inspected to understand the most influencing to temperature variance. This analysis is important to understand the extent of each feature affect temperature variation. The two methods employed assess this contribution differently. By averaging each features contribution to variance at each node across all trees, the Mean decrease impurity accurately assesses the importance of each feature in the model. This method has been accused of its bias in over inflating the importance of high cardinality features (such as elevation) as it provides more possible thresholds for splitting, though they may not be intrinsically the most important. Given this setback, the permutation feature importance method, which randomly permutates feature values and compares changes in accuracy as a means to understand feature's contribution to target variable was also used.

From the two plots (from Model1), the same set of variables are considered important, though at different ranks, signifying consistency of these variables in contributing to the target. In both cases, as seen in the plots below that ranks the top 20 variables, elevation consistently was selected as the most relatively important variable, with a significantly large gaps between the next important variable. This is consistent with the previously conducted correlation analysis, that revealed elevation having the highest correlation score of -0.7. This is not a novel discovery as several literature exist on the lapse rate phenomenon, that describes the general decrease in temperature with increasing altitude. This is also captured in the works of Janatian et al., (2017) and K. Wang et al., (2011), explaining that variables such at altitude that are related to the spatial variability of climatological factors tend to have significantly larger influences to surface air temperature. This notwithstanding, it is important to acknowledge that the model's biases to high cardinality features in the splitting stage can overshadow the influence of other features in the model.

Similarly, remote sensing variables, related to land surface and land cover properties (LST, NDVI, NDWI, NDBI) were ranked next important after elevation with an average range between 0.08 and 0.05. Morphological variables then followed with relative importance less than 0.05, except mean building height, which ranked higher than both NDVI, NDBI and NDWI, making it the most important morphological variable. It is interesting to note that all the selected important morphological variables are contextual with no primary character being of relevance. This finding is consistent with both the correlation and Principal Component analysis. This goes to show the importance of neighbourhood level characteristics to temperature.

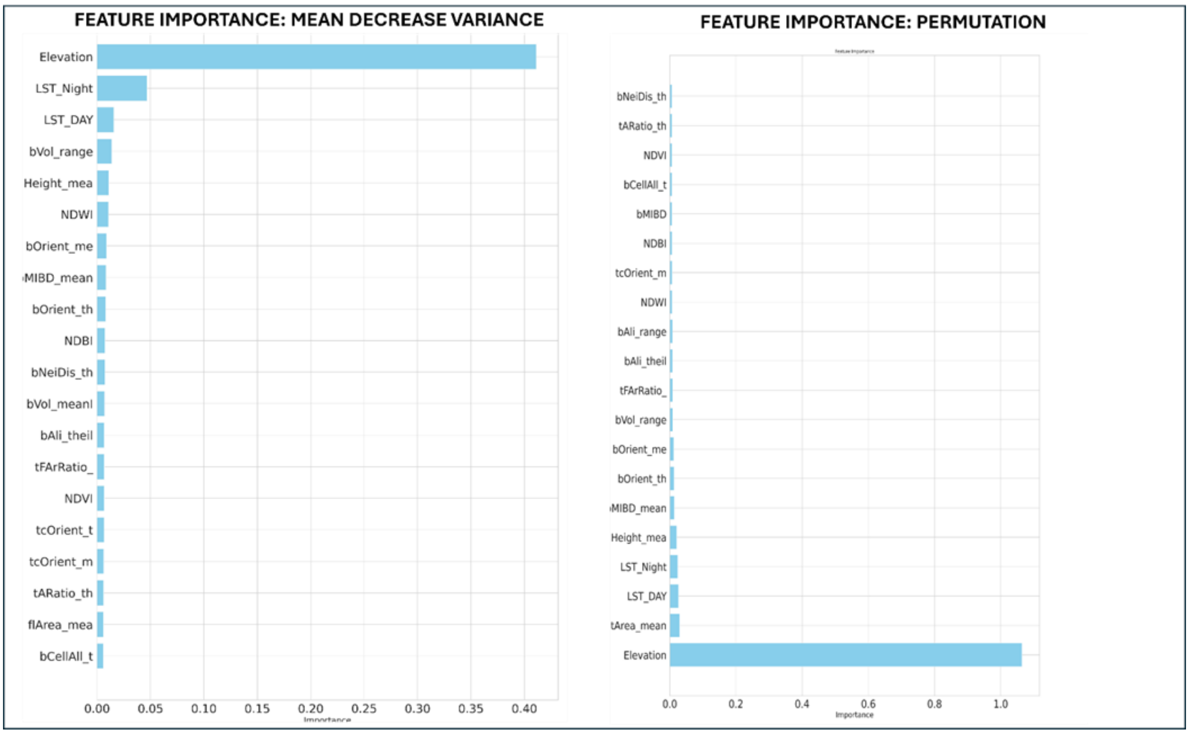


Figure 19: Feature importance of global model. From top; Elevation, LST Night, LST_DAY, building volume range, mean height, NDWI, mean building Orientation, mean interbuilding distance, building orientation, NDBI, building neighbour distance, mean building volume, building alignment theil, tessellation floor area ratio, NDVI, mean tessellation orientation, tessellation area ratio, mean floor area, building cell alignment.

Conversely, the two regional models shared similar feature importances with the global model, with elevation being the most important feature (see appendix 13). However, it was realised that in both models, the magnitude of importance was reduced significantly compared to the global model. With an importance score (mean decrease in variance) of about 0.5 in the global model, elevation importance was 0.26 and 0.07 in model 2 and model respectively. Due to the large disparities in elevation and locations between these two regions, the importance of elevation was over-emphasised in prediction in the global model as it provided a consistent linear relationship with temperature globally, whilst other morphological variable's relationship with temperature may change across different areas, though not making it any less important. This indicates that by restricting modelling to regional areas, as expected, the influence of elevation in prediction was decreased substantially, allowing to focus more detailed scale patterns of morphological influence at a more local level. Thus, similar to the modifiable areal unit problem (MAUP), this modelling was necessary to be executed at multi scale to understand the nuanced contribution of all variables in the prediction.
Though the selected most important features were similar to that of model 1, in the regional models, the apart from Elevation and LST_Night, morphological variables ranked consistently higher than the other land cover and land surface properties, similar to the findings of (Janatian et al., (2017) , stating that land surface and cover properties were found to have an almost negligible effect on temperature variations especially when there is limited variability and range of values, as expected in these regional models.

The results of these plots relate to the implication that individual urban morphometrics may not play significantly to temperature variability. This is supported by the partial dependency plots employed on all three models to ascertain the individual influence of these importance metrics on predicted temperature variations. The results of these are seen in Appendix 11 and 16.

4.3.3. Sensitivity Analysis: Analysing feature influences.

To explore the significance of the various predictor variables to temperature, sensitivity analysis was carried out. Three scenario models were developed by varying and changing the predictor sets used in training, while holding all hyperparameters constant. The evaluation metrics were then compared across these models to understand their significant contributions to temperature as well as ensure consistency of results. In the first model, all variables were considered and used for training, the second model was trained on all variables except elevation, leaving morphometrics and land cover indices. In the final scenario model, only morphometrics were used in training.

Table 7:Sensitivity Analysis

Scenario Model	**Training Variables**	**Hyperparameters**	**Results**			
			Training		**Test Set**	
			RMSE	**R^2**	**RMSE**	**R^2**
A	All Variables	max_depth=30, max_features=0.7, n_estimators=800	0.36	0.85	0.98	0.73
B	All Variables minus Elevation		0.39	0.71	1.4	0.36
C	All Variables minus Morphometrics		0.42	0.92	1.0	0.54
D	Only Morphometric Variables		0.45	0.62	1.5	0.29
E	All variables + composite Morphometrics		0.79	0.74	1.1	0.61

By varying the sets of features used in training, it is possible to gain more insights to the features significance to temperature as well as the robustness prediction. From the table above, there is a consistent drop in model performances as more variables are removed. There is a significant drop in R^2 from 0.73 to 0.36 by removing elevation as a predictor, affirming the importance of elevation to the model. In scenario C, when all variables are used without morphometrics, performance again reduced by 0.73 to 0.54, implying its influence in explaining variance. There is an even larger drop in R^2 when the model is trained only on morphometric variables; scenario D is only able to explain about 30% of the observed variance in temperature.

The greater the decrease in model performance, the higher the importance of the predictors removed. From this analysis, though morphology has some influence in temperature variation, it cannot be employed alone as the main contributor to temperature, instead a confluence of variables especially elevation; relating to actual topography and environmental factors contribute the most to temperature variance. This is corroborated in appendix 15 that reveals the higher importance of morphometrics when compounded and aggregated into a single index. Thus, individual urban morphometric characters have little to no impact to temperature variation, but as a whole, they can be quite significant, implying confounding interactions.

4.3.4. Analysing relationships between Morphology and predicted temperature.

Conversely, the relationships between the features and temperature are analysed by plotting the predicted values against each important feature. Though as established previously, there is a significant importance of elevation as compared to the other features, this analysis provides much detailed insights, example being the direction of linear relationships which feature importance does not provide.

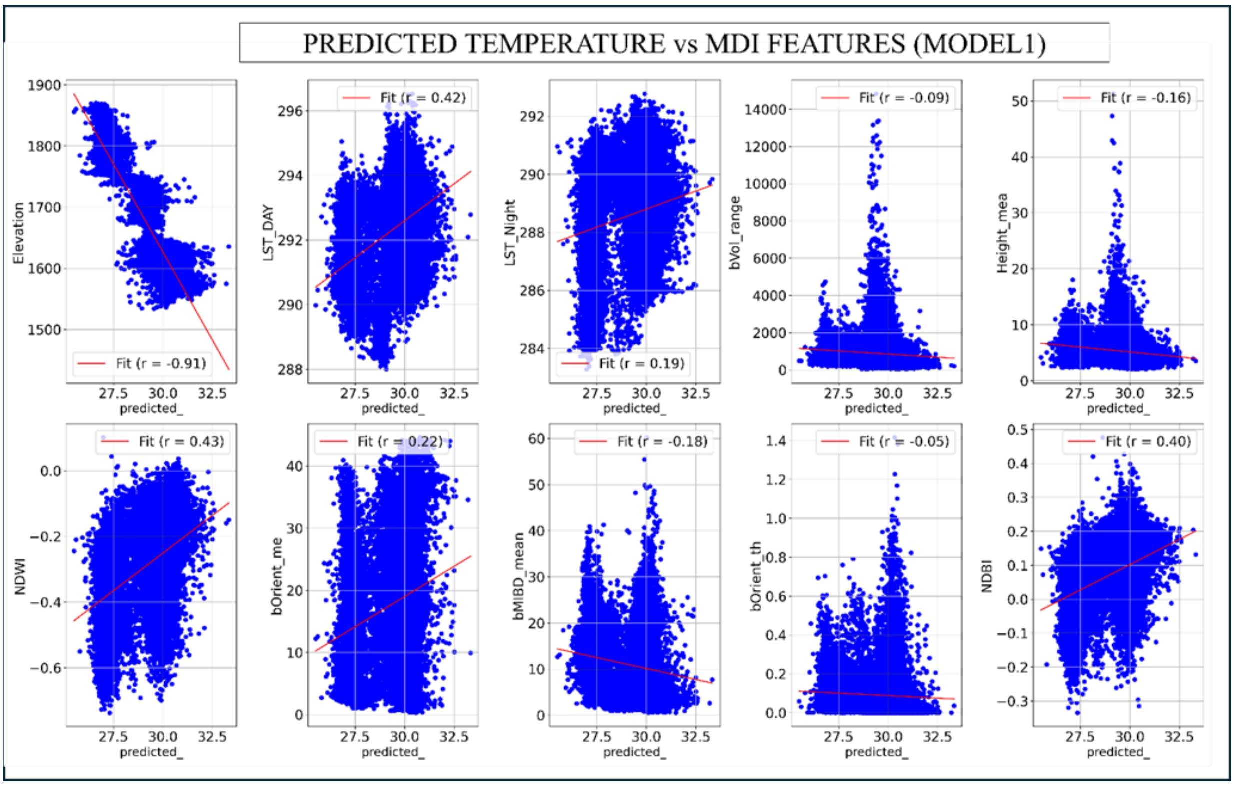


Figure 20; From top left; elevation, LST day, LST night, building volume, Mean height, NDWI, building orientation, NDBI. Feature relationship with predicted temperature Model 1.

The plot above verifies the feature importance results, showing that elevation has a signifying relationship with temperature. With a correlation r of -0.91, it is evident, the near perfect linear negative relationship; increasing elevation associated with decreasing temperature values. The other plots however indicate weaker relationships, though intuitive interpretations of the directions are relevant. For example, in the last plot, a positive relationship between NDBI and temperature Is expected as built-up surfaces like concreate tend to have lower albedo and absorb large amounts of heat from the sun, resulting in higher temperatures. Similarly, the Bmibd_mean plot (third to last plot) which represents the mean interbuilding distance shows a negative linear relationship, implying that higher building distances are associated with lower temperatures as greater spacing between buildings reduces the canyon effect thus allowing for better wind circulation better heat dispersion.

Given the weak relationships observed from model 1, the same analysis is conducted using predictions from the regional level (Model2) to ascertain if by zooming in regionally, the impact of elevation is reduced, thereby allowing better understanding of morphological relationship to temperature Appendix13. It was realised that despite the same negative linear relationship observed from model 1, the r of 0.24 observed from the relationship is substantially lower from model1 of 0.9. Instead, in this model building volume, mean interbuilding distance and NDBI had better R^2, indicating a better relationship between these features and temperature.

4.4. Spatio-Thermal Dynamics of Predicted Temperature

Given that model 1 had the highest accuracy and thus could explain the variance in temperature the most, it was adopted in prediction of air temperature across the entire study area. From the map below, the predicted temperature ranges between 25°C to about 33°C implying some significant variability. It is evident that lower temperature values tend to be located towards the western part of the map. Waruku and Kibera, in particular, exhibit generally lower temperature profiles compared to other areas, with median temperatures ranging between 27°C and 28.5°C as seen from the boxplots in figure 23 Kibera experiences higher variability in temperature distribution, with temperatures range of about 25°C reaching up to 31.5°C.

This increased variability in Kibera might be credited to its heterogeneous urban morphology within the settlement. According to Puche et al. (2023), morphological differences can lead to localized temperature fluctuations, resulting in the observed variability.

Additionally, both Waruku and Kibera, which exhibit lower temperatures, are situated in the western part of Nairobi. This region is characterized by higher altitudes, suggesting that topography plays a significant role in moderating temperatures as explained in the sections above. Higher altitudes are generally associated with cooler temperatures due to the adiabatic lapse rate (Janatian et al., 2017), which causes temperatures to decrease with elevation gain. This topographical influence, combined with the complex urban morphology, likely contributes to the cooler and more variable temperature profiles observed in these settlements.

Despite this influence, there still exists obvious variations even within these settlements, though not particularly extreme, central areas exhibit relatively higher temperatures compared to the outskirts as seen in Waruku. The pattern is observed amongst most settlements particularly in Mukuru, except Koriogocho which had much of the settlement exhibiting uniformly higher temperatures.

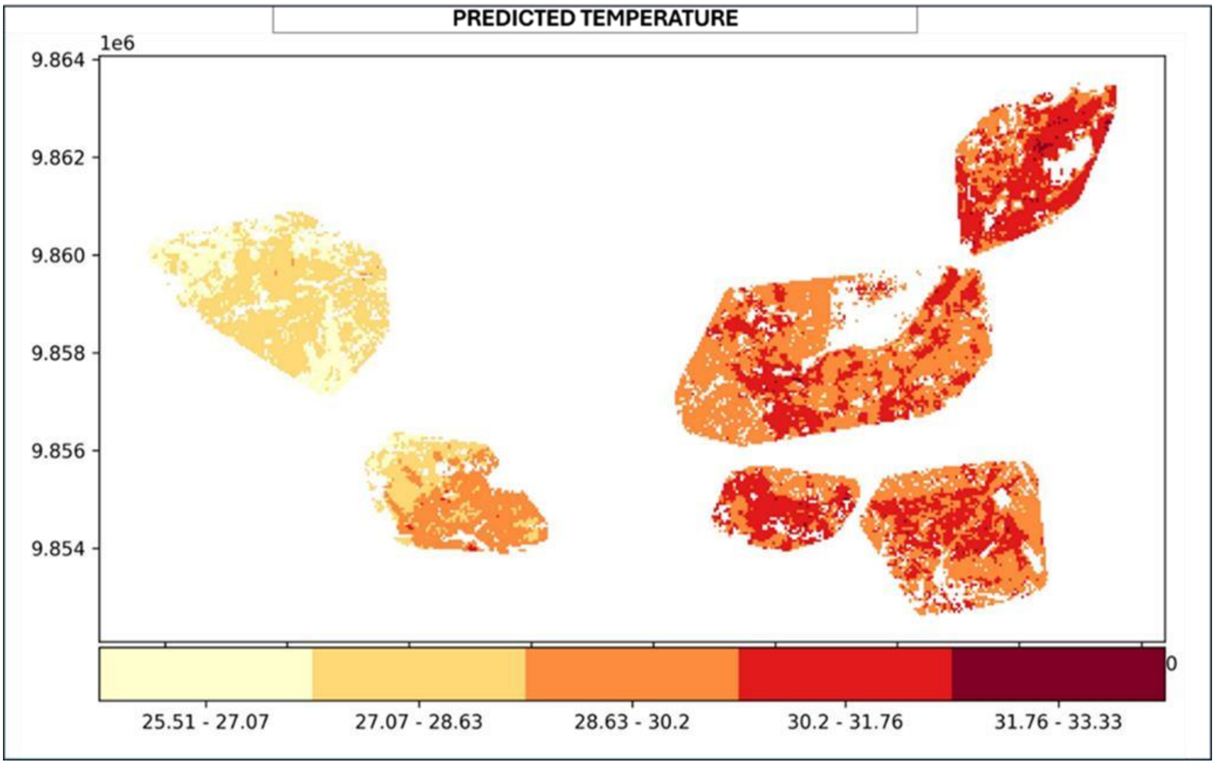


Figure 21: Predicted Air temperature

Using the settlement boundaries, temperature was compared across all informal and formal settlements to assess global temperature differences. The comparison reveal difference in temperature ranges between formal and informal areas, suggesting that the settlements generally do not experience extreme temperature disparities. Notably, formal areas tend to exhibit slightly lower temperatures, with approximately a 1°C difference in both median and minimum temperature values compared to informal areas. Statistical significance of this difference was assessed using an independent Mann-Whitney U statistic. This nonparametric test was chosen over the other popular tests due to non-normality of the data distribution within the two groups. With a significant p value of 1.14×10^{-98} far below the 0.05 accepted significance level, there is strong evidence of statistical difference in temperature between these groups. This supported by a U statistic of 52,675,461.0 indicating significant magnitudes in temperature differences between formal and informal areas.

To explore this at a settlement scale, the comparison is performed per settlement, allowing to evaluate the temperature differences between formal and informal areas within each settlements shown in the following boxplots in figure 23.This revealed that across all the settlements, informal areas consistently showed a higher median and minimum temperature as compared to formal areas within the same settlement though not with very large variations, reiterating the global assessment.

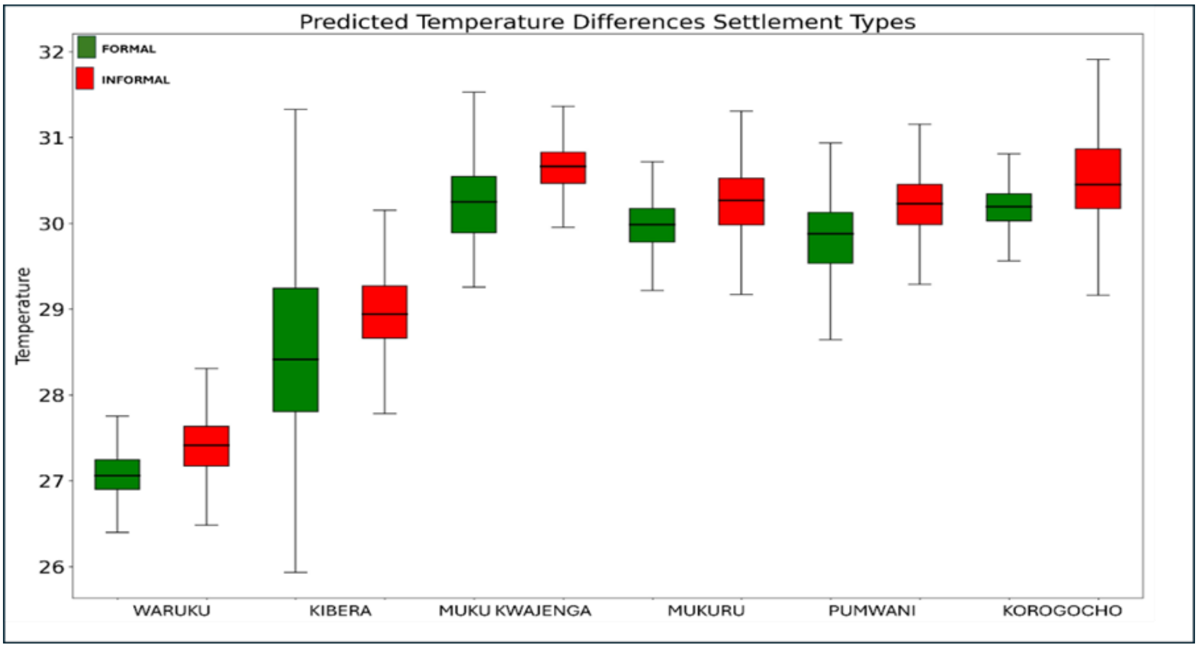


Figure 22: Boxplot of Predicted Air temperature

This analysis was replicated using morphological clusters identified in section 4.1. This step was crucial for investigating thermal variations across morphologically diverse areas and linking temperature variability with morphology. Utilizing clustering outcomes from the clustering analysis, differences between two distinct clusters: Cluster 0, primarily comprising informal areas, and Cluster 1, characterized by formal areas were assessed. As anticipated, the results indicated higher temperature values in Cluster 0 and lower temperatures in Cluster 1, consistent with expectations. Details are provided in the appendix 12 for further reference.

The identified important features to both temperature variations and morphological clusters are compared. This was necessary to further explore the association of morphology and temperature. From the plot, it is observed that morphometrics that were ranked with the highest correlation (between 0.29 to 0.27) were amongst the same ones seen as most important for distinguishing the morphological clusters. Similarly, most feature importance morphometrics were significant in distinguishing morphological clusters. The table below provides a summary of all important morphometric features and the association with distinguishing clusters.

Table 8: Important features across morphological distinction and Temperature prediction

Morphometrics	**Significant correlation (with predicted air temperature)**	**Feature Importance metrics**	**Distinguishing Morphometric Clusters**
Building volume range		✓	YES
Mean building Height		✓	YES

Building mean interbuilding distance		✓	YES
Theil building orientation		✓	NO
Cell alignment		✓	YES
Formfactor		✓	NO
Floor area ratio		✓	YES
Mean tessellation Area		✓	YES
Building circular compactness	✓		YES
Equivalent rectangular index	✓		NO
Building rectangularity	✓		NO
Building Corner mean	✓		YES

4.5. Citizen science Validation of Predicted Temperature Maps

Given that the study concerns the thermal comforts in the communities, the qualitative judgements of citizens concerning the relative thermal conditions in their settlements were of interest. A semi-structured interview with participants was used to this effect.

Overall, in all the settlements the general consensus was that the model could capture the intra-DUA temperature variations well, with evident temperature differences between DUA's and peripheral formal areas though with differing accuracies between settlements. They shared the optimisms that the modelled temperature highlighted higher temperatures within central DUA as compared to the outlying areas, which was confirmed to be true due to the densities and compactness. They also highlighted the model's accuracies in capturing differences even within DUA's, remarking the changes in temperature between differing urban characteristics. For instance, lower temperatures in central Mukuru, along the river as well as in open green pockets.

There were however some abnormalities observed between perceived and predicted thermal conditions. In Waruku especially, formal areas in the periphery, surrounded by lush greenery, and open spaces perceived to be cooler areas were predicted to be only slightly cooler than central DUAS. It was confirmed that these areas, though with higher greenery, were lying in much low-lying regions than the DUAS and other formal areas in Waruku. This was corroborated with the elevation map of the settlement, confirming that this area, though formal, was the lowest elevation point in Waruku and thus had contributed to slightly higher predictions. This notwithstanding, it was still slightly predicted to be lower than the DUA, thus given the impression that minus elevation, these DUAS would fare worse. Deviations in Pumwani revealed the importance of building materials and land use in prediction. In Kitui Village, a DUA within Pumwani, the model estimated quite high temperatures, however, this is an area that locals perceive to have lower temperatures. Despite Kitui Village being a DUA it is majorly constituting of mud and wooden huts and is residents in these areas are considered to be better off due to cooler temperatures even during the peak temperature periods. Because the training data did not include information on building materials, model predictions were made without this consideration and thus resulting in such deviations. Conversely, industrial areas around Koriogocho were predicted to have cooler temperatures relative to DUA's. The participant from Koriogocho however explained that despite this area's large buildings and relatively open spaces, the industrial activities release persistent hot fumes leading to the general warming of that area. It was therefore expected to have warmer conditions, highlighting the importance of considering land use types in temperature modelling.

Participants shared remarks on the promising results expressed pride in having contributed both in the data collection as well as its evaluation. They stressed on the usefulness of such maps especially in advocating for better living conditions within DUA's and the urgent need for upgrading in line with climate stress concerns.

5. DISCUSSIONS

This study aimed to explore local-scale temperature variations to uncover underlying patterns of urban morphological influences. This chapter presents an interpretation of the findings of the study, beginning with a summary of the proposed model, scope, and its applicability, indicating what the model is and what it is not. Subsequent sections provide insights to the patterns of temperature and its relationship with building morphology, offering a thorough interpretation of how urban form impacts thermal conditions within informal settlements.

Further, these insights are discussed in the broader context of vulnerability analysis and the societal benefits of the study. The chapter also characterizes uncertainties and limitations encountered during the research and provides recommendations for future work. Finally, the conclusions section explains the key takeaways, emphasizing the implications for urban planning and climate adaptation strategies.

5.1. Applications of Proposed Model

The air temperature prediction model in this study was successful in capturing local scale temperature variability and provided insights into understanding these variations across morphologically diverse areas in the informal settlements under study.

It is important to establish that the proposed model is not a physics-based model but a statistical regression model, that empirically models air temperature as a function of urban morphometrics, LST, land cover metrics and auxiliary variables. It is therefore intricately tailored to the specific context of the settlements under study's unique characteristics; thus, it provides accurate estimates within the spatial and temporal scope within which the research was carried.

Given the absence of physical parameters such as wind patterns, solar radiation, humidity etc, the model is based on observed patterns rather than underlying physical mechanics. The results of this model are, not to be interpreted as absolute measurements but to provide insights to relative thermal conditions within the study area. Consequently, the variations it captures are specific to the unique interactions and characteristics of the original environment, deeming the model's limitation to generalisability in different contexts. When applied to a different geographical or climatic setting, where these physical interactions vary significantly, the model may fail to accurately predict temperature variations.

Although it is primarily designed for the study area, the model has the potential to provide preliminary insights into thermal variations in other urban areas with similar climatic and morphological characteristics. For instance, other informal settlements even within Nairobi that were not studied or in in other tropical cities with comparable urban density and building structures might show analogous temperature patterns, though with wary interpretations.

Thus, the proposed model is not a one-size-fits-all solution, thus, while the model's framework can be transferred, its parameters and inputs need to be adjusted to suit local conditions for accurate results.

5.2. Air Temperature Variability at The Local Scale

Several studies on microclimate allude to significant variability in temperature even at a local scale (Li et al., 2023a; Stewart & Oke, 2012). Through this study, this variability is analysed within the context of deprived urban areas.

From the study, it was revealed statistically significant differences within and between deprived and nondeprived urban areas. The highest temperatures were consistently recorded in all DUA's, whilst the opposite was true for non-deprived areas. In all the settlements, except Kibera, the median temperatures of DUA's were between 0.6°C to 1.2 °C higher than their non DUA counterparts, with minimum temperatures in DUA's consistently lower or same to non DUA's. Spatially, this was reflected in the clustering of higher temperatures in central areas of the settlements, relative to the peripheral areas that tend to be non DUA. These findings are Similar to the work Scott et al., (2017), that reveals general warmer temperatures of about 1.8°C difference within informal settlements of Nairobi relative to neighbouring formal areas. The study highlights that this variability may partially be explained by differences in surface characteristics.

Expectedly, the same pattern was observed across the morphological clusters, revealing higher temperatures within morphological areas associated with DUA's and vice versa. Specifically, much of the hotspots were concentrated in clusters 0, the lowest temperatures on the other hand were populated in cluster 1(see Appendix 12). Comparing the mean temperatures across clusters, it was established, that despite the varied types of nonDUAS considered in the study, they consistently exhibited lower temperatures compared to their DUA counterparts.

Within the study area alone, these areas included gated residential communities (1), industrial areas (2), buildings within informal areas but with less deprivation (3), seen in the figures below. As such, the variance in temperature values is reflected in this diversity of morphology.

Figure 23Gated residential NonDUA in Waruku: Sateliteview view (in redframe) , (in greenframe): street view. Images obtained from Google earth Pro (Acquired 14/02/2024)

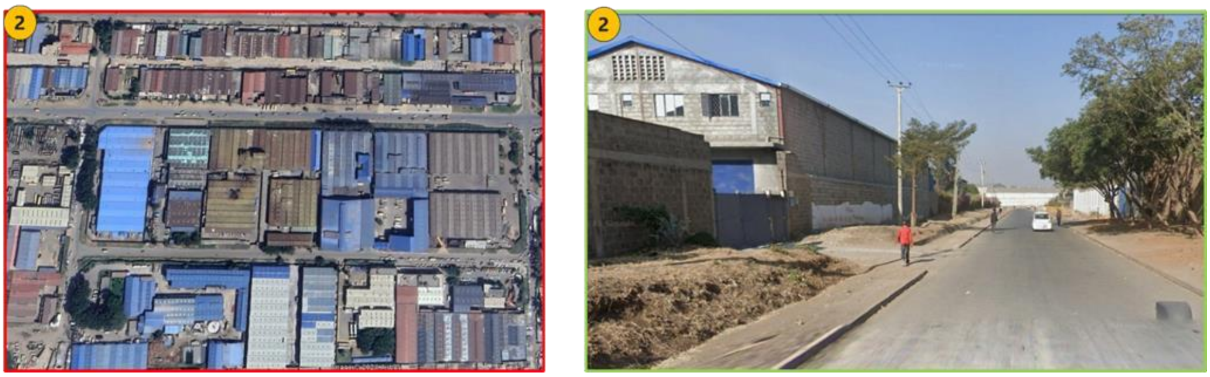


Figure24:Industrial NonDUA in Mukuru: Sateliteview view (in redframe) , (in greenframe): street view. Images obtained from Google earth Pro (Acquired 14/02/2024)

Figure 25: Less deprived within in informal area Kibera: Sateliteview view (in redframe) , (in greenframe): street view. Images obtained from Google earth Pro (Acquired 14/02/2024)

DUAs on the other hand tend to have similar physical characteristics rooted in shared experiences of multifaceted socio-economic challenges (Abascal et al., 2022; Isunju et al., 2016). These areas are often plagued by poverty and located in high-cost urban regions where access to affordable housing is limited. As a result, residents frequently resort to makeshift structures and overcrowded living conditions, with little or no greenery and open spaces. This reflects the economic struggles of poverty, unemployment, leading to a distinct and similar urban landscape characterized by substandard living conditions. This is reflected in the low separability and cophenetic distances of the two DUA classes (Appendix7). A summary of the two main morphological clusters as well as mean predicted temperatures is presented in the figure below.

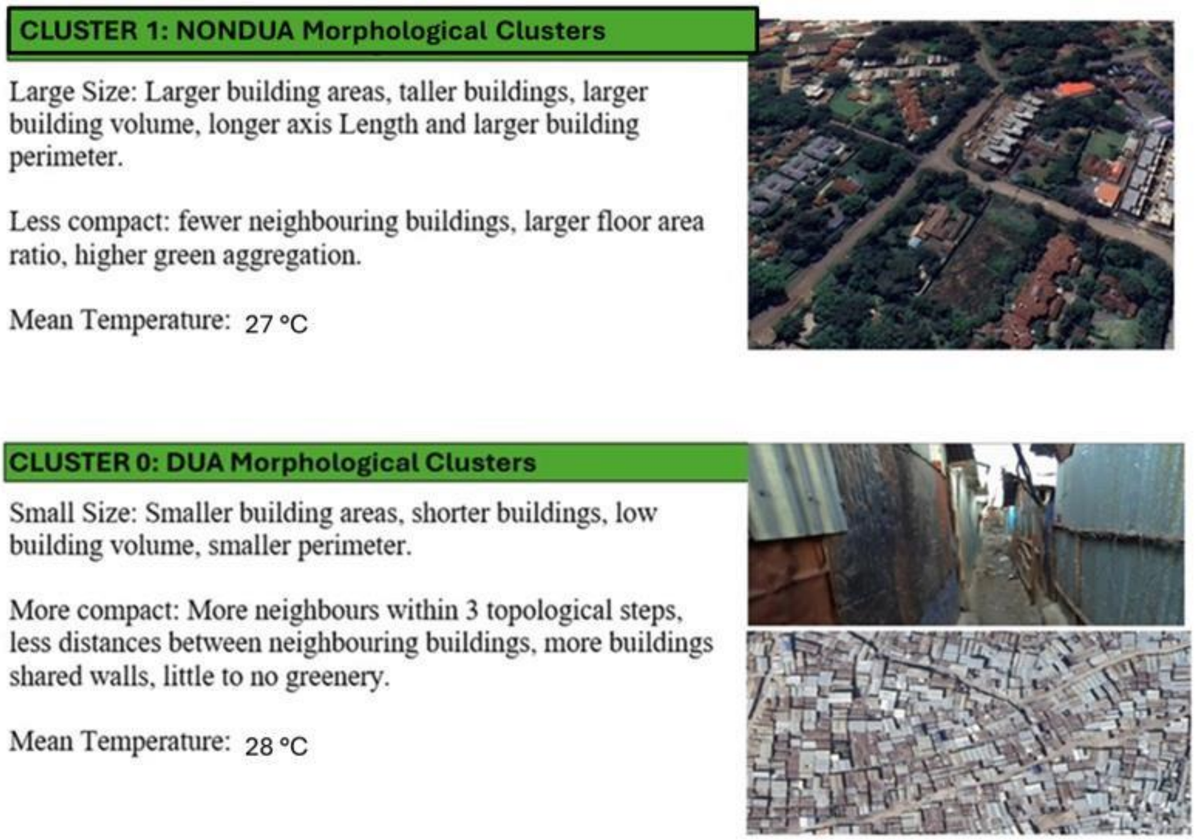


Figure 26: Derived cluster description

The findings from this study underscore significant thermal inequality within the informal settlements under investigation. The statistically significant differences in air temperature between deprived urban areas (DUAs) and non-deprived urban areas (non-DUAs) highlight how socio-economic factors and urban morphology converge to create disparate thermal environments. DUAs consistently recorded higher median temperatures compared to non-DUAs, reflecting the compounded effects of limited green spaces, overcrowded living conditions, and substandard housing structures that lack proper ventilation. This disparity is further exacerbated by the presence of taller buildings and larger building volumes in nonDUAs, which contribute to localized cooling through shading and altered wind patterns.

The spatial distribution of temperature hotspots within the settlements indicates that central areas, typically characterized by higher building densities and fewer green spaces, experience the highest temperatures. This spatial pattern aligns with morphological clusters associated with DUAs, where the lack of greenery and open spaces is most pronounced. These areas also exhibit less variability in temperature, suggesting a more stable yet persistently higher thermal environment. In contrast, non-DUAs show greater variance in temperature values, reflective of the diverse morphological features within these areas, including gated communities, industrial zones, and less-deprived segments of informal settlements. This morphological diversity contributes to a wider range of thermal conditions, as different urban forms interact uniquely with microclimatic factors.

5.3. Important Influences to Temperature Variation

The results in 5.4 imply elevation to be the most important feature to air temperature estimation. This was seen in all the model variations, though with differing magnitudes. This is likely attributed both models biases as well as the features predictive power. The random forest regression algorithm at each node, decides splits based on the feature that maximally reduces variance of the target variable within a randomly selected (bootstrap and bagging) subset of the data. Though this randomness is what reduces overfitting, it could potentially lead to biased selection of features. Thus, though unlikely, it is highly possible to consistently select the same set of features whilst other features never get chosen, leading to some information loss during the training process (Breiman, 2001). Aside from this, the splitting process can introduce biases, towards higher cardinality data such as elevation as they provide a unique value that provide consistent oneon-one relationship to the target. By providing distinct splitting opportunities, Site

On the other hand, elevation in itself tends to have a relevant predictive power in relation to temperature, especially in the absence of critical physics parameters. Several literatures exist on the study on the temperature lapse rate, which the describes the of temperature values with altitude. Due to the coupling effects of both adiabatic cooling and atmospheric pressure, generally, temperature tends to reduce with increasing elevation (Kattel et al., 2018). This was observed in this study, with the settlements in relatively higher elevations, such as Waruku, consistently with reduced temperature as recounted by Janatian et al., (2017), who also explains that elevation tends to include the spatial variability effects of climatological factors. This goes a long way to increase accuracies of statistical models which are not purely based on the energy-balance physics interactions. This gives a basis to the heavy loading of elevation in the model, as well as an explanation as to the failures of the scenario models, which saw drastic drops in accuracies, with the removal of elevation as a feature.

Following the plots in 5.4.5 elevation had the highest correlation with temperature, with a negative strong correlation score -0.9, indicating that temperature is highly dependent on elevation and reduces with increasing altitudes. It is thereby a reliable predictor to temperature over vertical distances.

Land cover indices have been employed as auxiliary variables by most air temperature estimation studies as they present the potential to explain variant surface types and their interactions with solar radiation, heat retention, and atmospheric processes. In this study, expectedly, all land cover indices used contributed significantly to the study and resulting in the top 10 most informative variable to air temperature. LST specifically, (day and night) consistently ranked top 5 variables amongst all the variants models, with an observed strong positive correlation to air temperature similarly observed by (Benali et al., 2012; Janatian et al., 2017; Zhang et al., 2016a). LST measures the heat emitted from the earth's surface, influenced by solar radiation, atmospheric conditions, and land cover. This direct surface heating significantly impacts the air temperature above it. During the day, the earth surfaces absorb solar energy, increasing in temperature and transferring this heat to the air through conduction and convection, thus elevating the air temperature. This phenomenon is particularly pronounced and more complex in urban areas, where materials like concrete and asphalt, with high heat capacities and low albedo, retain and slowly release heat, leading to a much complex relationship of LST with air temperature (Janatian et al., 2017).

The auxiliary variables explained above exhibited the most influence on air temperature prediction, however, the compounding efforts of morphometrics was substantial in explaining a relevant measure of variance. The impacts of these morphometrics are explained in this section.

Preliminary analysis using Principal component analysis revealed that confounding relationships that exist within the dataset, thus the variance in the dataset could not be explained by any 1 component. This was further in the sensitivity analyses, that showed the significant drop in R^2 by modelling and predicting air temperature without morphometrics, despite the low feature important scores of the individual metrics. This presents insights into the compounding importance of morphology as a whole. This suggests that individually, each metric may not necessarily influence air temperature very significantly, but with the interactions with other morphometrics as well as possibly auxiliary variables, the significant of morphometrics is enhanced. Thus, complex confounding relationship of morphometrics exist with air temperature, signifying the weak correlation scores of important morphological variables. This further explains the findings of other studies such as Kotharkar et al., (2023, Li et al., (2023a) and Stewart & Oke, (2012), that all find higher correlation of urban morphology to air temperature. All these studies do not us detailed measurements of morphometrics as done in this study. Morphological parameters are often limited to area-based and compound metrics such as green cover ratio, surface ratio etc similar to the local climate zone studies. The morphological metrics, usually much dependent on land cover indices, tend to have a more simple and linear correlation with air temperature. Thus, with such coarse morphological metrics, it is feasible to establish such direct association with air temperature, which further supports the findings of this study. To the best of the authors knowledge, no studies exist on air temperature predictive modelling with Morphometrics, which focuses on the detailed characterisation of building geometries.

Though not always, as Peng et al., (2022) finds that 3D morphological variables tend to almost always have a significance to air temperature though a much more varied and complex relationship. Li et al., (2023b), confirms this finding by explaining that building height, tends to have a much complex relationship with air temperature, explaining that increase in building height within low- and high-rise buildings, increased local temperatures, however, the opposite was true for multistorey and mid-rise buildings.

Similarly, in this study, the most important features to temperature were metrics associated with vertical development. For example, mean height building as well as building volume ranked among the top 10 most important features. Building heights relationship in this study was not very informative as a result of the limited diversity in height values within the study are. This is associated to the height data spatial scale of 100 metres, resulting in the aggregation and lack of detailed analysis in understanding, the relationship. Overall, the linear relationship indicated that buildings with lower volume and height, are associated with higher temperatures. However, all the few cases of relatively tall buildings above 20m were associated with between 29-30 °C of temperature. The same pattern was observed with building volume. This gives a glimpse into the complexity of 3D morphological relationships. This can be explained by the shade effect created by taller buildings that reduce the amount of direct solar radiation reaching the ground and lowerlevel structures, resulting in cooler temperatures, however, in other cases where there is more compactness and dense built up, taller buildings can alter wind patterns by blocking or channelling airflow, which affects the distribution of heat, leading to reduced ventilation and cooling (Li et al., 2023c; Puche et al., 2023; Yu et al., 2020). This varied relationships as well as data limitations contributed to limited understanding in the assessment of 3D metrics in this study.

Another key finding of this study was the importance of contextual or neighbourhood level metrics to air temperature. All the metrics that were consistently selected across all the models as important features were contextual in nature. Thus, they were computed across relative neighbourhood levels, within three topological steps of each other. They constitute the tendencies and diversity of metrics within a local neighbourhood scale. This implied that, though each buildings morphology may not necessarily be very influential in impacting temperature a combined efforts of neighbouring buildings characteristics instead influences temperatures significantly.

5.4. Limitations and recommendations

Several limiting factors were encountered in the study. This section elaborates on these challenges and potential effects on the results.

The WSF 3D despite being the highest resolution open dataset on building heights, at a resolution of 100 meters, it was still relatively coarse for the objectives of this study. This granularity is makes it difficult for capturing the fine-scale variations in building heights within the informal settlements. This leads to the aggregation of diverse height values, masking the detailed variations that can significantly influence local temperature patterns. For instance, the presence of tall buildings interspersed with low-rise structures creates microclimates that a coarse dataset cannot accurately represent. This limitation affects the model's ability to capture the complex interactions between building heights, urban morphology, and air temperature. The models thus overlooked critical micro-scale thermal variations. This is particularly limiting giving the extensive literature on the importance of heights in thermal variations.
Al-Mudhaffer et al., (2022) establishes the importance of building construction materials especially in regulating temperature. Despite the established knowledge that thermal conditions are significantly impacted by construction materials, the lack of available open datasets for this data is lacking. The study therefore did not extensively include data on building materials, which is a significant factor influencing heat dynamics, especially in informal settlements (Mehrotra et al., 2018), as different materials have different thermal properties, such as concrete, metal, or thatch and interact differently with solar radiation, affecting heat retention and dissipation. The absence of this meant that cannot account for the variations in heat absorption and emission characteristics of different structures, leading to less accurate temperature predictions especially in relation to absolute temperature.
Although the remotely sensed data used in the study acquired within the summer period, it did not coincide with the actual days or weeks of data collection. The sentinel and ECOSTRESS LST images were acquired in December 2023 and March 2024 whilst the data collection period was January and February. This temporal mismatch introduces uncertainties in correlating surface characteristics observed from satellite imagery with ground-based real-time temperature readings and influences the magnitude of generalisation that could be derived from results.
Similarly, the building footprint data used in this study was not based on current imagery, though it is the most current up-to date dataset on building footprints. This is a significant limitation given the dynamic nature of informal settlements, which tends to undergo rapid changes. Using outdated footprint data results may result in a mismatch between the actual morphological characteristics and measured characters leading to inaccuracies in capturing the true urban form and its impact on temperature patterns.
Lastly, the aggregation of air temperature data to 50 meters resulted in the loss of fine-scale details, crucial for understanding microclimatic variations within informal settlements. Despite the need for this aggregation for maximal morphometric information gain, as explained above, the aggregation smooths out localized extremes and nuances, leading to generalized temperature patterns that may overlook significant small-scale variations. This level of aggregation is less effective in capturing the heterogeneity of informal settlements, where short distances can exhibit substantial temperature differences due to varying building materials, vegetation cover, and urban morphology.

5.5. Uncertainties from citizen science data

Human influences are a primary source of uncertainty in citizen science data collection(Downs et al., 2021). Despite training, participants sometimes deviated from instructions, such as holding sensor sticks at varying heights, leading to inconsistencies in temperature readings. Physical contact with the sensors also posed a problem, as the exposed Kestrel sensors could register body temperatures instead of air temperatures, causing sudden spikes and drops in the readings. While the Kestrel sensor's sensitivity to small changes is beneficial, it also makes it susceptible to such interferences, complicating the interpretation of true air temperature readings.
Sensor inefficiencies further contribute to data uncertainties. The lack of a protective shield for the Kestrel sensor exposed it to direct sunlight, potentially causing artificially high readings. Citizens were instructed to place fixed sensors in shaded areas and away from metal materials to mitigate this issue. However, verifying the adherence to these instructions and the physical conditions of the sensor locations remains challenging. Despite participants' reports of compliance, the accuracy of the fixed sensor readings is uncertain, highlighting the need for more robust measures to ensure data reliability in future studies.

5.6. Random Forest Model Limitations

The use of Random Forest regression in this study, while effective for handling complex datasets, tends to smooth out predictions. This smoothing effect reduces the model's sensitivity to extreme temperature values, potentially overlooking localized hotspots or areas with significantly lower temperatures. Random Forest regression operates by averaging multiple decision trees, which can dampen the impact of outliers and extreme values, leading to more generalized predictions (Breiman, 2001).

While smoothing can be beneficial for overall model stability, it limits the model's ability to capture sharp temperature gradients and microclimatic variations. In informal settlements, where small-scale variations can significantly impact residents' thermal comfort, this limitation is crucial. The model's tendency to generalize results may obscure important localized thermal phenomena that are critical for targeted interventions.

6. CONCLUSION

This study focused on exploring local-scale temperature variations within informal settlements, emphasizing the influence of urban morphological factors on air temperature. To achieve this three objectives and research questions were outlined. The first section of this chapter highlights how the findings of this research answers the outlined questions.

The first research question concerns determining if there are morphological patterns observed within and between informal and formal areas and what morphological characteristics are key in distinguishing them. To this effect, this study utilised momepy toolkit in measuring detailed descriptions of some 117 primary

and contextual characters across 5 informal settlements within Nairobi. The k-means clustering analysis revealed two distinct morphological groups within the study area. Further evaluations revealed an intersected with reference formal and informal areas, indicating that indeed formal areas are morphologically unique to informal areas. The findings highlighted that those morphological characters related to size and compactness such as building perimeter, building area, building volume, mean height, mean interbuilding distance and number of neighbours were particularly different amongst these two areas.
In answering the research question on spatial patterns of air temperature observed within and across informal and formal settlements, this study revealed, first, that there are thermal differences across the informal settlements. Particularly, Kibera and Waruku tend to experience relatively cooler temperatures than the other informal settlements. This was found to be associated with the fact that they are located in higher elevation and thus the temperature lapse rate influences the thermal conditions. The study also revealed significant temperature differences between formal and informal areas with about 0.5°C to 1.2°C higher median temperatures within informal settlements as compared to their formal counterparts.
Finally, the last research question pertaining to the accuracy of the Air temperature model and key morphological characteristics that contribute significantly to the temperature variations, the final air temperature predictive model successfully predicted air temperature across the study area with an R^2 of 0.7 thus, it could efficiently explain about 70% of the variance in air temperature. Using feature importance elevation was identified to be the most important variable. This study found that, morphometrics played only a supplementary role in the prediction. Various analysis such as the sensitivity analysis and PCA confirmed that despite the fact that morphometrics contributes to air temperature variations, it may not necessarily be the most influencing factor.
The most influencing morphometric character to temperature variation however were features relating to height. Specifically, mean building height and building volume at the neighbourhood level contributed the most to temperature variations.

The findings of the study underscored significant temperature disparities between deprived urban areas (DUAs) and non-deprived urban areas (non-DUAs). Lessons learned from this study implies, that urban planners and civil society organisations interested in improving thermal conditions of informal settlements, adopt a holistic approach that considers the entire urban fabric, including greenery, building materials, and neighbourhood context, beyond just building geometries. This comprehensive perspective is crucial in targeted interventions to mitigate climate-related risks.

6.1. Ethical Considerations

Several aspects of this study may raise ethical concerns, particularly regarding its potential influence on policy decisions. The detailed physical characterization of informal settlements in relation to morphometrics, as well as the identification of hazard-related differences between formal and informal areas, may pose risks to community privacy. While the aggregation of data at a 50-meter resolution may buffer some details, it is crucial to acknowledge the potential misuse of the study's findings. There is a risk that the data could be used to segregate and target informal settlements for evictions and slum destruction, which would defeat the purpose and aim of this research.
The primary objective of this study is to provide insights into the conditions within informal settlements to support and enhance slum upgrading efforts. Therefore, it is essential to recognize the distinction between association and causation. While this study offers valuable insights into temperature variability through its association with urban morphology, it does not establish direct causation. Consequently, any policy influence should be approached cautiously to avoid drawing causal inferences without further corroborative studies.
It is also important to consider the inherent biases associated with any modelling technique. Choices related to model selection, hyperparameters, and feature selection can significantly influence model predictions. Specifically, for the random forest algorithm used in this study, these biases can impact the interpretation and findings of the results. As discussed in Section 5.3, such biases must be acknowledged and addressed to ensure the validity and reliability of the conclusions drawn from the model.

While this study seeks to inform and improve slum upgrading efforts, it is critical to handle the findings responsibly and ethically to ensure that the primary goal of enhancing the quality of life in informal settlements is achieved.

6.2. Recommendations For Further Work

Though this study provided relevant insights into the interactions of urban morphometrics with air temperature and successfully estimated temperature variations within informal areas, more insights is still needed concerning the comprehensive characterisation of morphology in informal settlements. Recommendations are provided below that can help build upon the work that has been done in this study.

- Exploring the integration of physical models such as energy balance models and datasets with statistical models such as the random forest model used in this study, will further increase accuracy and the generalisability of temperature estimates. This will ensure a healthy balance in capturing complex associations of surface characteristics whilst modelling the underlying physical interactions. This will go a step beyond this study as highlighted previously, the importance of physical parameters to air temperature estimation.
- The aggregation of insitu temperature data to 50metres resulted in a less detailed analysis, though it was necessary to capture more morphological information. Further research can investigate this study at a point scale and relate point measurements to surrounding morphometric characteristics within a specified buffer. Apart from allowing a much-detailed analysis, such a study can also associate the influence of urban morphometric to air temperature within varying distances.
- Finally, improved datasets and covariates, such as building materials, higher resolution height dataset and temporally aligned covariates will go a long way to reduce biases by ensuring a more realistic characterisation of covariates thereby improving accuracy of predictions.

LIST OF REFERENCES

Abascal, A., Rodríguez-Carreño, I., Vanhuysse, S., Georganos, S., Sliuzas, R., Wolff, E., & Kuffer, M. (2022). Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Computers, Environment and Urban Systems*, *95*, 101820. https://doi.org/10.1016/J.COMPENVURBSYS.2022.101820

Al-Mudhaffer, A. F., Saleh, S. K., & Kadhum, G. I. (2022). The role of sustainable materials in reducing building temperature. *Materials Today: Proceedings*, *61*, 690–694. https://doi.org/10.1016/J.MATPR.2021.08.249

Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, *124*, 108–121. https://doi.org/10.1016/J.RSE.2012.04.024 Breiman,

L. (2001). *Random Forests*. *45*, 5–32.

Cao, Q., Huang, H., Hong, Y., Huang, X., Wang, S., Wang, L., & Wang, L. (2022). Modeling intra-urban differences in thermal environments and heat stress based on local climate zones in central Wuhan. *Building and Environment*, *225*, 109625. https://doi.org/10.1016/J.BUILDENV.2022.109625

Carranza, C., Nolet, C., Pezij, M., & van der Ploeg, M. (2021). Root zone soil moisture estimation with Random Forest. *Journal of Hydrology*, *593*, 125840. https://doi.org/10.1016/J.JHYDROL.2020.125840

Egondi, T., Kyobutungi, C., Kovats, S., Muindi, K., Ettarh, R., & Rocklöv, J. (2012). Time-series analysis of weather and mortality patterns in Nairobi's informal settlements. *Https://Doi.Org/10.3402/Gha.V5i0.19065*, *5*(SUPPL.), 23–32. https://doi.org/10.3402/GHA.V5I0.19065

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2020). Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale. *Computers, Environment and Urban Systems*, *80*, 101441. https://doi.org/10.1016/J.COMPENVURBSYS.2019.101441

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2022). Methodological foundation of a numerical taxonomy of urban form. *Environment and Planning B: Urban Analytics and City Science*, *49*(4), 1283–1299. https://doi.org/10.1177/23998083211059835/SUPPL_FILE/SJ-PDF-1-EPB-10.1177_23998083211059835.PDF

Fung, W. Y., Lam, K. S., Nichol, J., & Wong, M. S. (2009). Derivation of Nighttime Urban Air Temperatures Using a Satellite Thermal Image. *Journal of Applied Meteorology and Climatology*, *48*(4), 863–872. https://doi.org/10.1175/2008JAMC2001.1

Gholamnia, M., Alavipanah, S. K., Darvishi Boloorani, A., Hamzeh, S., & Kiavarz, M. (n.d.). *remote sensing Diurnal Air Temperature Modeling Based on the Land Surface Temperature*. https://doi.org/10.3390/rs9090915

Han, J., Kamber, M., & Pei, J. (2012). Cluster Analysis. *Data Mining*, 443–495. https://doi.org/10.1016/B9780-12-381479-1.00010-1

He, Y., Chen, C., Li, B., & Zhang, Z. (2022). Prediction of near-surface air temperature in glacier regions using ERA5 data and the random forest regression method. *Remote Sensing Applications: Society and Environment*, *28*, 100824. https://doi.org/10.1016/J.RSASE.2022.100824

IPCC. (2023). *AR6 Synthesis Report: Climate Change 2023*. https://www.ipcc.ch/report/ar6/syr/

Isunju, J. B., Orach, C. G., & Kemp, J. (2016). Hazards and vulnerabilities among informal wetland communities in Kampala, Uganda. *Environment and Urbanization*, *28*(1), 275–293. https://doi.org/10.1177/0956247815613689

Janatian, N., Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *International Journal of Climatology*, *37*(3), 1181–1194. https://doi.org/10.1002/JOC.4766

Janatian Nasime, Sadeghi, M., Sanaeinejad, S. H., Bakhshian, E., Farid, A., Hasheminia, S. M., & Ghazanfari, S. (2017). A statistical framework for estimating air temperature using MODIS land surface temperature data. *INTERNATIONAL JOURNAL OF CLIMATOLOGY Int. J. Climatol*, *37*, 1181–1194. https://doi.org/10.1002/joc.4766

Kattel, D. B., Yao, T., & Panday, P. K. (2018). Near-surface air temperature lapse rate in a humid mountainous terrain on the southern slopes of the eastern Himalayas. *Theoretical and Applied Climatology*, *132*(3–4), 1129–1141. https://doi.org/10.1007/S00704-017-2153-2

Keung Tsin, P., Knudby, A., Krayenhoff, E. S., Brauer, M., & Henderson, S. B. (2020). *Land use regression modeling of microscale urban air temperatures in greater Vancouver, Canada*. https://doi.org/10.1016/j.uclim.2020.100636

Kisters, P., Ngu, V., & Edinger, J. (2022). Urban Heat Island Detection Utilizing Citizen Science. *Communications in Computer and Information Science*, *1617 CCIS*, 94–98. https://doi.org/10.1007/978-3-03123298-5_9/FIGURES/2

Kjellstrom, T., Freyberg, C., Lemke, B., Otto, M., & Briggs, D. (2018). Estimating population heat exposure and impacts on working people in conjunction with climate change. *International Journal of Biometeorology*, *62*(3), 291–306. https://doi.org/10.1007/S00484-017-1407-0

Koppe, Christina., Kovats, Sari., Jendritzky, Gerd., Menne, Bettina., Breuer, D. J., Deutscher Wetterdienst., London School of Hygiene and Tropical Medicine., & European Commission. Energy, E. and S. Development. (2004). *Heat waves : risks and responses*. 124.

Kotharkar, R., Dongarsane, P., & Keskar, R. (2023). Determining influence of urban morphology on air temperature and heat index with hourly emphasis. *Building and Environment*, *233*, 110044. https://doi.org/10.1016/J.BUILDENV.2023.110044

Li, X., Yang, B., Liang, F., Zhang, H., Xu, Y., & Dong, Z. (2023a). Modeling urban canopy air temperature at city-block scale based on urban 3D morphology parameters– A study in Tianjin, North China. *Building and Environment*, *230*, 110000. https://doi.org/10.1016/J.BUILDENV.2023.110000

Li, X., Yang, B., Liang, F., Zhang, H., Xu, Y., & Dong, Z. (2023b). Modeling urban canopy air temperature at city-block scale based on urban 3D morphology parameters-A study in Tianjin, North China. *Building and Environment*, *230*, 110000. https://doi.org/10.1016/j.buildenv.2023.110000

Li, X., Yang, B., Liang, F., Zhang, H., Xu, Y., & Dong, Z. (2023c). Modeling urban canopy air temperature at city-block scale based on urban 3D morphology parameters-A study in Tianjin, North China. *Building and Environment*, *230*, 110000. https://doi.org/10.1016/j.buildenv.2023.110000

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. https://doi.org/10.1016/J.BUILDENV.2022.109910

Liu, L., Lin, Y., Liu, J., Wang, L., Wang, D., Shui, T., Chen, X., & Wu, Q. (2017). Analysis of local-scale urban heat island characteristics using an integrated method of mobile measurement and GIS-based spatial interpolation. *Building and Environment*, *117*, 191–207. https://doi.org/10.1016/J.BUILDENV.2017.03.013

Liu, Z., Anderson, B., Yan, K., Dong, W., Liao, H., & Shi, P. (2017). Global and regional changes in exposure to extreme heat and the relative contributions of climate and population change. *Scientific Reports 2017 7:1*, *7*(1), 1–9. https://doi.org/10.1038/srep43909

Mahabir, R., Agouris, P., Stefanidis, A., Croitoru, A., & Crooks, A. T. (2020). Detecting and mapping slums using open data: a case study in Kenya. *International Journal of Digital Earth*, *13*(6), 683–707. https://doi.org/10.1080/17538947.2018.1554010

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Mehrotra, S., Bardhan, R., & Ramamritham, K. (2018). Urban Informal Housing and Surface Urban Heat Island Intensity: Exploring Spatial Association in the City of Mumbai. *Environment and Urbanization ASIA*, *9*(2), 158–177. https://doi.org/10.1177/0975425318783548

Morrow, B. H. (2008). Community Resilience: A Social Justice Perspective. *CARRI Initiative: Oak Ridge National Laboratory*. https://www.academia.edu/4313849/Community_Resilience_A_Social_Justice_Perspective

Nag, P., Nag, A., … P. S.-N. I. of, & 2009, undefined. (2009). Vulnerability to heat stress: Scenario in western India. *Danida.Vnu.Edu.Vn*. http://danida.vnu.edu.vn/cpis/files/Refs/Heat%20Waves/Heat_Stress/VULNERABILITY%20TO%20HEAT%20STRESS-%20SCENARIO%20IN%20WESTERN%20INDIA%20.pdf

Nascetti, A., Monterisi, C., Iurilli, F., & Sonnessa, A. (2022). A NEURAL NETWORK REGRESSION MODEL FOR ESTIMATING MAXIMUM DAILY AIR TEMPERATURE USING LANDSAT-8 DATA. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B3-2022), 1273–1278. https://doi.org/10.5194/ISPRS-ARCHIVES-XLIII-B3-2022-12732022

NCCG. (2020). *Nairobi City County Climate Action Plan 2020-2050 | Nairobi City County*. https://nairobi.go.ke/download/nairobi-city-county-climate-action-plan-2020-2050/

Ochola, E. M., Fakharizadehshirazi, E., Adimo, A. O., Mukundi, J. B., Wesonga, J. M., & Sodoudi, S. (2020). Inter-local climate zone differentiation of land surface temperatures for Management of Urban Heat in Nairobi City, Kenya. *Urban Climate*, *31*, 100540. https://doi.org/10.1016/J.UCLIM.2019.100540

Onačillová, K., Gallay, M., Paluba, D., Péliová, A., Tokarčík, O., & Laubertová, D. (2022). Combining Landsat 8 and Sentinel-2 Data in Google Earth Engine to Derive Higher Resolution Land Surface Temperature Maps in Urban Environment. *Remote Sensing 2022, Vol. 14, Page 4076*, *14*(16), 4076. https://doi.org/10.3390/RS14164076

Peng, W., Wang, R., Duan, J., Gao, W., & Fan, Z. (2022). Surface and canopy urban heat islands: Does urban morphology result in the spatiotemporal differences? *Urban Climate*, *42*, 101136. https://doi.org/10.1016/J.UCLIM.2022.101136

Puche, M., Vavassori, A., & Brovelli, M. A. (2023). Insights into the Effect of Urban Morphology and Land Cover on Land Surface and Air Temperatures in the Metropolitan City of Milan (Italy) Using Satellite Imagery and In Situ Measurements. *Remote Sensing 2023, Vol. 15, Page 733*, *15*(3), 733. https://doi.org/10.3390/RS15030733

Purevdorj, T. S., Tateishi, R., Ishiyama, T., & Honda, Y. (1998). Relationships between percent vegetation cover and vegetation indices. *International Journal of Remote Sensing*, *19*(18), 3519–3535. https://doi.org/10.1080/014311698213795

Qi, J., Yu, Y., Wang, L., & Liu, J. (2016). K*-means: An effective and efficient k-means clustering algorithm. *Proceedings - 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*, 242–249. https://doi.org/10.1109/BDCLOUD-SOCIALCOM-SUSTAINCOM.2016.46

Ramsay, E. E., Duffy, G. A., Burge, K., Taruc, R. R., Fleming, G. M., Faber, P. A., & Chown, S. L. (2023). Spatio-temporal development of the urban heat island in a socioeconomically diverse tropical city. *Environmental Pollution*, *316*, 120443. https://doi.org/10.1016/J.ENVPOL.2022.120443

Rathi, S. K., Desai, V. K., Jariwala, P., Desai, H., Naik, A., & Joseph, A. (2017). Summer temperature and spatial variability of all-cause mortality in Surat city, India. *Indian Journal of Community Medicine*, *42*(2), 111– 115. https://doi.org/10.4103/0970-0218.205216

Reisinger, A., Howden, M., Vera, C., Garschagen, M., Hurlbert, M., Kreibiehl, S., Mach, K. J., Mintenbeck, K., O'neill, B., Pathak, M., Pedace, R., Pörtner, H.-O., Poloczanska, E., Rojas Corradi, M., Sillmann, J., Van Aalst, M., Viner, D., Jones, R., Ruane, A. C., & Ranasinghe, R. (n.d.). *The concept of risk in the IPCC Sixth Assessment Report The concept of risk in the IPCC Sixth Assessment Report: a summary of cross-Working Group discussions Guidance for IPCC authors The concept of risk in the IPCC Sixth Assessment Report The concept of risk in the IPCC Sixth Assessment Report: a summary of cross-Working Group discussions Guidance for IPCC authors The concept of risk in the IPCC Sixth Assessment Report Contents*.

Ren, H., Guo, W., Zhang, Z., Kisovi, L. M., & Das, P. (n.d.). *Population Density and Spatial Patterns of Informal Settlements in Nairobi, Kenya*. https://doi.org/10.3390/su12187717

Scott, A. A., Misiani, H., Okoth, J., Jordan, A., Gohlke, J., Ouma, G., Arrighi, J., Zaitchik, B. F., Jjemba, E., Verjee, S., & Waugh, D. W. (2017). Temperature and heat in informal settlements in Nairobi. *PLOS ONE*, *12*(11), e0187300. https://doi.org/10.1371/JOURNAL.PONE.0187300

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Salah, Y., Bouchareb, E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). *Continental-Scale Building Detection from High Resolution Satellite Imagery*. https://arxiv.org/abs/2107.12283v2

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, *93*(12), 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

UN-HABITAT. (2005). *NAIROBI URBAN SECTOR PROFILE Rapid Urban Sector Profiling for Sustainability (RUSPS) UN-HABITAT United Nations Human Settlements Programme*. http://www.unhabitat.org

United Nations Environment Programme. (2003). *Impacts of Summer 2003 Heat Wave in Europe - Environment Alert Bulletin 2*. https://wedocs.unep.org/xmlui/handle/20.500.11822/40942

Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., & Porta, S. (2023). EO + Morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning*, *233*. https://doi.org/10.1016/j.landurbplan.2023.104691

Wang, J., Kuffer, M., Sliuzas, R., & Kohli, D. (2019). The exposure of slums to high temperature: Morphologybased local scale thermal patterns. *Science of The Total Environment*, *650*, 1805–1817. https://doi.org/10.1016/J.SCITOTENV.2018.09.324

Wang, K., Sun, J., Cheng, G., & Jiang, H. (2011). Effect of altitude and latitude on surface air temperature across the Qinghai-Tibet Plateau. *Journal of Mountain Science*, *8*(6), 808–816. https://doi.org/10.1007/S11629-011-1090-2/METRICS

World Cities Report 2022. (n.d.). Retrieved September 4, 2023, from https://unhabitat.org/wcr/

Wu, W., Ren, H., Yu, M., & Wang, Z. (2018). Distinct Influences of Urban Villages on Urban Heat Islands: A Case Study in the Pearl River Delta, China. *International Journal of Environmental Research and Public Health 2018, Vol. 15, Page 1666*, *15*(8), 1666. https://doi.org/10.3390/IJERPH15081666

Yang, D., Zhong, S., Mei, X., Ye, X., Niu, F., & Zhong, W. (2023). A Comparative Study of Several Popular Models for Near-Land Surface Air Temperature Estimation. *Remote Sensing 2023, Vol. 15, Page 1136*, *15*(4), 1136. https://doi.org/10.3390/RS15041136

Yang, Y. Z., Cai, W. H., & Yang, J. (2017). Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. *Remote Sensing*, *9*(5). https://doi.org/10.3390/RS9050410

Yoon, D. K. (2012). Assessment of social vulnerability to natural disasters: A comparative study. *Natural Hazards*, *63*(2), 823–843. https://doi.org/10.1007/S11069-012-0189-2/TABLES/8

Yu, Z., Chen, S., Wong, N. H., Ignatius, M., Deng, J., He, Y., & Hii, D. J. C. (2020). Dependence between urban morphology and outdoor air temperature: A tropical campus study using random forests algorithm. *Sustainable Cities and Society*, *61*. https://doi.org/10.1016/J.SCS.2020.102200

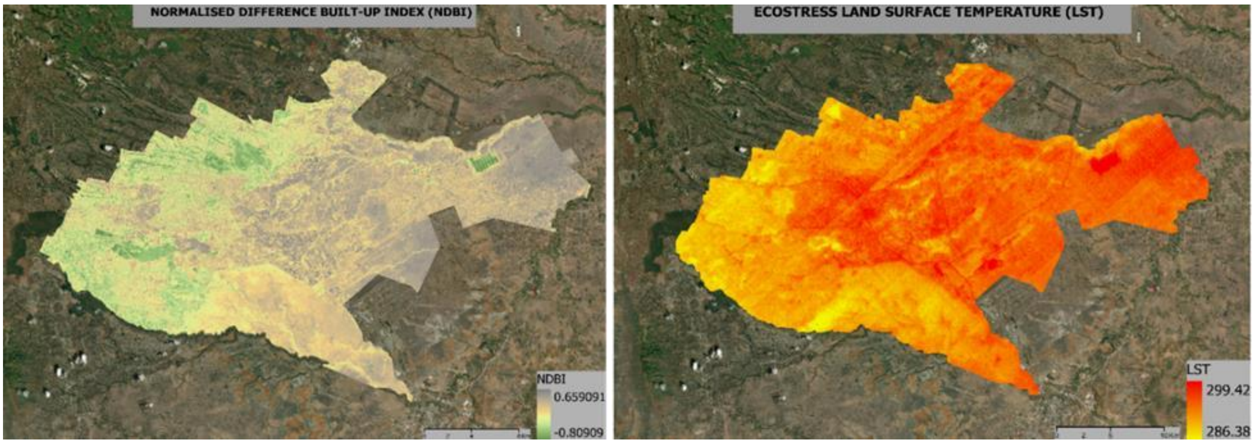
Zha, Y., Gao, J., & Ni, S. (2010). *International Journal of Remote Sensing Use of normalized difference built-up index in automatically mapping urban areas from TM imagery Use of normalized difference built-up index in automatically mapping urban areas from TM imagery*. https://doi.org/10.1080/01431160304987

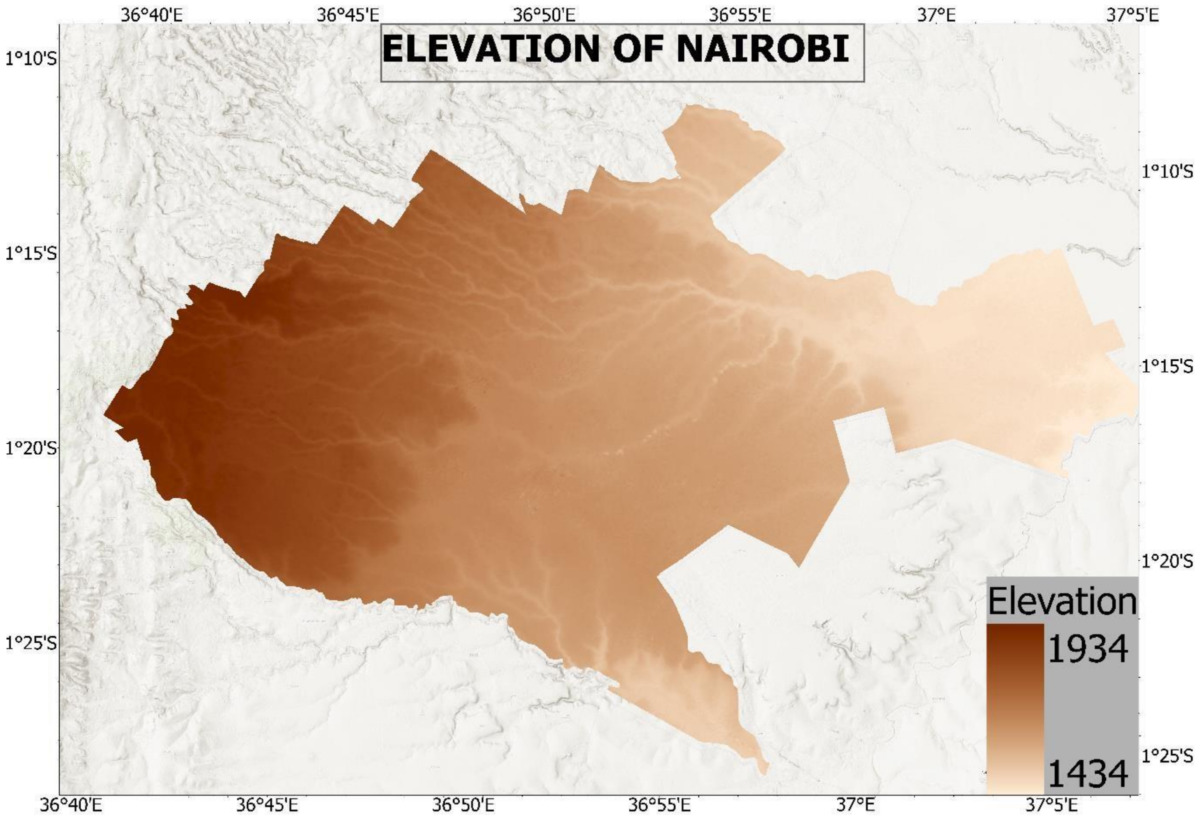
Zhang, H., Zhang, F., Ye, M., Che, T., & Zhang, G. (2016a). Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. *Journal of Geophysical Research*, *121*(19), 11425–11441. https://doi.org/10.1002/2016JD025154

Zhang, H., Zhang, F., Ye, M., Che, T., & Zhang, G. (2016b). Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. *Journal of Geophysical Research*, *121*(19), 11425–11441. https://doi.org/10.1002/2016JD025154

7. APPENDIX

ANNEX 1: STUDY AREA CHARACTERISTICS: NDBI and LST and DEM





ANNEX 2: DATASETS AND SOURCES

DATA	SOURCE	FORMAT	RESOLUTION	YEAR	Access

Building Footprint	Google Open Buildings Version 3	Geojson		2023	Open Buildings V3 Polygons | Earth Engine Data Catalog | Google for Developers .
Building Height	World Settlment Footprint 3D	Rater	90m	2015	EOC Geoservice Maps - World Settlement Footprint (WSF) 3D - Global, 90m (dlr.de)
Digital Elevation Model	ALOS PALSAR	Raster	12.5m	2022	Earthdata Login User Registration (nasa.gov)
NDVI, NDBI,NDWI	Sentinel 2	Raster	10m	2023-2024	Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | Earth Engine Data Catalog | Google for Developers
Urban Greenery Characters	Local Climate Zone(WUDAPT)	Raster	100m	2023	Submitted LCZ Maps (rub.de).
Land Surface Temperature	ECOSTRESS	Raster	70m	2023-2024	Earthdata Login User Registration (nasa.gov)

ANNEX 3: URBAN MORPHOMETRICS CHARACETERS

Morphological Characters As Adopted From Fleischmann et al., (2022)			
Index	**Element**	**Definition**	**Category**
Area	Building	area covered by a building footprint in m2.	Dimension
Height	Building	building height in m(required input for other metrics)	Dimension
Volume	Building	Building footprint multiplied by its height in m3	Dimension
Form factor	Building	captures 3D unitless shape characteristic of a building	Shape
Volume to Façade ratio	Building	proxy of volumetric compactness.	Shape
Perimeter	Building	sum of lengths of the building exterior walls in m.	Dimension
Courtyard Area	Building	sum of areas of interior holes in footprint polygons in m2 .	Dimension
Circular compactness	Building	Area of footprint / area of minimal enclosing circle	Shape
Squareness	Building	Sum of deviation of angle of corner from 90 degrees / number of corners.	Shape
Corners	Building	Vertex of building exterior shape with an angle between adjacent line segments ≤ 170 degrees	Shape

Equivalent rectangular index	Building	Measure of shape complexity of building footprint	Shape
Elongation	Building	Indirectly measures the deviation of the shape from a square	Shape
Centroid-corner distance deviation	Building	It captures potential circularity of object and topological imprecision of building polygon.	Shape
Centroid - corner mean distance	Building	Captures the dimension of the object dependent on its shape	Shape
Solar orientation	Building	Captures the deviation of orientation from cardinal directions.	Distribution
Cell alignment	Building	Reflects the relationship between a building and its cell.	Distribution
Longest axis length	Tessellation Cell	Proxy of plot depth for tessellation-based analysis.	Dimension
Area	Tessellation Cell	Area covered by a tessellation cell footprint in m2 .	Shape
Circular compactness	Tessellation Cell	Captures the relation of tessellation cell footprint shape to its minimal enclosing circle	Shape
Solar orientation	Tessellation Cell	Captures the deviation of cell orientation from cardinal directions	Distribution
Coverage Area Ratio	Tessellation Cell	Captures intensity of development.	Intensity
Equivalent rectangular index	Tessellation Cell	Measure of shape complexity of tessellation Cell	Shape
Shared walls ratio	Adjacent buildings	Captures the amount of wall space facing the open space	Distribution
Alignment	Neighbouring buildings	Calculates the mean deviation solar orientation of buildings on adjacent cells from a building	Distribution
Mean distance	Neighbouring buildings	Captures the average proximity to other buildings.	Distribution
Weighted neighbours	Tessellation cell	Reflects granularity of morphological tessellation.	Distribution
Area covered	Neighbouring cells	Captures the scale of morphological tessellation	Dimension
Reached area	Neighbouring segments	It captures an accessible area.	Dimension
Perimeter wall length	Adjacent buildings	Length of an exterior ring of a polygon composed of footprints of adjacent buildings	Dimension
Mean inter-building distance	Neighbouring buildings	Captures the average proximity between buildings.	Distribution
Building adjacency	Neighbouring buildings	Number of joined structures/ number of buildings	Distribution

ANNEX 4: CONTEXTUAL CHARACTER CALCULATIONS

1. *Interquartile mean (IQM)*, the most representative value cleaned of the effect of potential outliers.
2. *Interquartile range (IQR)*, as local measure of statistical dispersion, describes the range of values cleaned of outliers:

$$IQR_{ch} = Q3_{ch} - Q1_{ch}$$

where $Q3_{ch}$ and $Q1_{ch}$ are is the third and quartiles of the primary character.

3. *Interdecile Theil index (IDT)* describes the local (in)equality of distribution of values

$$IDT_{ch} = \sum_{i=1}^{n} \left(\frac{ch_i}{\sum_{i=1}^{n} ch_i} ln \left[N \frac{ch_i}{\sum_{i=1}^{n} ch_i} \right] \right)$$

where ch is the primary character.

4. *Simpson's diversity index (SDI)* captures the local presence of classes of values compared to the global structure of the distribution

$$SDI_{ch} = \frac{\sum_{i=1}^{R} n_i(n_i - 1)}{N(N-1)}$$

where R is richness, expressed as number of bins, n_i is the number of features within i-th bin and N is the total number of features.

Adapted from (Fleischmann et al., 2022)

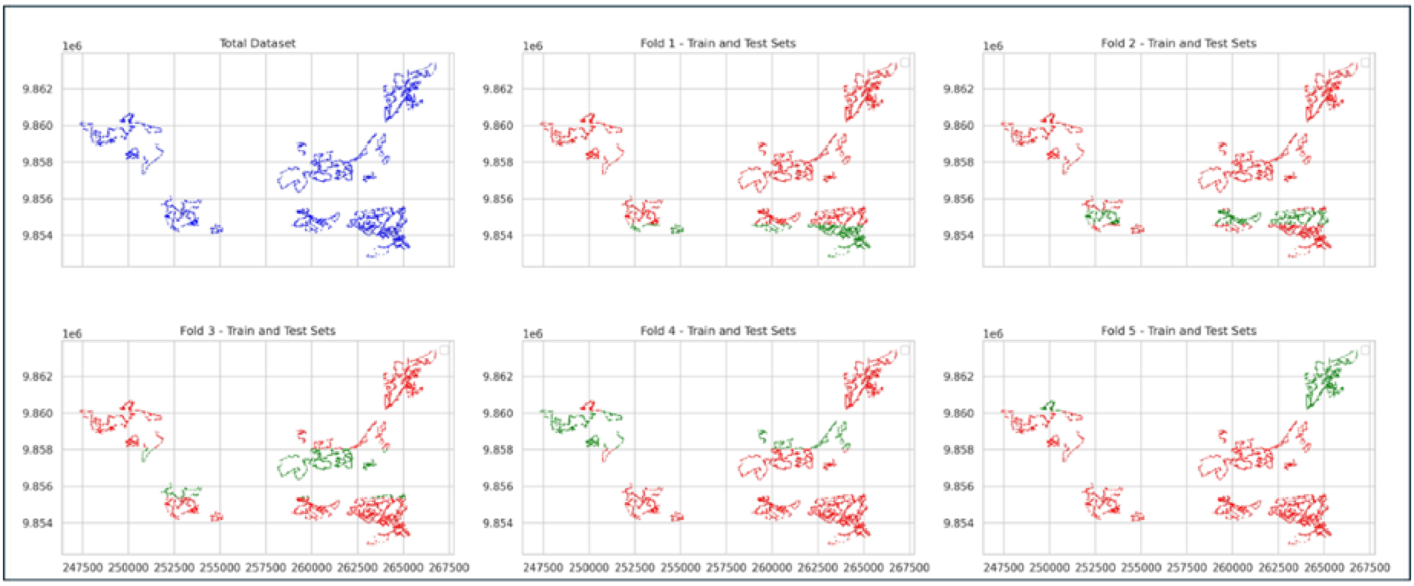
APPENDIX 5 : SENSORS USED FOR THE STUDY

From left, Kunak Stationary sensor, Garmin etrex 10, Kestrel Drop 2 and walking sensor sticks.

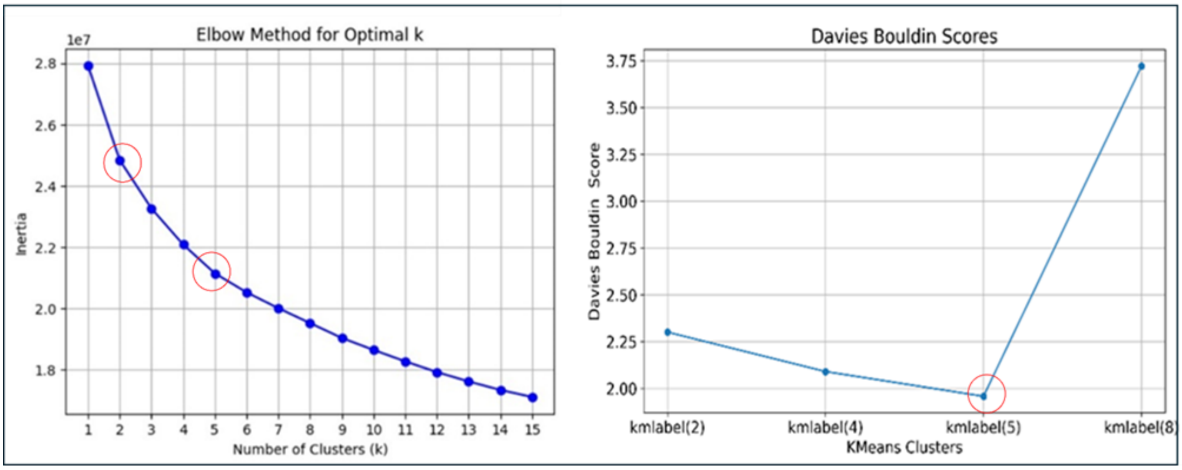
ANNEX 6 : Local CLIMATE ZONE CLASSES

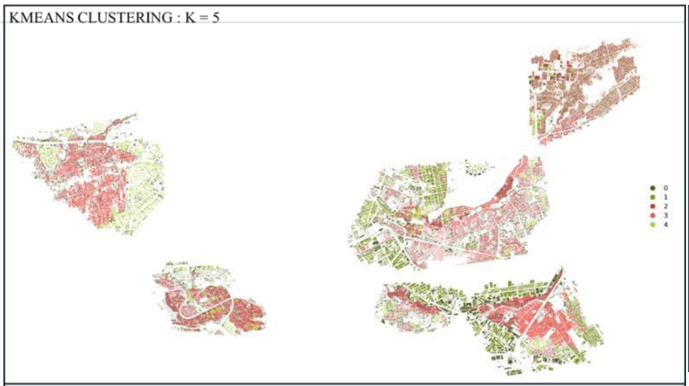
Adapted from (Stewart & Oke, 2012)

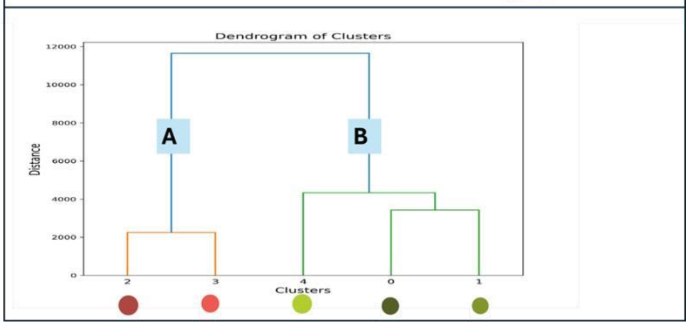
ANNEX 7 : KFOLD CROSS VALIDATION



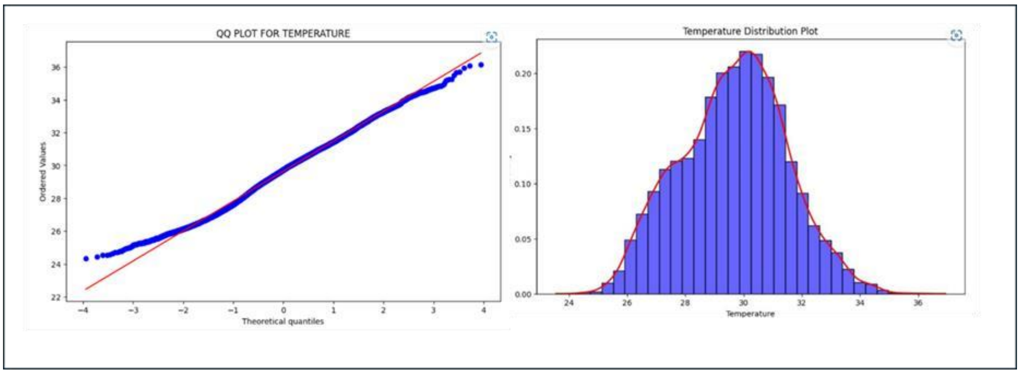
ANNEX 7: DETAILED KMEANS ANALYSIS

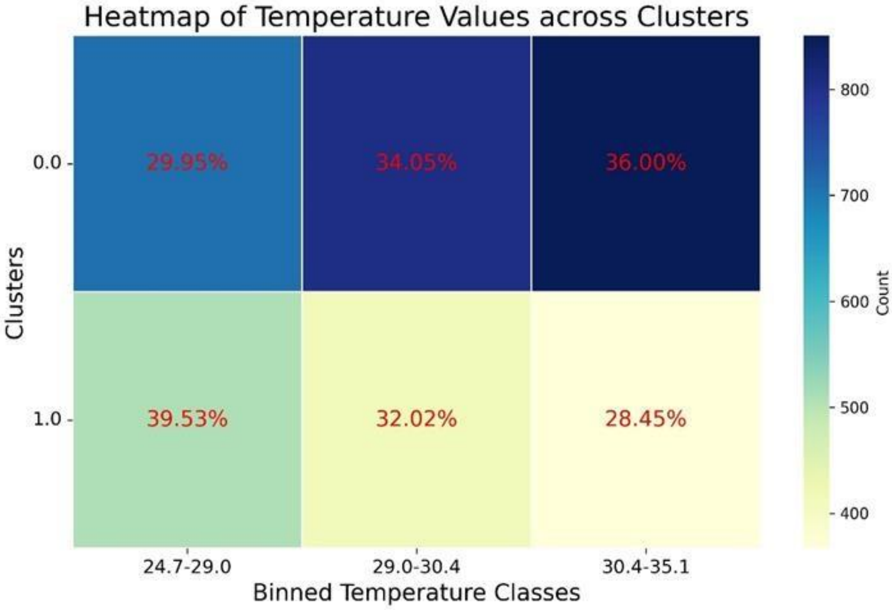


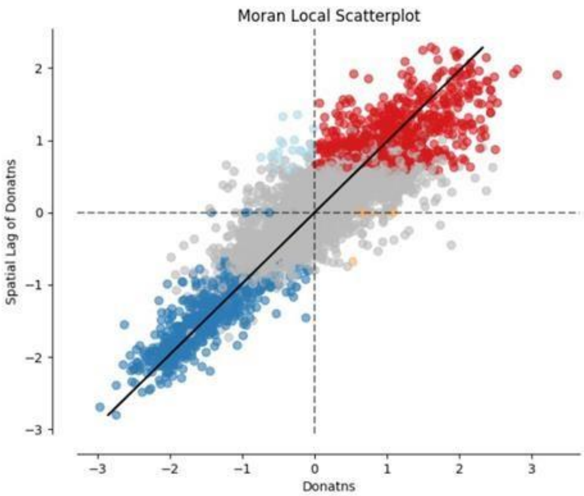




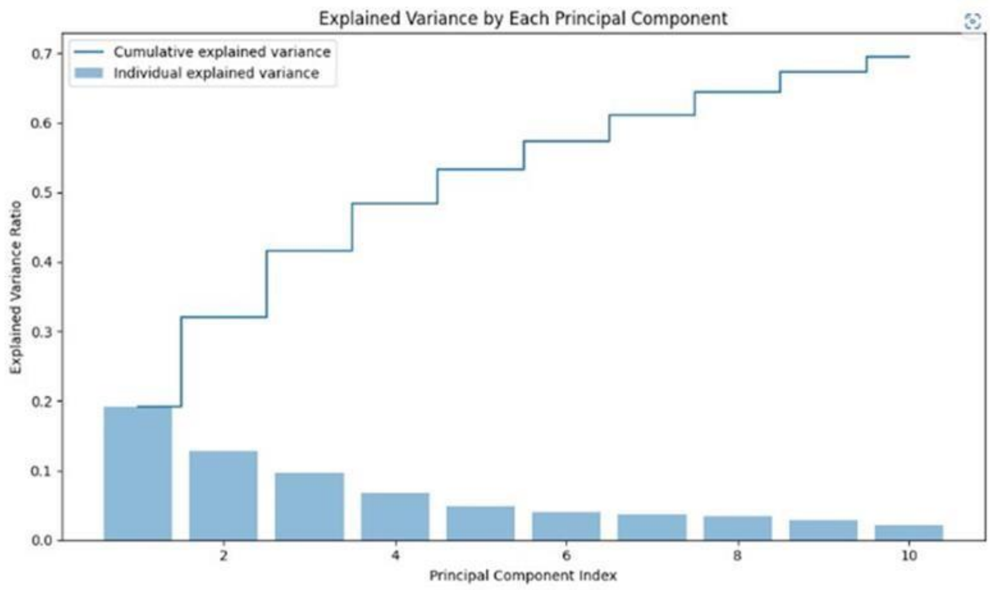
ANNEX 8; INSITU MEASUREMENT DESCRIPTIVES





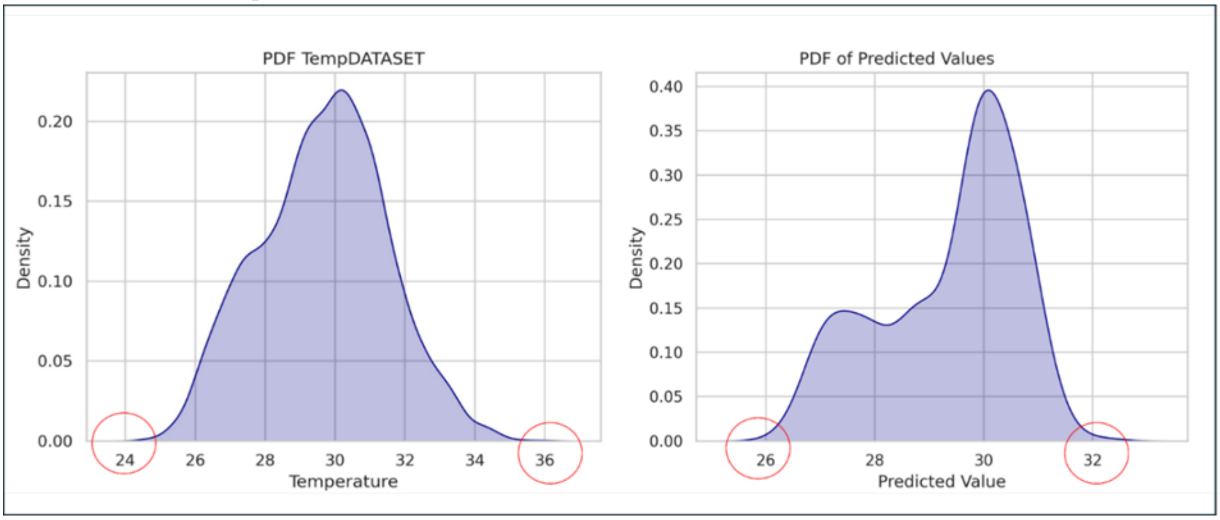


ANNEX 9; PRINCIPAL COMPONENT ANALYSIS

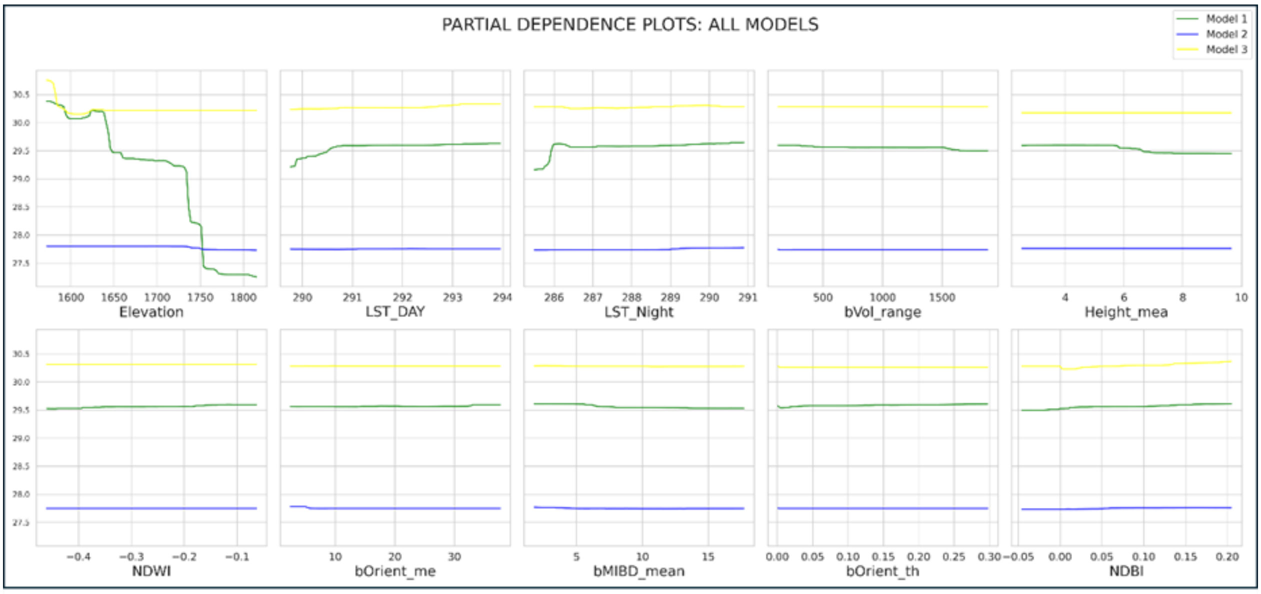


	Component	Most Important Feature
0	PC0	tWNeigh_me
1	PC1	bShpInd_me
2	PC2	BvolFaca_1
3	PC3	bRect_rang
4	PC4	bPeri
5	PC5	bCCM_theil
6	PC6	tcOrient_m
7	PC7	tWNeigh_ra
8	PC8	bOrient_me
9	PC9	NDWI

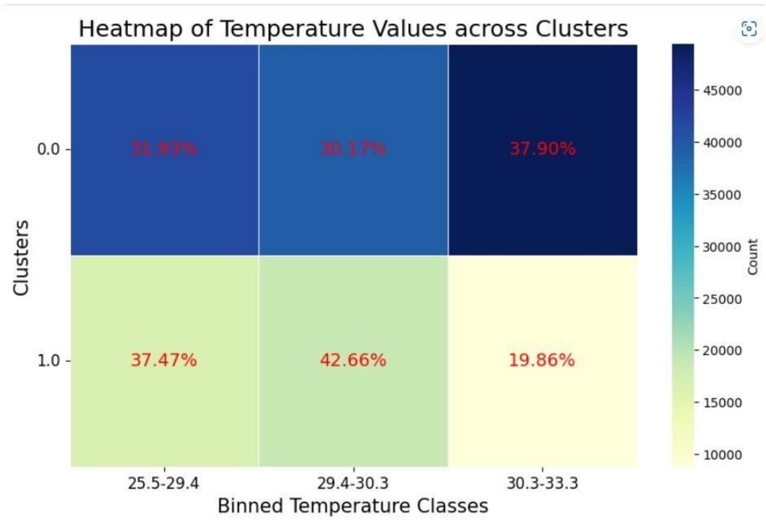
ANNEX 10: PDF of predicted and actual values

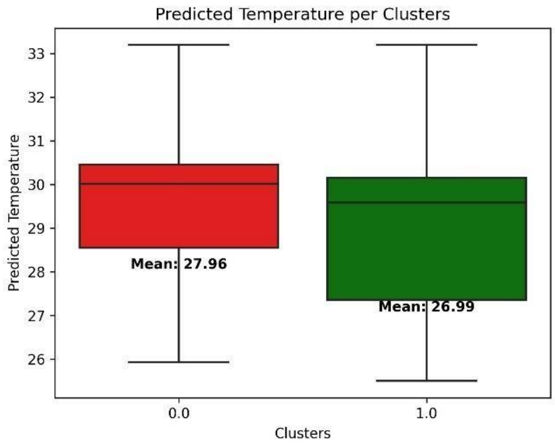


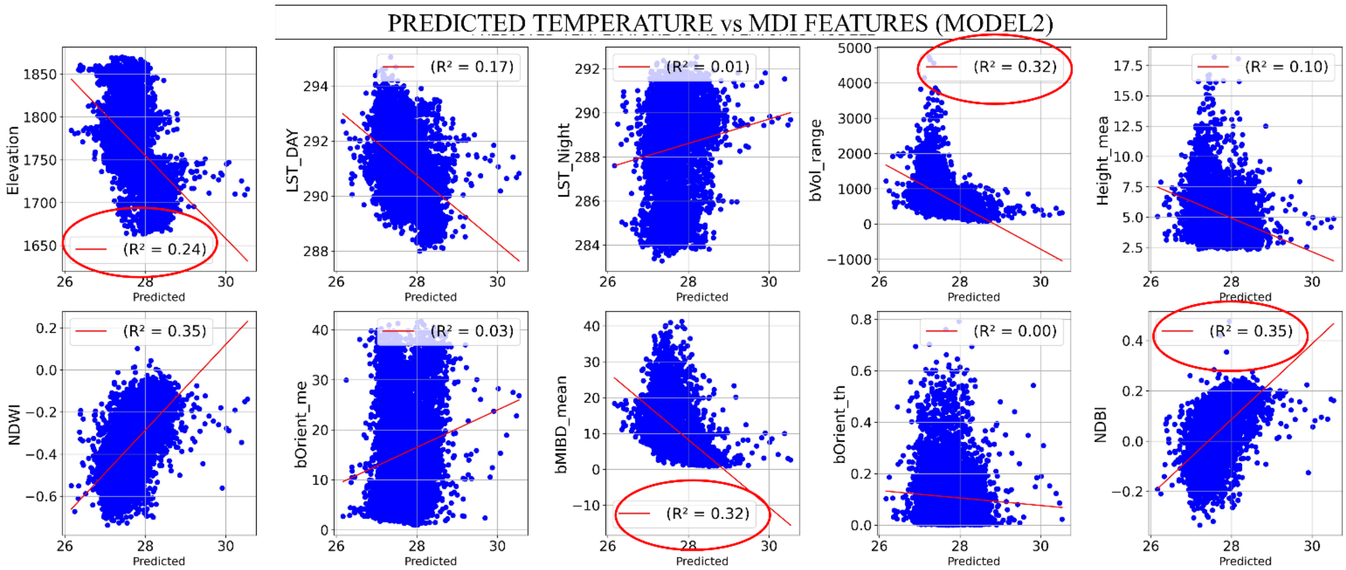
ANNEX 11; PARTIAL DEPENDENCY PLOTS



ANNEX 12; TEMPERATURE VARIANCE ACROSS CLUSTERS ANNEX 13; TEMPERATURE RELATIONSHIPS MODEL 2

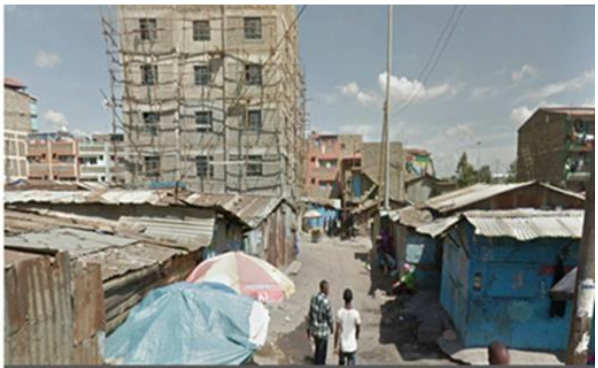






ANNEX 14 ; Study areas

Aerial View: Kibera

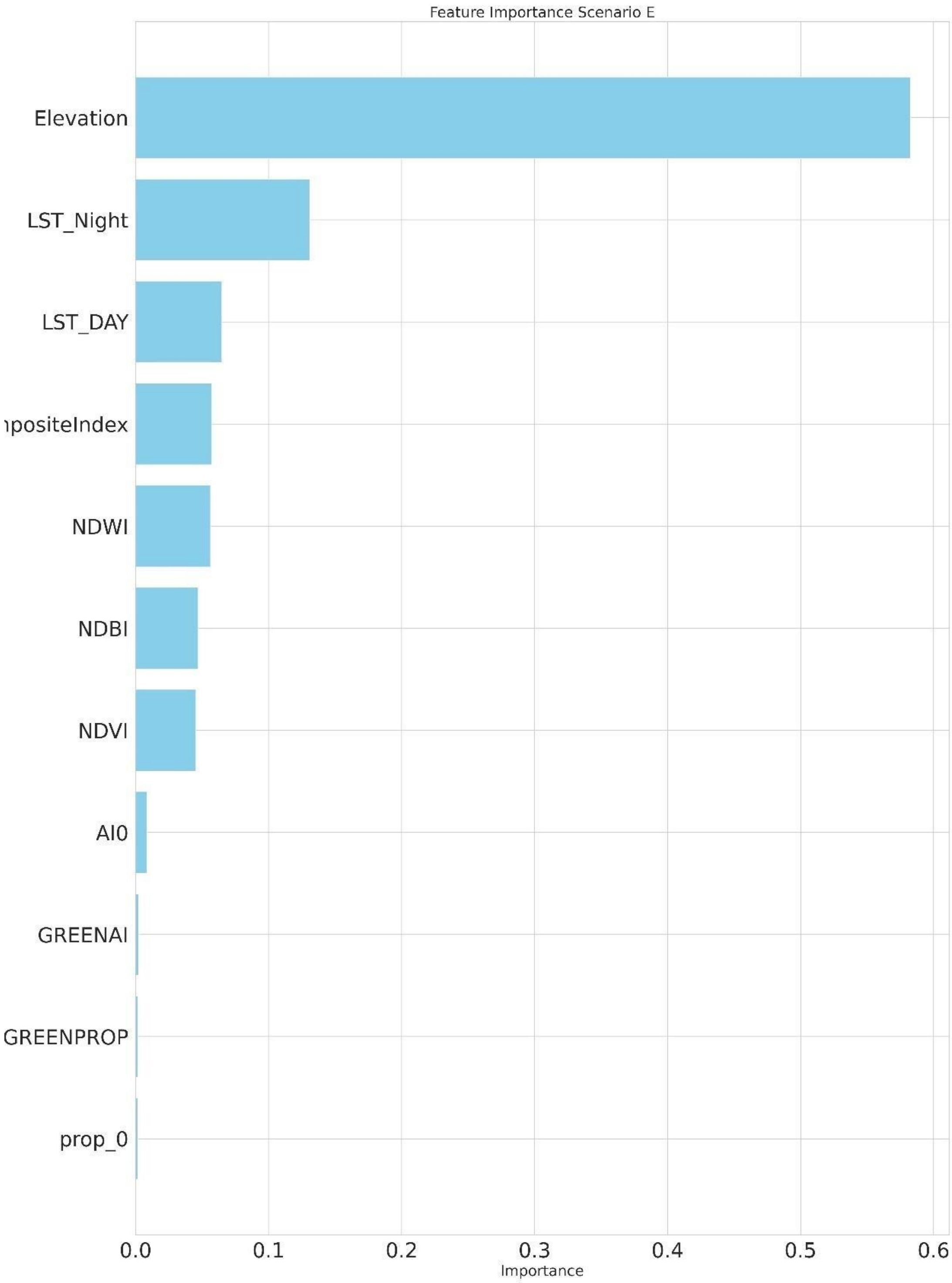


Ground View: Korogocho

Aerial View: Mathare

Ground View: Mukuru

ANNEX 15: Feature Importance Composite Morphometrics



Appendix 16; PARTIAL DEPENDENCY PLOT MODEL 2

PARTIAL DEPENDECY PLOTS MODEL 2