ANALYSING THE RELATIONSHIP BETWEEN HAZARDS AND DEPRIVATION USING MACHINE LEARNING

PRISCILLA NGIMA KABIRU
August, 2021

SUPERVISORS:

Dr. M. Kuffer
Prof.dr. R.V. Sliuzas

ANALYSING THE RELATIONSHIP BETWEEN HAZARDS AND DEPRIVATION USING MACHINE LEARNING

PRISCILLA NGIMA KABIRU
Enschede, The Netherlands, August, 2021

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
Prof.dr. R.V. Sliuzas

THESIS ASSESSMENT BOARD:
Dr. Diana Reckien (Chair)
MSc. Sabine Vanhuysse (External Examiner, Universite Libre de Bruxelles, Belgium)

Netherlands, August, 2021

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

According to literature, slums, herein referred to as deprived settlements, are located in hazardous areas. However, there have been very few studies that examine this notion. Studies that have analyzed this relationship (between hazards and deprived settlements) have primarily focused on single-hazards. In contrast, the analyses of multi-hazards have been hindered by a lack of sufficient methods and data. However, technological advancements in geospatial data and techniques present an opportunity to empirically investigate the relationship between hazards and deprivation. This study identifies multi-hazards in the select case study area of Nairobi through literature review and expert interviews. Using geospatial data, we identify proxies used to construct a city-wide index to investigate the location of deprived settlements and multi-hazards. We contrast morphologically identified deprived settlements to non-deprived settlements. We find that settlements in the inner city are more exposed to hazards than those located in the periphery. Further, physical traits determine the degree of susceptibility to hazards that a neighbourhood faces. Therefore, in partial agreement to literature, deprived settlements in the inner city are highly exposed to hazards, but so are formal planned high to mid-density settlements. On the other hand, deprived settlements in the urban periphery are less exposed except to hazards influenced by the neighbourhood characteristics, such as fire. Additionally, we test the predictability of deprivation using multi-hazards. We find that despite obtaining a high OA of 74%, the classification results by multi-hazards appear generalized. In contrast, though obtaining a lower OA by 2%, texture features result in more realistic land use classification. Lastly, we conduct household interviews in two deprived settlements to contrast the findings of the index. The index proxies used adequately capture the hazards. However, more localized data can improve multi-hazard index performance. Moreover, the cross-cutting approach of hazard assessment from the city to the household level lead to the detection of hidden patterns of deprivation – intra-settlement socio-spatial marginalization.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisors dr. Monika Kuffer and Prof.dr. R.V. Sliuzas who have journeyed with me through the phases of my research. I truly appreciate the support, including sharing your networks and bits of your life, critical feedback and undoubted trust you bestowed on me as the research unfolded organically. I have truly enjoyed being under your mentorship.

Further, I am grateful to Prof. Diana Reicken, the chair of my assessment board, for her light-hearted mood while giving critical feedback and the mystery behind her existence. I also note the worthy contributions of Nicholus Mboga and Maxwell Owusu during the initiation of my research. The conversations and feedback helped to shed light on the processes of academic research.

I also thank Sabine Vanhuysse, Stefanos Georganos of Universite Libre de Bruxelles, Belgium and Angela Abascal of IDEAMAPs for the help in refining my ideas and especially for providing me with data for my research; the team at the Department of Geography; King's College London, Strand Campus: Prof. Mark Pelling, Bruce D. Malamud, Robert Sakic Trogrlic and Sebastiaan Beschoor Plug who provided helpful input on hazard analysis, including providing me with relevant literature for my study and; the local organizations – Community Mappers and Spatial Collective for the help in collecting data. I extend my appreciation to my family, friends and classmates who've supported me in different ways through this period.

Last but not least, I wish to thank the ITC Foundation for financially investing in me through my Masters' education and; myself for making everything possible.

The research used data generated with funding from the Belgian Federal Science Policy (BELSPO) according to subsidy no. (SR/11/380) (SLUMAP: http://slumap.ulb.be/).

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

ALOS-PALSAR	Advance Land Observation Satellite - Phased Array type L-band Synthetic Aperture Radar
AOI	Area of Interest
ASM	Angular Second Moment
CARTs	Classification and Regression Trees
CBD	Central Business District
CO	Carbon Monoxide
CRED	Centre for Research on the Epidemiology of Disasters
CRS	Coordinate Reference System
DEM	Digital Elevation Model
EM-DAT	Emergency Events Database
EO	Earth Observation
ESA	European Space Agency
ESRI	Environmental Systems Research Institute
FOSS4G	Free and Open Source Software for Geoinformatics
GADM	Database of Global Administrative Areas
GEE	Google Earth Engine
GHG	Greenhouse Gases
GIS	Geographic Information System/Science
GLCM	Grey-Level Co-Occurrence Matrix
GOK	Government of Kenya
GRASS	Geographic Resources Analysis Support System
GSO	Generic Slum Ontology
HAND	Height Above Nearest Drainage
HH	Household
HR	High Resolution
IDEAMAPs	Integrated Deprived Area Mapping system
IPCC	International Panel on Climate Change
LST	Land Surface Temperature
LULC	Land Use Land Cover
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NGO	Non-governmental Organization
NIR	Near Infrared
NO_2	Nitrogen Dioxide
O_3	Ozone
OOB	Out Of Bag
OSM	Open Street Map
QGIS	Quantum Geographic Information System
RFC	Random Forest Classifier
RS	Remote Sensing
RTC	Radiometric Terrain Corrected

SAPs	Structural Adjustment Programs
SAR	Synthetic Aperture Radar
SDGs	Sustainable Development Goals
SO_2	Sulphur Dioxide
SSA	Sub-Saharan Africa
UNECE	United Nations Economic Commission for Europe
UNEP	United Nations Environment Programme
UNFPA	United Nations Fund for Population Activities
UN-Habitat	United Nations Human Settlements Programme
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VHR	Very High Resolution
VIS	Visible
VSURF	Variable Selection Using Random Forest
WGS	World Geodetic System
WHO	World Health Organization

1. INTRODUCTION

1.1. Background and Justification

Globally, disasters cause millions in economic losses and thousands of fatalities annually (Dilley et al., 2005; EM-DAT, 2009). Presently, cities are affected by more than one hazard, and the frequency of disasters is reportedly increasing (Dilley et al., 2005). Disasters refer to sudden accidents, potentially causing damage and losses, while hazards are defined as physical phenomena that can lead to disasters (Gallina et al., 2016). Yet, cities are currently home to more than 50% of the world's population (United Nations, 2019). Continued rapid urbanization aggravates the issue since cities are located in hazard-prone areas and contribute to increased hazards. Urbanization has also been spatially expansive, characterized by increased impervious surfaces and less vegetation due to mass land cover changes (Seto, Sánchez-Rodríguez, & Fragkias, 2010). These characteristics make cities heat sources and poor water storage and drainage systems (Seto & Shepherd, 2009). They have also destroyed natural ecosystems, led to environmental stresses, and degradation (Seto & Shepherd, 2009). Additionally, urban areas have led to increased heat-trapping greenhouse gases (GHG) due to fossil fuel combustion (Revi, Satterthwaite, et al., 2014). Carbon Dioxide (CO_2) emissions from cities account for over 70% of the anthropogenic GHG (UNEP, 2020). Collectively, these anthropogenic causes have significantly contributed to global warming, a phenomenon characterized by the increase in the earth's average surface temperature. Global warming's associated impacts are reported to be already influencing the climate system, thus posing 'new threats' to urban areas (Hoegh-Guldberg, Jacob, & Taylor, 2018).

In addition to adversely affecting the climate system, inequality characterizes many cities globally. Inequality is presented as an economic polarization between the wealthy and the poor; and is perpetrated by inequitable distribution of resources and insufficient anti-poor policies (Phillips et al., 2007). As a result, urban poverty (a set of socio-economic difficulties brought about by systemic inequality) is a looming phenomenon in cities. The magnitude of inequality is particularly dire in the Global South, where urban poverty manifests as slums settlements (non-exclusively) (Baker, 2008; UN-Habitat, 2015). Presently, one in eight urban dwellers live in a slum (UN-Habitat, 2015); and in Sub-Saharan Africa (SSA), 59% of the urban population are slum residents (UN-Habitat, 2015). Slums are defined by UN-Habitat using five household deprivations - the lack of access to improved water services, sanitation facilities, sufficient living area, durable housing, and tenure security (UN-Habitat, 2003). Similar to other studies (see Kuffer et al., 2020, 2018; Thomson et al., 2020) and conscious that the term slum bears a negative connotation and has been politicized, we adopt the term deprived areas (and its variants) to refer to slums in this study (Borie, Pelling, Ziervogel, & Hyams, 2019; Mayne, 2017). Specifically, "deprivation implies a standard of living or a quality of life below that of the majority in a particular society, to the extent that it involves hardship, inadequate access to resources, and underprivilege" (p.362, Herbert, 1975). It is also important to note that not all who live in deprived areas are poor, and poverty exists beyond the boundaries of deprived settlements (Calder, Medland, Dent, & Allen, 2009).

Furthermore, deprived settlements have been linked to housing inadequacy and unaffordability (UN-Habitat, 2015; United Nations, 2019a). Specifically, inadequate housing supply and unaffordability have resulted from a failure by urban authorities and institutions to meet the demand for housing and service provision (UNFPA, 2007). Therefore, in the absence of adequate and affordable housing, the urban poor, lacking land access and tenure security (which affords access to financial mechanisms), put up shelter in hazardous areas (UN-Habitat, 2015). Additionally, the quality of the housing structure in deprived areas is often precarious and offers insufficient protection from climate and weather elements (UN-Habitat, 2015). Collectively, these challenges are captured by UN-Habitat's domain of durable housing that considers (i) structure permanency - an evaluation of the type and quality of building material, compliance with building codes, and state of a structure; and (ii) location of structure - evaluated based on whether or not a dwelling is located in hazardous areas (on or near toxic waste, a geologically hazardous zone, high-industrial pollution areas, or other unprotected high-risk zones) (UN-Habitat, 2018). Consequently, disasters in cities represent a significant source of risk, especially for the urban poor (Dilley et al., 2005; Revi, Satterthwaite, et al., 2014). Despite this, there has been systemic failure to assess the physical and environmental living conditions of the urban poor since many studies have focused on a set of social and economic factors such as income, consumption, and expenditure of households to define the phenomenon (Sanusi, 2008; UNFPA, 2007).

Additionally, the primary assumption held by urban authorities and development agencies was that urban poverty is a "transient phenomenon of rural-to-urban migration and will disappear as cities develop" (p.14, Phillips et al., 2007). This assumption has been held for decades and transferred from the industrial cities of the 1800s into the 21st century (Mayne, 2017). As a result, the degree of urban poverty and its spatial patterns have remained masked for decades. Urban poverty studies have progressively shifted focus to incorporate processes leading to urban poverty and the heterogeneity and multi-dimensional nature of the phenomenon (Cano, 2019). In particular, Geospatial and Earth Observation sciences have been beneficial in analyzing urban poverty by investigating urban areas' spatial patterns.

1.2. Research Problem

Deprived settlements represent urban poverty, a high degree of deprivation, and socio-spatial marginalization where the inhabitants are severely disadvantaged and subjected to life-threatening conditions (UN-Habitat, 2015). As the frequency of disasters increases in cities, there is a dire need to effectively mainstream disaster risk reduction strategies into development agendas (Dilley et al., 2005); and develop tools primarily targeted to protect those living in deprivation (United Nations, 2017). To do this, adequate and timely data of deprived areas is imperative. However, data on deprived areas have been missing from official records for years - a matter attributed to the political connotation around their existence. In addition, efforts to capture their presence and conditions have been mainly through household surveys. These are often limited in scope, lacking geo-locational and spatial characteristics, time and resource-intensive, aggregated at pre-defined administrative boundaries and collected after long periods, e.g., national censuses (Kohli, Sliuzas, Kerle, & Stein, 2012; Martínez, Pfeffer, & Baud, 2016; UN-Habitat, 2018). A representation that only gives a partial view of deprivation.

Looking at hazards in cities, the scope of the investigation has been limited due to the focus on single hazards (e.g., J. Wang, Kuffer, Sliuzas, & Kohli, 2019; S. Wang, Wang, Fang, & Feng, 2019). Additionally, many studies rely on household survey data and are operationalized at very localized scales (e.g., Mulligan, Harper, Kipkemboi, Nobi, & Collins, 2017). However, advancements in remote sensing and machine learning can be used to address these challenges. Earth observation data are spatial and offer many advantages over the traditional data collection methods such as timeliness, high spatial and temporal resolutions, wide coverage, and higher accuracy (Kuffer, Pfeffer, & Sliuzas, 2016). They also capture environmental phenomena indicative of hazards that have been incorporated in studies to investigate the relationships between deprivation and different types of hazards, for example: using air quality (S. Wang et al., 2019) and temperature (J. Wang et al., 2019). On the other hand, machine learning techniques provide the advantage of being computationally powerful; thus, they can handle large datasets. They also help solve complex problems. Hence, they have been found helpful for intra-city mapping and analysis of deprivation (e.g., Ajami, Kuffer, Persello, & Pfeffer, 2019; Liu, Kuffer, & Persello, 2019; Mboga, Persello, Bergado, & Stein, 2017).

Therefore, leveraging the advantages of remote sensing and machine learning, this study analyzes the relationship between hazards and deprivation using a multi-hazard approach and employs it at three spatial levels (city, settlement, and household level). By considering multi-hazards, defined as "the totality of relevant hazards in a defined area" (p.7, Kappes, 2011) e.g. within an administrative boundary, we anticipate to identify the hazards which deprived settlements are predisposed to and that also hidden deprivation is uncovered. Furthermore, we present a crosscutting approach to understanding how urban residents interact with hazards by considering three spatial levels for analysis. Moreover, the study aims at being flexible on the data used. For example, Müller et al. (2020) illustrated, the slope can be used as a proxy for indicating susceptibility to landslides. Due to their affordability, transferability, and ease of replicability, free, open-source data and machine learning algorithms are used in this study. The study and the approaches employed are seen as necessary in the wake of climate change risks in cities and for the reporting on slum indicators that aid in informing decision-making, guiding efficient planning, and developing impactful policies and programs.

1.3. Research Objectives and Questions

1.3.1. General objective

To analyse the relationship between hazards and deprivation using machine learning.

1.3.2. Sub-objectives

1. Identify geospatial data indicators of hazards to be used as predictors of deprivation.
2. Apply machine technique using identified data to predict deprivation.
3. Investigate the intra-settlement disbursal of hazards

1.3.3. Research Questions

1.3.3.1. Sub-objective 1

Which hazards are deprived areas predisposed?

Which open geospatial data can be used as hazard indicators?

1.3.3.2. Sub-objective 2

Are deprived areas more likely to be located in hazardous areas in relative comparison to formal settlements?

What share of deprived areas are located in hazard-prone areas?

Can a multi-hazard dataset be used to predict deprivation?

How do multi-hazard datasets compare to textural features in the prediction of deprivation?

1.3.3.3. Sub-objective 3

How are hazards disbursed within a deprived settlement?

2. LITERATURE REVIEW

In this chapter, findings from other studies are presented to justify the aim of our research and the choice of data and methodologies that we employ.

2.1. Multi-Hazards

The acknowledgement of the existence of multiple hazards was first made at the Agenda 21 conference (UNEP, 1992), recognizing the importance of multi-hazard analysis as part of pre-disaster planning. It is also where the concept - multi-hazards was first mentioned (Gallina et al., 2016). Since then, the progressive increase of disaster risks has emphasized the need for an integrated approach to hazard analysis (Dilley et al., 2005; Greiving, 2006). From the definition of multi-hazards presented earlier, two key elements are captured: (i) *totality* and (ii) *relevancy* of the hazards (Melanie Simone Kappes, 2011). Hence, the definition implies that all hazards relevant within a study area should be considered in assessing multi-hazards. However, this has not been the case since multi-hazards are diverse and require different data and methods for their assessment (Melanie S. Kappes, Keiler, von Elverfeldt, & Glade, 2012).

Further, the lack of interdisciplinary approaches to identify hazards and develop suitable methods presents additional challenges (Melanie S. Kappes et al., 2012). These challenges are captured by Gallina et al. (2016) in their review of multi-risk methodologies. The study reveals that even in cases of multi-hazard assessment, many studies focus on one type of hazard, e.g., natural hazards or technological hazards. Still, attempts at integrating climate change-induced hazards have not been made. We further note that hazards such as air pollution, a significant global public health threat (WHO, 2021a), are not considered in multi-hazard analyses, despite their infamous research in urban and health studies.

To address the challenge of heterogenous hazard data (for natural disasters), Dilley et al. (2005) create a simple hotspot multi-hazard index. Similarly, a multi-risk index developed by Greiving (2006) constitutes a spatial weighted multi-hazard index. Indices are widely used methodologies in studies that rely on heterogenous data since their primary function is to compile different data into a single metric. The approach has proven helpful in assessing various phenomena, especially in social sciences, e.g., the development of the human development index (HDI), further adopted to measure deprivation (Sanusi, 2008). Furthermore, Ajami, Kuffer, Persello, & Pfeffer (2019) developed a methodological framework combining surveyed and earth observation data, providing a more holistic deprivation index. Their research implemented a deprivation index using machine learning and very high resolution (VHR) images. Therefore, indices demonstrate their ability to handle heterogenous data while providing meaningful results in multi-hazard analyses, social and urban studies, and interoperability with different methods. Thus, in this study, we use a simple equal-weighted index for the assessment of multi-hazards.

2.2. Hazards and Deprivation Mapping

Despite the increase in multi-hazard assessments, very few studies have empirically investigated the location of deprived settlements in hazardous areas. This could be attributed to the reliance on standard socio-economic surveys method for collecting data on deprived settlements. Nonetheless, with increased climate change-driven hazards, innovative household surveys have shown that deprived settlements are located in hazardous areas (e.g., Mulligan, Harper, Kipkemboi, Nobi, & Collins, 2017). In their study, carried out in parts of Kibera, Nairobi – Kenya's largest informal settlement, 50% of the surveyed households reported that they experienced flooding during the long rainy season. These results, however, still face the mentioned challenges (section.1.3.) of using household surveys. Additionally, using such methods, it is only implied that deprived settlements are located in hazardous areas since the spatial component remains a miss.

Remote sensing has, however, been used to address this data gap. Recently, satellite imagery has been used to provide quantifiable evidence on the location of deprived settlements in hazardous areas. For example, Müller et al. (2020) assessed deprived settlements on landslide-prone areas using slope proxy. Their study, carried out across seven cities, found that deprived settlements are relatively more likely to be located in landslide-prone areas than formal settlements. Another study investigating heat exposure in urban areas found deprived settlements in places with higher temperatures (J. Wang et al., 2019). These studies stress the need for empirical investigation of the presence of deprived areas in hazardous areas. Despite this, single-hazard approaches remain limited in that they neither present the overall degree of hazardousness nor allow for the understanding of interactions between hazards (Melanie S. Kappes et al., 2012).

2.3. Prediction of Deprivation Using Multi-Hazard Index

Research on deprivation mapping is progressively expanding and improving in the development of methods and frameworks used. Robust image processing techniques, like machine learning, show that, similar to traditional deprivation indices, multisource geospatial data can be compounded into indices for assessing different dimensions of deprivation. Initially designed for pattern recognition, machine learning techniques have been incorporated in remote sensing, with the main advantage of automatically detecting patterns in data (Goodfellow, Bengio, & Courville, 2016). They also use multisource data as input in the models and have thus been found effective for land use and land cover classification (Gislason, Benediktsson, & Sveinsson, 2006).

For this reason, we use the traditional machine learning method - Random Forest Classifier (RFC) in our study. RFC is based on the combination of automatic learning algorithms and hand-crafting feature engineering techniques (LeCun, Bottou, Bengio, & Haffner, 1998). Hand-crafted techniques are limiting, especially in processing large data, since the process is laborious, non-transferable, non-scalable, and prone to biases which can compromise the models performance (LeCun et al., 1998; Persello & Stein, 2017). Despite these limitations, RFC offers the following advantages in comparison to other ML models: (i) achieving high classification accuracy with fast processing speed; (ii) they are robust to little training data compared to more conventional ML models like Neural Networks and (iii) they are interoperable with data from different sensors (multi-source data) (Belgiu & Drăgu, 2016; Gislason et al., 2006). The operations of RFC are discussed below in detail.

2.3.1. Ensemble Classifiers: Multisource Data Analysis Using Random Forest Classifier (RFC)

Ensemble classification is a machine learning (ML) technique that combines several base classifiers to produce one optimal model (Gislason et al., 2006). A commonly used base classifier is Decision Trees (Belgiu & Drăgu, 2016). In an ensemble of trees(a collection of decision trees built sequentially where each succeeding tree recovers the loss of the previous (Nagpal, 2017)), each classifier is trained and the results aggregated through a voting process (Belgiu & Drăgu, 2016; Gislason et al., 2006). This approach has yielded better accuracies than using single decision trees (Belgiu & Drăgu, 2016). In training the classifiers, the most commonly used techniques are boosting and bagging (Bootstrap AGGregating) (Belgiu & Drăgu, 2016; Gislason et al., 2006). The boosting approach employs an iterative re-training and re-weighting (for incorrectly classified samples) process using all the training samples (Belgiu & Drăgu, 2016; Gislason et al., 2006). On the other hand, the bagging approach draws subsamples of the entire training set (Belgiu & Drăgu, 2016; Gislason et al., 2006). Both methods have been found to offer the advantage of reduced classification variance (Belgiu & Drăgu, 2016; Gislason et al., 2006). In contrast, boosting has been found to produce higher accuracies than bagging: while bagging offers the advantage of requiring less computational resources and effect on classification bias (Belgiu & Drăgu, 2016; Gislason et al., 2006).

2.3.1.1. Random Forest Classifier

Random Forest Classifier (RFC) uses an ensemble of Decision Tree-type supervised classifiers called Classification and Regression Trees (CARTs) (Belgiu & Drăgu, 2016). CARTs are trained using a similar approach to bagging: with a tweak in how the splitting of trees occurs. While in the standard bagging approach, the trees break at similar features in each tree; in RFC, random subsamples of the training set (with replacement) are used for training the classifiers (Belgiu & Drăgu, 2016). Randomizing this process reduces the correlation between trees, and using a subset of the features with replacement reduces the computational costs (Belgiu & Drăgu, 2016). Thus, the trees are split at different features (nodes), creating bigger ensembles. Then, the class prediction is made based on the majority vote in the ensemble (Belgiu & Drăgu, 2016). Consequently, they produce a predictor model with the advantages of bagging and greater accuracies comparable to the booting approach without its shortcomings (Belgiu & Drăgu, 2016; Gislason et al., 2006).

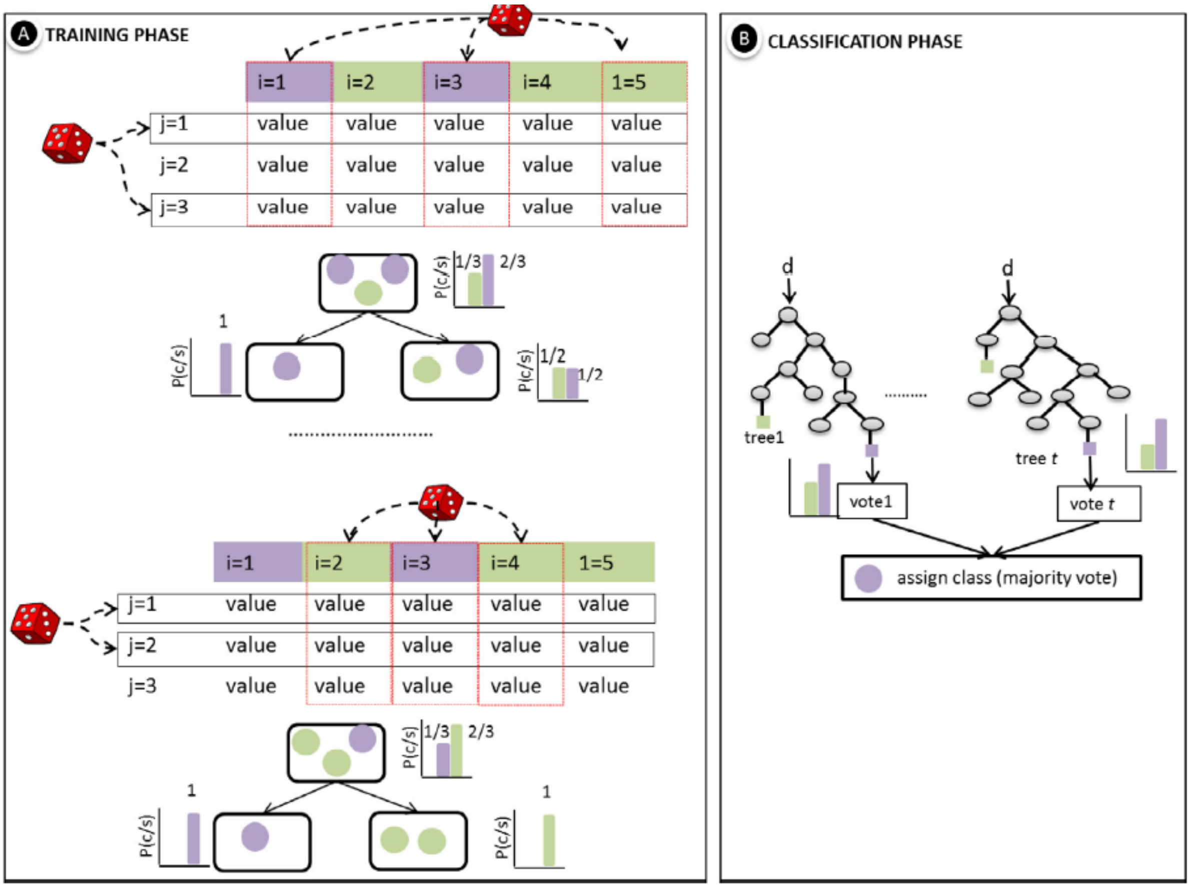


Figure 1: Training and classification phases of Random Forest classifier
i = samples, j = variables, p = probability, c = class, s = data, t = number of trees, d = new data to be classified, and value = the different values that the variable j can have.
Source: Belgiu & Drăgu, (2016)

2.3.1.2. Multisource Data Analysis

CARTs are non-parametric, meaning they do not assume normal data distribution (Belgiu & Drăgu, 2016; Gislason et al., 2006). Therefore, RFCs can effectively analyze multivariate data, e.g., multispectral imagery for LULC classifications that rarely have a normal frequency distribution (Belgiu & Drăgu, 2016; Gislason et al., 2006). Additionally, supervised classifiers are robust in learning class characteristics from training sample data and subsequently identifying them in unclassified data (Belgiu & Drăgu, 2016). Specifically, RFCs perform well with noisy and imbalanced training samples (Belgiu & Drăgu, 2016). RFCs also offer the advantage of assessing variables' ability to classify target classes and rank them in order of importance by determining their collinearity (Belgiu & Drăgu, 2016). These characteristics are particularly useful in this study, where multisource earth observation and geographic data are analysed in the form of a multi-hazard index.

2.3.1.3. RFC Model Operation And Validation

Two user-defined parameters are required in an RFC. *Ntree* determines the number of trees that grow, and mtry determines the number of splits at each tree node (Belgiu & Drăgu, 2016). Additionally, RFC has an internal performance evaluation technique that uses *out-of-the-bag* (OOB) samples to produce an error estimate, called the OOB error (Belgiu & Drăgu, 2016). The *out-of-the-bag* samples constitute approximately one-third of the input samples. The rest of the (*in-bag*) samples are used in the training of the trees (Belgiu & Drăgu, 2016).

2.3.2. Linear Canonical Discriminant Analysis

RFC, despite its interoperability with multi-source data, is considered a 'black-box' model. Efforts towards transforming the model into a white-box model have been made using programming libraries such as treeinterpreter in python, which decompose the model's predictions (Saabas, 2014). The processes are, however, complex to implement, and our skills and knowledge level are limited. For this reason, we incorporate an additional method –statistical discriminant analysis to perform classification of deprivation using multi-hazards.

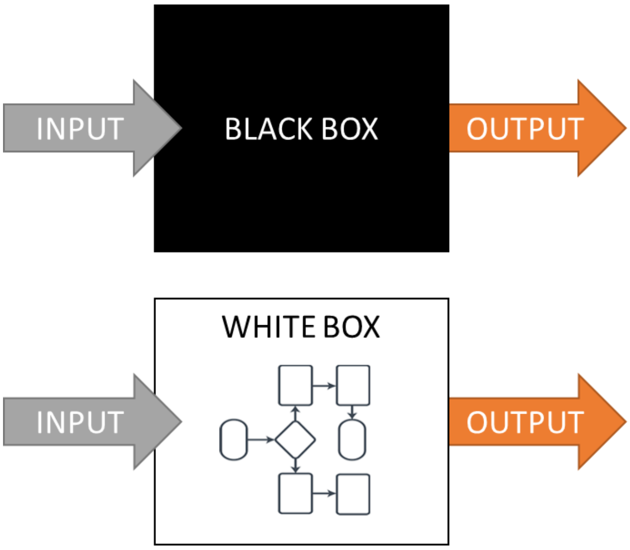


Figure 2: Illustration of black-box and white-box models

Discriminant analysis is a nonparametric linear model which is used for multi-variate analysis (Field, 2017). It has been mainly used in health and environmental/ecological studies, including those with a spatial component (e.g., Hall, Evanshen, Maier, & Scheuerman, 2014; Reitz, Hemric, & Hall, 2021). One of the mentioned studies analyses the nonpoint source of contamination in watersheds by comparing the effects of different land use/land covers (Reitz et al., 2021). The study demonstrates how discriminant analysis uses cases comprising categorical variables and predictor variables as input. Each case is plotted on a feature space, labeled by its class (Field, 2017). Therefore, by plotting the cases, the model forms decision boundaries by class and can thus make predictions of new cases based on where they lie in the feature space when plotted (Field, 2017). The popularity of discriminant analysis in ecological studies is implied to originate from the need for localized analysis dependent on the detection of subtle categorical distinctions, e.g., of transitional zones whose dynamics are critical indicators of environmental change (Lobo, 1997). Based on this logic, we consider canonical discriminant analysis an appropriate method for distinguishing residential land uses (deprived vs. non-deprived) with subtle distinctions. Additionally, given the study's attempt to predict deprivation using multi-hazard data, we deem it important to understand the local interpretation of the predictions (implied by the predictor variables).

2.4. Theoretical Framework

The theoretical framework below summarizes interrelationships of the theories and concepts discussed in the above sections - explaining the phenomenon of deprivation and its relationship with hazards. Additionally, the framework captures how we take advantage of geospatial data that offer promising opportunities. They have been used to map deprivation and capture the different phenomena indicative of hazards. Furthermore, geospatial data are interoperable with machine learning techniques that have proven robust in analysing complex phenomena (including deprivation and hazards).

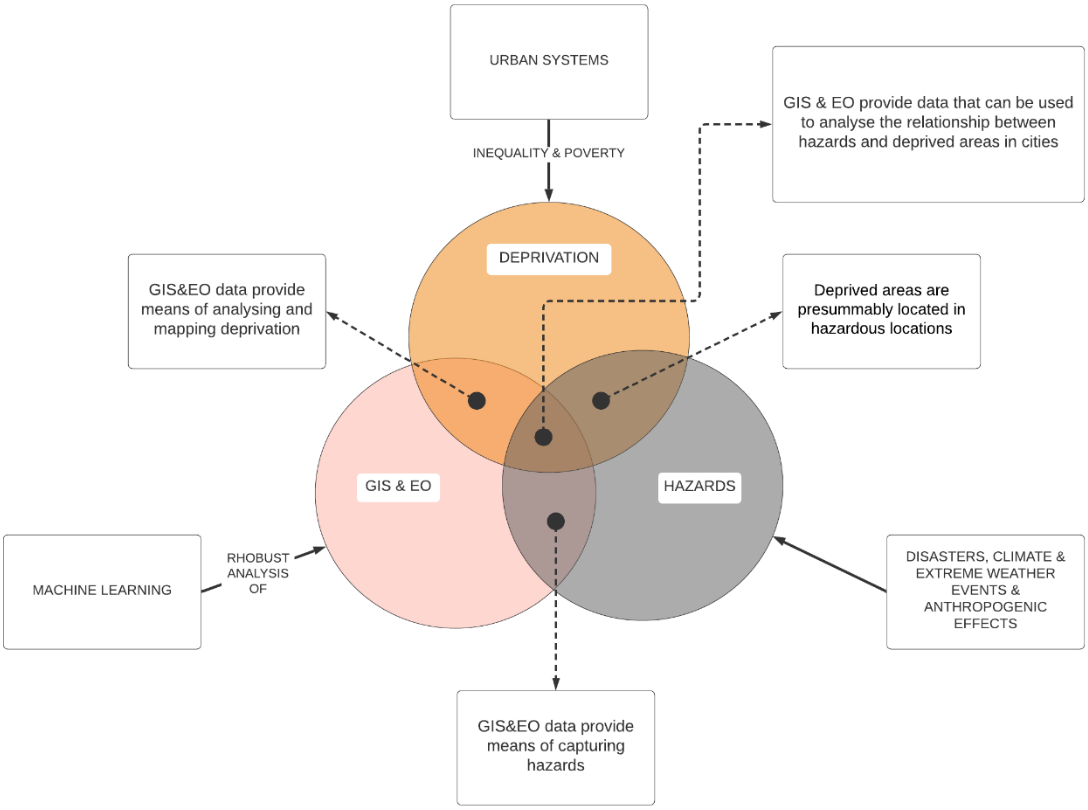


Figure 3: Theoretical framework – a logical representation of theories framing the study

3. RESEARCH METHODOLOGY

This chapter discusses the research process used to identify data, including the data collection techniques and analysis processes. Since we are undertaking a case study approach for the research, the study area is also presented in this section. We also analyse four different types of residential settlements in our study area, differentiating deprived and non-deprived settlements to compare the degree of hazardousness across settlements. These data are also used as the label data in our predictive classification process. Specific to the data collection and analysis, we start by constructing the multi-hazard index. The process is informed by extensive literature review and key informant interviews – a consultative participatory approach (Vaughn & Jacquez, 2020). By consulting the experts, we gain insights into the hazards present in the study area and thus refine the theoretical multi-hazard index. Next, we conduct extensive literature and database search to identify geo-data to construct the index. We then apply descriptive statistics to analyse the relationship between hazards and the different settlements in our study area. Afterwards, using machine learning methods, we test the ability of multi-hazards to predict deprivation. Lastly, we contract a local research group (comprising residents from deprived settlements) to conduct household interviews, employing consultative and inclusive participatory research principles (Vaughn & Jacquez, 2020). The results are contrasted to the multi-hazard index outcomes. The household survey outcomes are also used to analyse the inter-settlement hazard dispersal and household-level exposure to hazards.

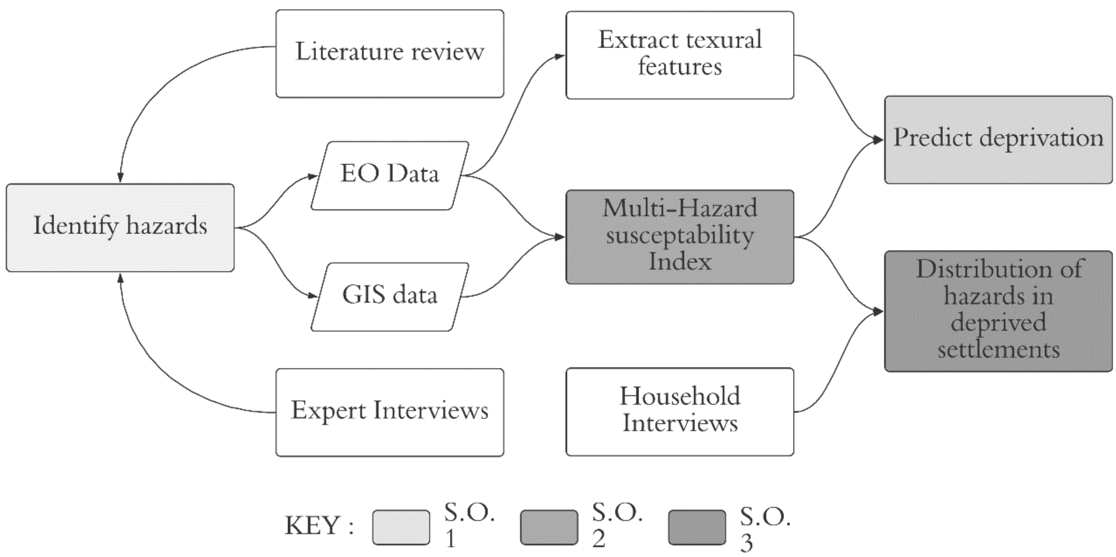


Figure 4: Research process

3.1. Case Study Area – Nairobi

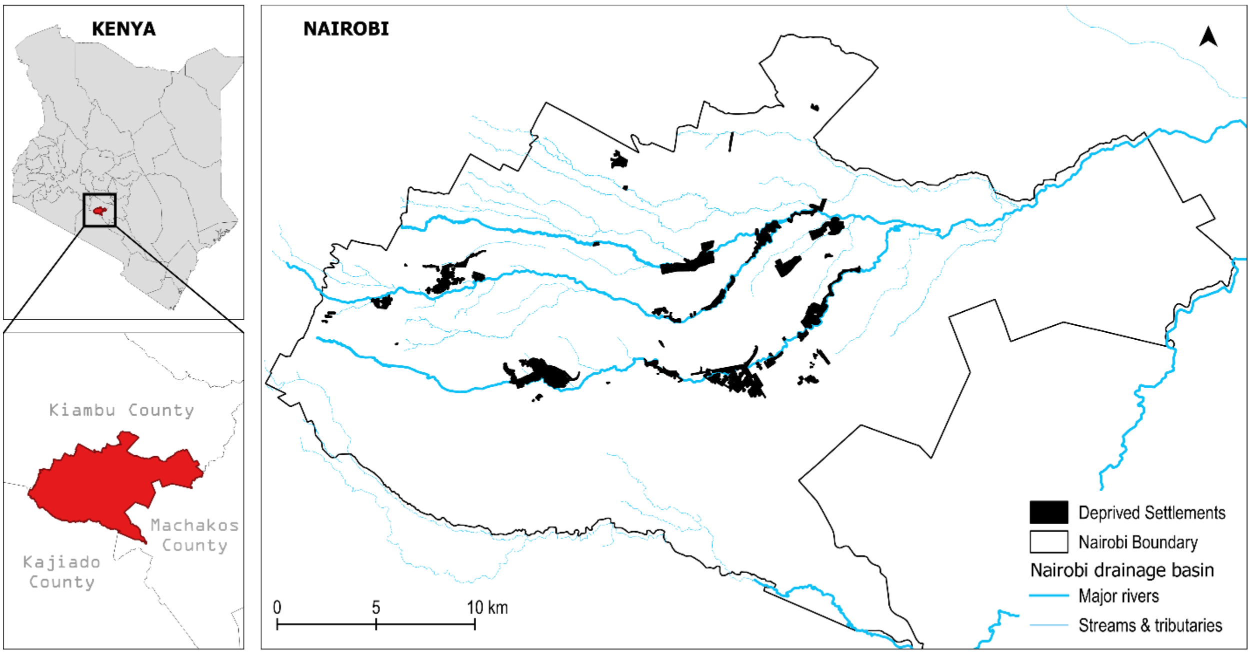


Figure 5: Case study area

The study area (Fig.5) is Nairobi, the capital city of Kenya and a central economic hub in East Africa. Nairobi has a population of 4.4 million people, which accounts for approx. 9% of the country's populace (Kenya National Bureau of Statistics, 2019). Similar to other colonial towns, Nairobi is still faced with the long-standing effects of residential racial zoning and rigid building standards, which were entrenched in the city's master plan of 1948 entitled "Nairobi Master Plan for A Colonial Capital" (Gatabaki-Kamau & Karirah-Gitau, 2004; Pamoja Trust, 2009). The plan remained the city's sole master plan until 2015 (Gatabaki-Kamau & Karirah-Gitau, 2004) (Fig.6). Such processes reflect poor urban governance systems, including the conduction of city boundary extensions in the absence of concrete master plans (Fig.7).

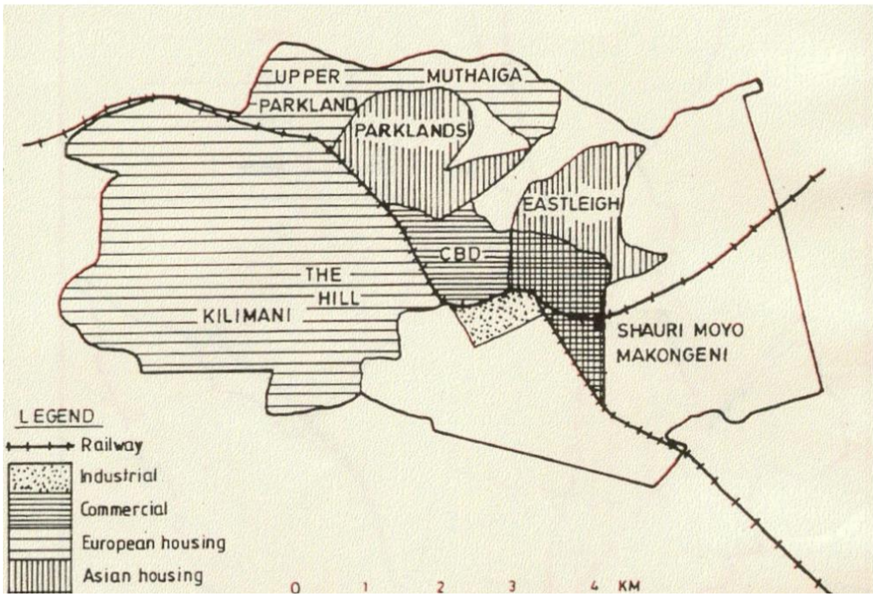
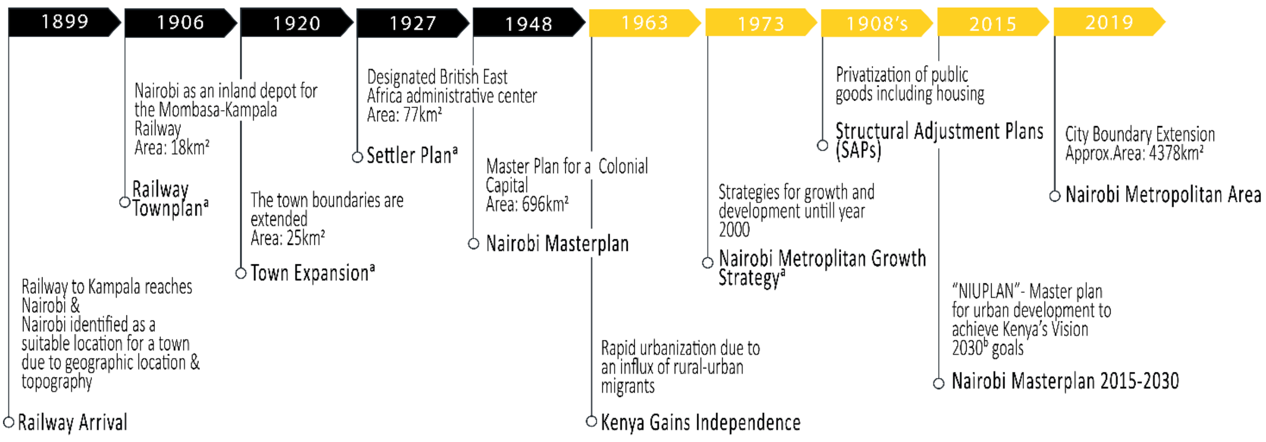


Figure 6: Nairobi Masterplan of 1948
Source: (Pamoja Trust, 2009)

The 1948 master plan was developed to establish Nairobi as a colonial capital from a railway headquarter. The plan carried on the concept of racially stratified residential neighborhood schemes introduced by the 1927 settler plan. As a result, the Europeans occupying more than 50% of the residential land - with the lowest densities - located in the north-western regions of the city, which have higher altitudes and well-drained soils (Gatabaki-Kamau & Karirah-Gitau, 2004; Gattoni & Patel, 1974). The Asians occupied the areas near the city center and industrial area, and; the Africans occupied rental hostel-type quarters in the low-lying eastern region of the city near the industrial area and train station, on land characterized by soils of poor drainage that are prone to flooding (Gatabaki-Kamau & Karirah-Gitau, 2004; Pamoja Trust, 2009). To date, many deprived neighborhoods are located in the eastern region of the city (Mwau & Sverdlik, 2020). In addition to the stack discrepancies in the location of non-native settlements, the housing conditions of Africans during the colonial era were temporary. Thus, the city is seen to have been typically designed for non-natives and enforced through anti-native policies. Notably, the *kipande* system restricted access of Africans into the city; and land and property rights were exclusive to non-natives (Home, 2014).



[a]Expounded on in a study by Driekx.D (2019), capturing pre-independence plans and policies in Kenya

[b]A long-term development blueprint for Kenya launched in 2008

Figure 7: Historical development of Nairobi

Upon attaining independence in 1963, African movement restrictions into the city were lifted, resulting in an influx of rural-urban migrants (Gatabaki-Kamau & Karirah-Gitau, 2004; Mitullah, 2003). In the absence of a new plan, Nairobi experienced a housing crisis that primarily affected the migrants and urban poor. Furthermore, the introduction of Structural Adjustment Programs (SAPs) in the 1980s led to housing privatization; thus, deprived settlements proliferated the cities (Mwau & Sverdlik, 2020). At present, approximately 95% of Kenya's urban housing stock is supplied by the private sector (individual and companies), including slumlords, with Nairobi having the highest proportion (86.4%) of households renting residential units (KNBS, 2018).

The post-independence administration also inherited discriminatory planning, and socio-economic segregation simply replaced racial segregation. The effects are reflected in the recorded 1972 residential densities within the European (8persons/acre), Asian (32 persons/acre), and African zones (400 persons/acre)(Dierkx, 2019). The trend continued, and as of 2009, slums occupied only 1% of the land in Nairobi and yet provided accommodation for 50% of the city's residents (Pamoja Trust, 2009). Furthermore, the 1948 master plan had been accompanied by the establishment of building codes and standards. These were rigid and unaffordable to the migrants and urban poor leading to the continuity of the hostel-type housing (single room with shared facilities) (Mwau & Sverdlik, 2020). Notably, this type of housing is still prevalent, sheltering approx. 67% of Nairobi households (KNBS, 2018) characterizing many deprived settlements (Mwau & Sverdlik, 2020).

3.2. Conceptualizing Settlements Using Earth Observation (EO) Data

To predict deprivation and compare the degree of hazardousness of deprived and non-deprived settlements, we identify four types of residential settlements within our study area. We use the generic slum ontological (GSO) framework, a hierarchical grouping framework for morphological deprivation (slums) developed by Kohli et al. (2012). The hierarchical order enables context-specific identification and differentiation of deprived settlements from the rest of the city. The framework has also been used to describe non-deprived settlements (e.g., Owusu, 2020).

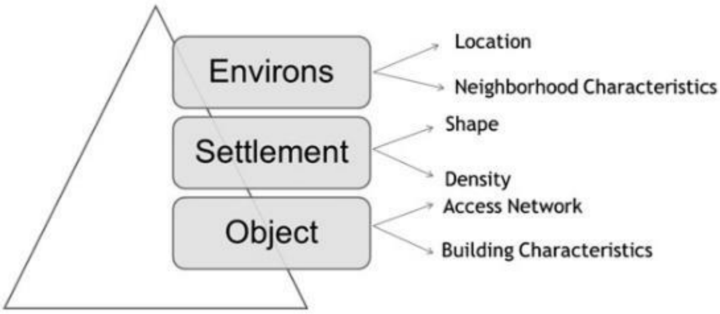


Figure 8: The hierarchical (spatial) conceptualization of slums with associated indicators. Source: Kohli et al. (2012)

Therefore, we use the GSO to describe different types of residential settlements considered in our study. The settlement selection is based on the availability of scientifically (stratified random sampling) generated label data for our machine learning processing (Vanhuysse et al., 2021). The label data comprise five classes: (i) high to mid-density built area, (ii) low density built area, (iii) deprived area (type I), (iv) deprived area (type II), and (v) large buildings/complexes (industrial/commercial) (Vanhuysse et al., 2021). Below we describe the four identified residential settlements using the GSO and show their locations within the city.

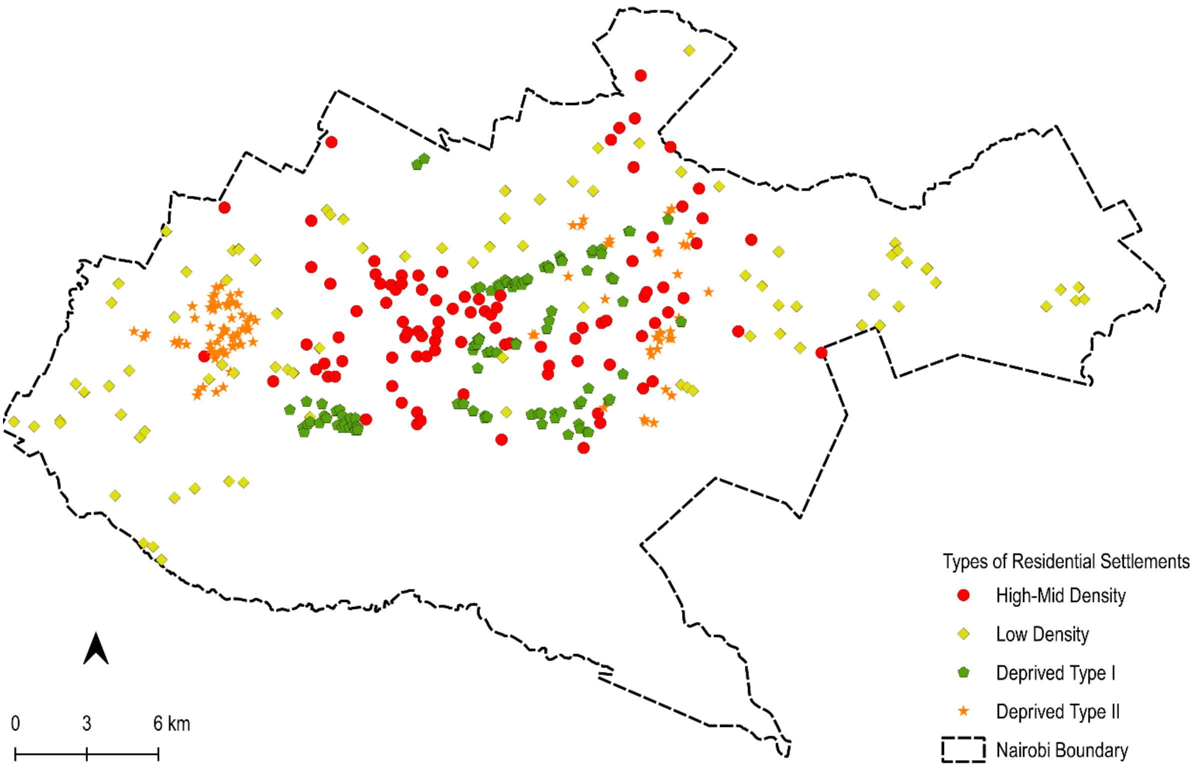
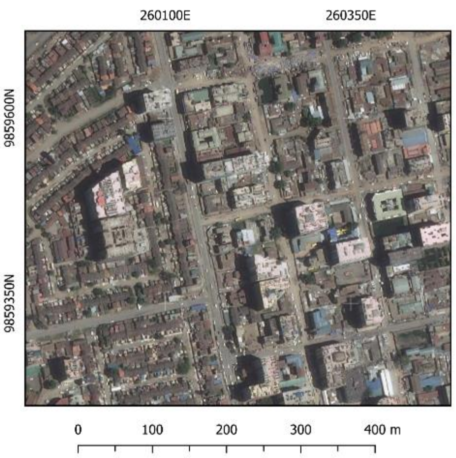


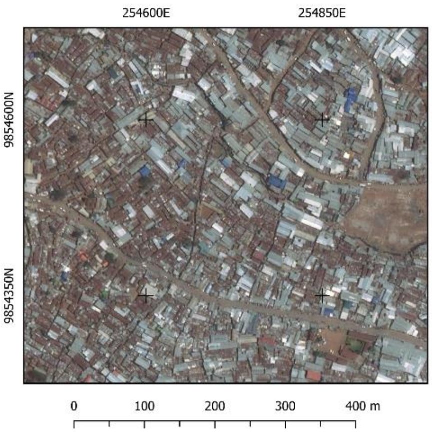
Figure 9: Distribution of different types of residential settlements in Nairobi

Table 1: Residential densities and slum settlement characteristics

Type of Residential settlement	Level	Indicators
Class 1: High to mid-density built area		
	Environs	**Location:** Close to the CBD **Neighborhood characteristics:** Single-family housing with compounds and apartment complexes.
	Settlement	**Shape:** Elongated street blocks. **Density:** High roof coverage with very little vegetation (usually trees). Mid-density settlements are located further from the CBD and have more vegetation (trees).
	Object	**Buildings**: Permanent building material with terrace roofing, tiles and coated iron sheets. **Access Network**: Well defined, regular street pattern.

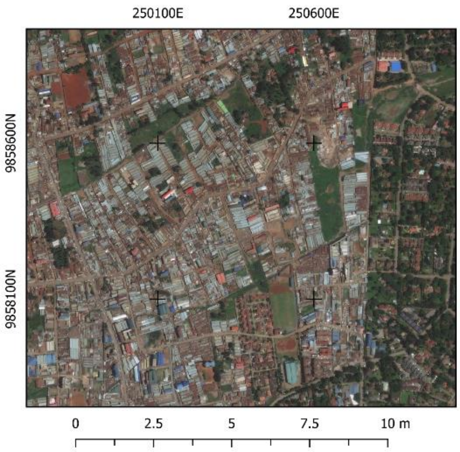
Type of Residential settlement	Level	Indicators
Class 2: Low density built area		
	Environs	**Location**: Urban periphery/suburbs. **Neighbourhood characteristics:** Large single-family housing
	Settlement	**Shape**: Large, regular street blocks. **Density**: Low roof and high vegetation (trees and lawns) coverage.
	Object	**Buildings**: Permanent building material with mostly tiles or coated iron sheets. **Access network**: Well defined, regular street pattern.

Class 3: Deprived urban area (Type I)



Environs	**Location**: Inner city and many are in 'evidently' hazardous locations-near the city's major drainage channels (river).
	Neighbourhood characteristics: Near the CBD and industrial area (employment).
Settlement	**Shape**: Tend to follow the shape of natural and man-made features e.g. rivers, roads and rivers.
	Density: Compact, with high roof coverage (>70%) and no (or very little) vegetation coverage.
Object	**Buildings**: Mainly temporary buildings, single-storey housing made of corrugated iron sheet roofing.
	Access network: Irregular street pattern. Very few paved/quality roads as they mainly rely on footpaths.

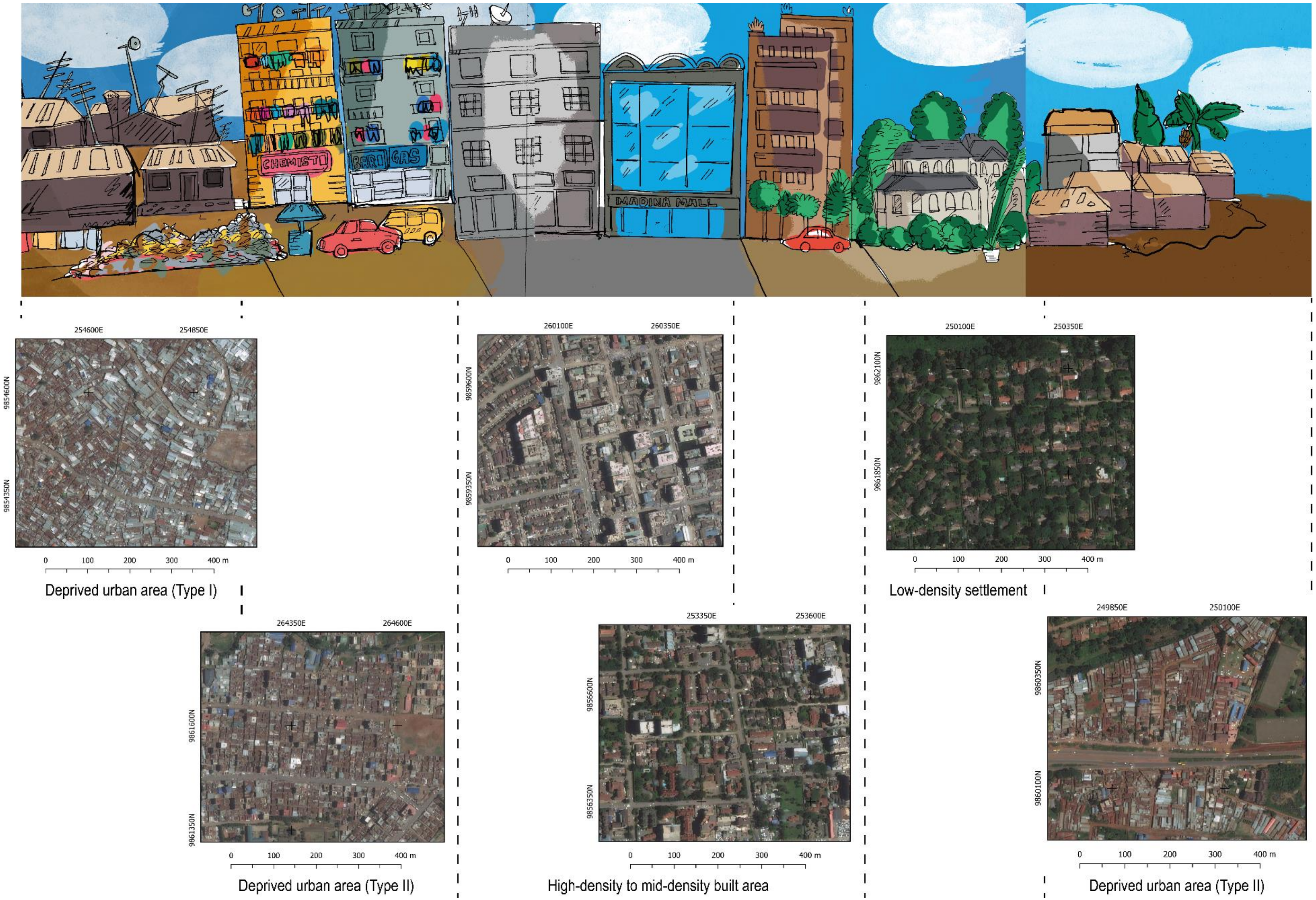
Class 4: Deprived urban area (Type II)



Environs	**Location**: More towards the city's periphery.
	Neighbourhood characteristics: Near higher-income neighbourhoods.
Settlement	**Shape**: Slightly regular with elongated street blocks.
	Density: Less compact than (type I), high to mid-density (>40%) and presence of some vegetation (trees, undeveloped plots or small farms depending on location).
Object	**Buildings**: Mix of temporary and permanent building material, single as well as multi-storey buildings. A mix of roofing material (corrugated iron sheets, tiles and terraces).
	Access network: Narrow, slightly regular streets.

We also sketch the horizontal transect of the settlements used in our study (fig.10) to conceptualize the *environs'* transition from one settlement to another. The sketch is only a representation and doesn't capture the entirety of the situation.

Figure 10: Horizontal section (sketch) representation of the transition among conceptualized settlements in Nairobi



3.3. Data and Software

In defining our area of interest (AOI), we select Nairobi county's (which is also the city's) political-administrative boundary (fig.5). Additionally, we obtain cloud-free Sentinel 2 surface reflectance multispectral imagery from European Space Agency (ESA). The imagery is downloaded using Google Earth Engine (GEE) for 2019, where the annual mean values are computed, and cloud masking is also undertaken. A similar approach is undertaken to acquire Land Surface Temperature from MODIS and air pollution data from Sentinel 5P, where the annual maximum values are computed. The Digital Elevation Model (DEM), a Synthetic Aperture Radar (SAR) Radiometric Terrain Corrected (RTC), imagery is obtained from the National Aeronautics and Space Administration (NASA) Earth Data portal.

Nairobi's land use map and building outlines were generated by Columbia University's Center for Sustainable Urban Development in 2010 and obtained through the World Bank data portal. The slum boundaries were obtained from a local company – Spatial Collective, and represent morphological slums. Ancillary data was obtained from Open Street Map (OSM). ESRI satellite imagery, accessed through QGIS, is used as a base map and conceptualizes settlements. Free and Open Source Software for Geoinformatics (FOSS4G) solutions are employed in our study. Specifically, QGIS is used for raster and vector data manipulation, KoBo Toolbox for primary data collection (household questionnaires), and R studio for advanced statistical manipulation, i.e., texture extraction and machine learning (annex.8.5). We, however, also use commercial software: ArcGIS 10.8.1-Topography Toolbox (Tom Dilts, 2015) for extracting the Height Above Nearest Drainage (HAND); ZOOM – a video teleconferencing platform for conducting key informant interviews; MS-Excel and SPSS for statistical analysis of our data. MS Excel is also used to present the outcomes of the statistical analysis.

Table 2: Data sources and description

Data	Resolution	Type	Description	Date	Source
Sentinel 2	10m	Multispectral	Multi-spectral	2019	ESA
Sentinel 5P	5.5km		Air pollution (CO,SO_2, NO_2 & O_3)	2019	ESA
MODIS	1km		Land surface temperature (LST)	2019	NASA
ALOS PALSAR	12.5m		Digital Elevation Model (DEM	2009/2007	NASA
Slum boundaries	-	Shapefiles	Morphological Boundaries of Nairobi's deprived settlements	-	Spatial Collective
Land use map	-	Shapefiles	Land use cover map of Nairobi	2010	Columbia University
Ancillary	-	Shapefiles	Polygon and line features	-	OSM
Building outlines		Shapefiles	Outlines of buildings in Nairobi	2010	Columbia University
Administrative Boundaries		Shapefiles	Politically administered boundaries, including AOI	2019	GADM
ESRI Satellite	-		Base map satellite imagery in QGIS		ESRI

3.4. Multi-Hazard Index

3.4.1. Identification of Hazards

To develop and localize a multi-hazard index, we review UN-Habitat's durable housing domain to identify hazards relating to deprived settlements. As a first step, we check the Emergency Events Database (EM-DAT) (https://www.emdat.be/) classification of disasters to UN-Habitats' measures of durable housing (Table 1), where we identify two broad hazard domains, i.e., natural and technological hazards (table 3). EM-DAT is a global disaster database operated and maintained by the Centre for Research on the Epidemiology of Disasters (CRED) (https://www.cred.be/).

Table 3: Hazard domain derivation from UN-Habitat 'Durable Housing' measures

Hazard group	**Hazard sub-group**	**Description** (EM-DAT, 2009)	**UN-Habitat durable housing measures**
Natural	Hydrological, e.g., floods and landslides	'A hazard caused by the occurrence, movement, and distribution of surface and subsurface freshwater and saltwater.'	Housing in geologically hazardous zones (landslide/ earthquake and flood areas)
	Geophysical, e.g., earthquake and volcanic activity	'A hazard originating from solid earth. This term is used interchangeably with the term geological hazard.'	
	Biological e.g., epidemic	'A hazard caused by the exposure to living organisms and their toxic substances (e.g. venom, mould) or vector-borne diseases that they may carry.'	Housing on or under garbage mountains
	Meteorological e.g. extreme temperature	'A hazard caused by short-lived, micro- to mesoscale extreme weather and atmospheric conditions that last from minutes to days.'	Quality of construction (e.g. materials used for wall, floor and roof)
Technological	Transport e.g. air, rail and road	'A hazard caused by transport-related accidents or incidents.'	Housing around other unprotected high-risk zones (e.g. railroads, airports, energy transmission lines)
	Industrial e.g. pollution and explosions	'A hazard caused by industry-related accidents or incidents.'	Housing around high-industrial pollution areas
	Miscellaneous e.g. fire and building collapse	'Any other hazard which may cause harm to a population or destruction of assets/property.'	Compliance with local building codes, standards and bylaws.

Next, we review the country's National Policy for Disaster Management (Government of Kenya, 2009). Despite the lack of city-specific categorization of disasters/hazards, we find that the main threats in the country are: "droughts, fire, floods, terrorism, technological accidents, diseases and epidemics" (Government of Kenya, 2009). Next, review the EM-DAT database for recorded disasters in Nairobi over ten years (2009-2019) (annex 8.2). From this, we find that; floods, fire, building collapse, transport accidents, epidemics, and industrial accidents (explosion) occurred. It is also noted that EM-DAT primarily focuses on the national scale and reported disasters. Thus, hazards like air pollution are not captured, and 'smaller' incidents may go unrecorded. We, therefore, review projects that focus on the city and settlement scale.

At the city scale, we identify the "Tomorrow's Cities: Urban risk in transition" project under current implementation in Nairobi. Their initial results indicate the following single hazards affect Nairobi: "(i) geophysical (earthquakes, volcanic eruptions, landslides), (ii) hydrological (floods and droughts), (iii) shallow earth processes (regional subsidence, ground collapse, soil subsidence, ground heave), (iv) atmospheric hazards (storm, hail, lightning, extreme heat, extreme cold), (v) biophysical (urban fires), and vi) space hazards (geomatic storms, and impact events)" (Malamud et al., 2021).

At settlement level, we review the IDEAMAPs framework of Domains of Deprivation (Abascal et al., 2021). Although their scope is global, they identify studies operationalized at the settlement level. Of interest to our study, we identify two categories: Contamination and physical hazards and assets. Under the physical hazards and assets, identified hazards are natural (flood zone, weather, and slope), ecological, natural assets, and nonspecific/multiple. Contamination comprises air pollution, garbage accumulation, industrial pollution, noise pollution, water pollution, and non-specific/multiple. Notably, the hazard categorization by both the Tomorrow's Cities project and IDEAMAPs Domains of Deprivation bare similarities with those used by EM-DAT. Differences are also noted and can be attributed to the difference in scale and the scope of focus.

Lastly, we conduct expert interviews to identify hazards affecting our study area at the city and deprived settlement levels. The outcomes are presented in the results section. The interview questions were prepared beforehand based on the literature review. They covered four main topic areas: (i) deprived settlements, (ii) hazards, (iii) durable housing and, (iv) ethical concerns/considerations (Table 4) (annex 8.3). The data from the interviews were analysed by identifying key themes. The response summaries are compiled, and descriptive analysis is undertaken.

Table 4: Interview topic areas and key questions

Topic	Main Question
Deprived Settlements	What data is helpful in slum mapping, and who are the involved actors in data use and generation?
Hazards	What hazards affect Nairobi and the informal settlements in the city?
Durable housing	Does the location and type of housing protect against hazards?
Ethical concerns	Does slum mapping using improved technology (e.g., AI and VHR imagery) pose a threat to the privacy of slum residents?

Key informant interviews were conducted with experts working in the urban or disaster risk fields and residents of informal settlements. The experts were selected due to their experience working in deprived settlements or analysing urban poverty in different types of organizations and professions to capture diverse views on the subjects (Table 5).

Table 5: Experts on urban and/or disaster risk and their roles in slums

Designation	Role
Urban Policy Analyst	Evidence provision to inform decision making (advisory role)
Urban Systems Officer	Developing and implementing urban acupuncture projects (e.g. space activations)
Program Officer (Human Settlements)	Capacity building for municipalities and communities and systematic approach to slums interventions-strategies, plans etc.
Spatial Data Expert	Spatial data production and analysis for monitoring progress towards achieving SDGs
Professor in Geography	Researcher in community-based vulnerability risk assessment
Kibera resident (1)	Community leader – born and raised in Kibera
Kibera resident (2)	Data collection enumerator

3.4.2. Construction of the Multi-Hazard Index

We identify spatial multi-hazard index construction principles outlined by Greiving (2006) to inform this study. Although developed for a multi-risk index, three of the four principles are relevant for developing a multi-hazard index. The principles are also in line with the definition of multi-hazards considered in this study (Melanie Simone Kappes, 2011). These are: (i) non-sectoral, meaning the consideration of hazards should incorporate different sectors; (ii) the hazards should have spatial relevancy; and (iii) collective hazards are what should be considered (Greiving, 2006). Under the second characteristic, Greiving (2006) highlights that 'ubiquitous' risks (including epidemic and traffic accidents) should not be considered. These two components are fundamental in urban/spatial planning[1]; therefore, we still consider them in our study.

To construct the index, we identify open geospatial data indicative of hazardousness following extensive literature search and outcomes of the expert interviews (section 4.1.1). We identify appropriate geographic and EO-based variables that capture the hazards through comprehensive data search in global and local repositories. Due to data limitations, we drop some previously identified hazards, e.g., geophysical hazards, and building collapse, resulting in 6 hazard indicators and 18 variables (Table 6). The selected variables include primary variables (i.e., temperature, air pollution) and proxy variables (e.g., geomorphon).

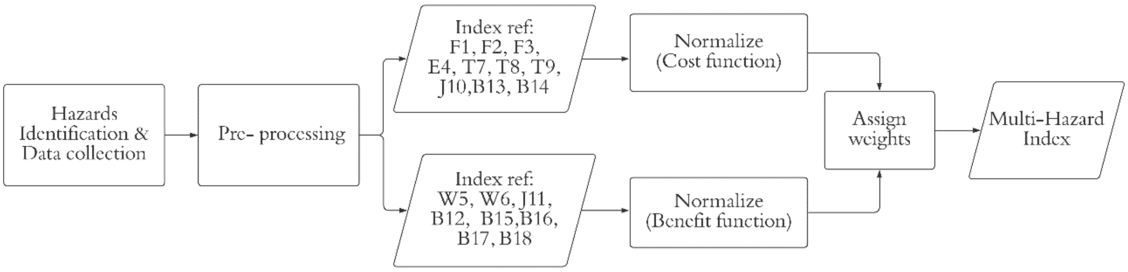


Figure 11: Susceptibility to hazards index workflow

After identifying relevant data for the indicators, we follow the steps as outlined (figure 13). We start by pre-processing the data, including projecting the data to Nairobi's Coordinate Reference System(CRS) - EPSG:32737 - WGS 84 / UTM zone 37S, masking, and clipping to our AOI, and cleaning the data. Next, we process all the vector data into raster format. All the data are resampled to 10m, our chosen unit of operation, for consistency purposes- given that we rely on Sentinel 2 data (10m resolution) for further analysis (following sections). We also code the indicators by assigning the first alphabet of the sub-hazard group to which it belongs, followed by a number (from 1-18, the total number of indices present). For industrial hazards, we use the letter J instead of I to avoid confusion with number 1.

All the data are normalized, resulting in values ranging from 0 to 1 (lowest to highest indication of hazardousness). Normalization of data is essential since it minimizes complexity and allows us to compare the indicators. Data normalization's cost and benefit functions are used for the study and operationalized using the raster calculator tool in QGIS.

$$[(value - min)/ \ range]$$

Equation 1: Benefit normalization formula

$$[1 - ((value - min)/ \ range)]$$

Equation 2: Cost normalization formula

Specifically, the LST, air pollution, and building and industries densities data are normalized using the benefit function since higher values indicate a higher likelihood of hazardousness. On the other hand, the road network density, NDVI, proximity data, and HAND are normalized using the cost function since lower values indicate a higher likelihood of hazardousness. Lastly, we assign equal weights to each of our six main hazard groups. Thus, each hazard group is accorded equal importance. Equal weightage is considered since we lack access to data that could be used to compute the weights (e.g., frequency of hazards). The weights are then distributed among the sub-hazard groups and all 18 data indicators (table.6).

[1] John Snow: the cholera epidemic of London in 1854 & Transport Oriented Development

Table 6: Hazard indicators, their descriptions and properties

Hazard	Weight	Sub-Hazard	Weight	Code	Data	Weight	Data Description	Resolution	Measurement
Flood	0.167	Riverine floods	0.0835	F1	Height Above Nearest Drainage (H.A.N.D)	0.042	Height relationship of locations to nearest natural drainage (extract of DEM)	12.5m	Vertical distance (m)
				F2	Proximity to Rivers	0.042	Distance from major river drainage system of the city	10m	Euclidean distance (m)
		Run-off	0.0835	F3	Geomorphons	0.0835	Terrain form of the city extracted from DEM	12.5m	Classified absolutes values
Epidemic	0.167	Epidemic	0.167	E4	Proximity to Garbage dump sites	0.167	Distance from the city's dumpsite/landfill	10m	Euclidean distance (m)
Weather and climate	0.167	Extreme temperatures	0.167	W5	Day Land Surface Temperature (LST)	0.0835	Daytime radiative temperature of the city	1000m	Kelvin (K)
				W6	Night Land Surface Temperature (LST)	0.0835	Night-time radiative temperature of the city		
Transport accidents	0.167	Rail accidents	0.0557	T7	Proximity to Railway lines	0.0557	Railway lines cutting through the city	10m	Euclidean distance (m)
		Road accidents	0.0557	T8	Proximity to major roads	0.0557	Major roads and highways of the city	10m	
		Aero accidents	0.0557	T9	Proximity to Airports	0.0557	Airport boundaries/runways and infrastructure	10m	
Industrial accidents	0.167	Industrial accidents	0.167	J10	Proximity to Industries	0.0835	Proximity to industries	10m	
				J11	Density of industries	0.0835	Number of industries outlines per unit area ($1km^2$)	10m	Density (No. of features/km^2)
Biophysical hazards	0.167	Fire	0.0835	B12	Density of buildings	0.0278	Number of building outlines per unit area ($1km^2$)	10m	
				B13	Road density	0.0278	Number of roads (line segments) per unit area ($1km^2$)	10m	
				B14	NDVI	0.0278	Density of plant growth	0.01 arc degrees (approx. 11.1 km)	mol/m^2
		Air pollution	0.0835	B15	Sulphur Dioxide (SO_2)	0.021	Vertical pollutant column density - a ratio of pollutant and total air mass factor		
				B16	Nitrogen Dioxide (NO_2)	0.021			
				B17	Ozone (O_3)	0.021			
				B18	Carbon Monoxide (CO)	0.021			

3.4.3. Indicator Description and Relevance

3.4.3.1. Flooding

Kenya is affected by flooding, and it accounts for approximately 60% of disaster-related fatalities (UNDP, 2012). From our evaluation of the EM-DAT disaster database of Nairobi and expert interviews, flooding events are attributed to riverine flooding and run-off due to heavy rainfall.

I. Riverine floods

Most of Nairobi's slums are located along the city's natural drainage system (fig. 5). Many settlements encroach on the riparian reserves, and there are hardly any protection measures to prevent flooding into the settlement. Due to this, we select two data indicators for assessing the susceptibility of settlements to flooding. Proximity to major river tributaries is considered a factor of Euclidean distance. We assume that locations near the river are more likely to be affected by riverine floods than those further away.

To compute the Euclidean distance, we filter OSM water body features in QGIS. Using Boolean operations, we select only the rivers and major streams. Due to incompleteness, we substitute the OSM data with rivers classification from the LU dataset. For polygon to line conversion, we use the HCMGIS plugin to extract centrelines. This data is supplemented using the Nairobi Land Use data. For correctness, we check the filtered results against the city maps. Next, we convert the features from vector to raster, and using the Proximity (Raster Distance) function in QGIS, we compute the Euclidean distances of the city from major rivers.

In addition to considering Euclidean distances from rivers, we also compute the *Height Above The Nearest Drainage* (HAND). HAND is the vertical distance measure of locations to the nearest drainage channel (Nobre et al., 2011). The vertical distances are computed from a DEM and drainage channel. We use HAND since river inundation is more likely to affect flatter areas, despite their Euclidean distance from the drainage channel. To create HAND, we use the major rivers and DEM as input and operationalize the process in ArcMap, using the Riparian tool within the Topography Toolbox developed by Tom Dilts (2015). The process entails: (i) flattening the DEM, (ii) ingraining drainage lines into the flattened DEM, (iii) computing the height above nearest drainage for each cell.

II. Runoff

Runoff is affected by: (i) meteorological factors and (ii) physical characteristics (USGS, n.d.). Under the meteorological factors, precipitation and its different factors such as intensity, amount, and duration are identified (USGS, n.d.). These are important since the IPCC (Revi, David E., et al., 2014) projects an increase in weather events in East Africa. The effects are already being felt in Kenya, with increased annual temperatures and more intense rainfall reported (GoK), 2010). With the increase in rainfall, flooding due to surface runoff is anticipated to increase. Therefore, the physical characteristics of our study area are considered indicators of runoff flooding.

Under physical characteristics, we consider the soils and topography. Run-off occurs when the absorption capacity of the soil is saturated (Kim Rutledge et al., 2011). The clay material in the soils in our study comprises kaolinites, interstratified, and montmorillonites. Kaolithic clay soils are prominent in the northwestern region of the city, while the montmorillonites are in the central and eastern regions. Of the three clay mineral components, montmorillonites cover 68% of the city and are more prone to floods since they expand while wet hence retain more surface water (Aksu, Bazilevskaya, & Karpyn, 2015). Data on the soils are, however, too coarse and are exempted from our analysis. Therefore, we focus on the topography by creating geomorphons. Geomorphons are terrain forms comprising most common landforms: PK-peak; RI-ridge; SP-spur; SL-slope; SH-shoulder; FS-foot slope; FL-flat; HL-hollow; VL-valley; and PT-pit.

Additionally, studies have found soils correlated to geomorphons (Jasiewicz & Stepinski, 2013; Silva et al., 2016).

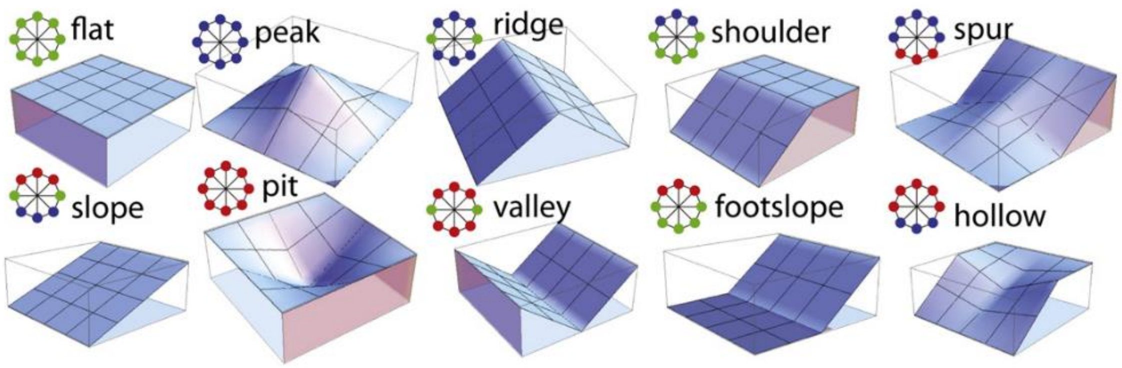


Figure 12: Geomorphons - the ten most common landforms and respective 8-tuple representations of a pixel's surface relief in relation to its neighbours within line of sight.
Source: Jasiewicz & Stepinski, (2013)

To create the Geomorphons, we first resample the 12.5m ALOS PALSAR DEM to 10m using the GRASS raster resample (r.resample) function in QGIS. Next, using the GRASS geomorphon function in QGIS, we vary the search radius (L) as (3,5,10,15,20,25,30,35,40,45,50,100,150 and 200 pixels). The search radius (L) determines the scope of the terrain search. Large L radius gives a global overview of the terrain, whereas smaller L values are localized. Geomorphon with L=100 is selected based on visual representation. This approach was undertaken due to the limited time and scope of this study.

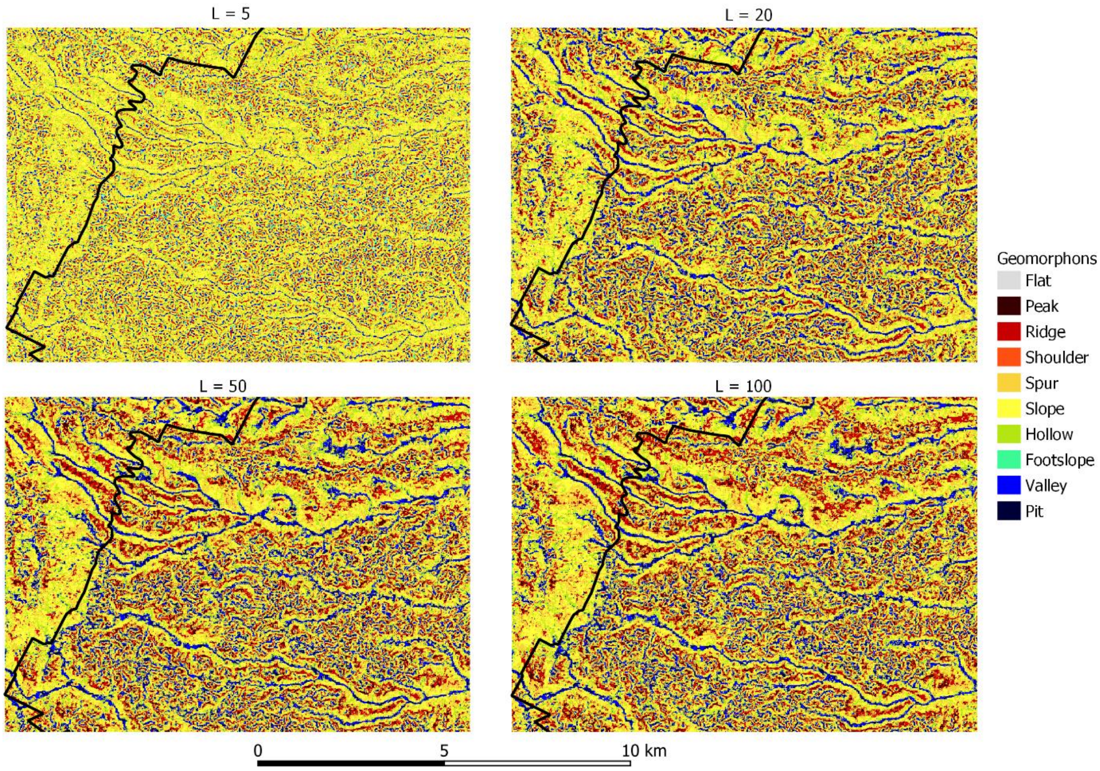


Figure 13: Different number of cells as radius for creating Geomorphons

Next, we vary the skip radius (inner radius) from 0 to 40 with intervals of 10. Skip radius helps eliminate irregularities (Jasiewicz & Stepinski, 2013). We maintain the system default for the flatness threshold (10) and flatness distance (0) parameters through the process. At a skip radius of 50, we obtain the geomorphons which we categorize in order of most to least susceptible to surface run-off floods (PK-peak; RI-ridge; SP-spur; SH-shoulder; SL-slope; FL-flat; FS-foot slope; HL-hollow; VL-valley; and PT-pit). Our logic guides the categorization that high relief terrains are less prone to flooding than lower relief terrains.

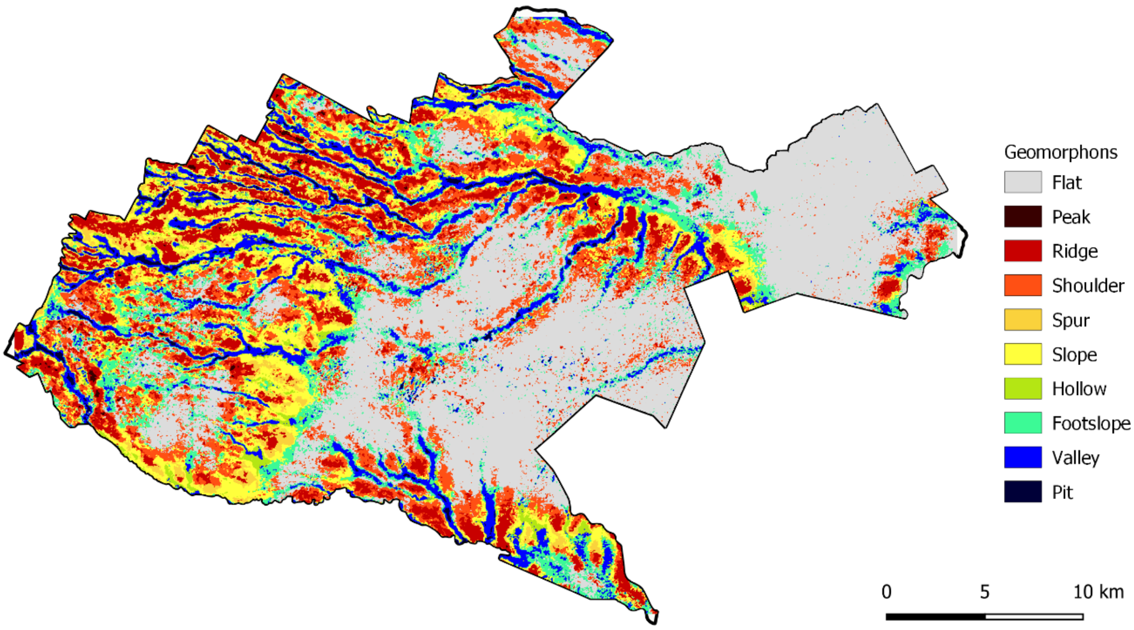


Figure 14: Geomorphons created using L=100 and Skip radius = 50

The categorization is based on the 8-tuple representations of each of the ten generated terrain forms. We first consider and rank pit, flat, and peak terrains (represented having pure tuples) from most to least likely to flood (respectively). Next, we check the degree of purity of the remaining seven terrain forms to the peak and pit terrain forms resulting in the following ranking: PK, RI, SP, SL, HL, VL, and PT from least to most susceptible to runoff. We have SL as the medium terrain form and are left with SH, FL, and FS. Next, provided we have an even number of terrain forms and cannot equally separate them (to place the SH, FL, and FS in the ranking). We consider FL (having pure tuples) to rank SH and FS. Next, provided SL and FL are the terrains in the middle, we rank SL as being less likely to be adversely affected by runoff in comparison to FL (since runoff would quickly accumulate on flat terrain).

Table 7:Summary of terrain forms ranking process from least to most likely to be affected by runoff

1st Ranking	Peak					Flat				Pit
2nd Ranking				Shoulder		Flat	Foot slope			
	Peak	Ridge	Spur	Slope				Hollow	Valley	Pit
3rd Ranking	Peak	Ridge	Spur					Hollow	Valley	Pit
Final Ranking	Peak	Ridge	Spur	Shoulder	Slope	Flat	Foot slope	Hollow	Valley	Pit

3.4.3.2. Epidemic

Infectious diseases are a major health concern since they are one of the leading morbidity causes globally (Adiga et al., 2018). Infectious diseases include water-borne (e.g. cholera), vector-borne (e.g. malaria) diseases as well as respiratory diseases (e.g. COVID-19). For our study, we consider "proximity to garbage sites as an indicator" of infectious diseases and for the data we use the location of Nairobi's sole landfill – Dandora landfill (UNEP, 2018). Garbage sites are 'hot-spots' for infectious diseases. The sites act as breeding grounds for living organisms such as insects and rodents, which are carriers of vector-borne disease. They have also been found responsible for respiratory diseases caused by air pollution resulting from burning garbage (e.g. UNEP, 2018); and water-borne disease outbreaks such as cholera - which is one of the global public health threats, due to contamination of water sources (e.g. UNEP, 2018). Therefore, high proximity to dumpsites can be considered hazardous. We extract the vector feature of the city's landfill from the land use data for the index. We then rasterize the feature and compute Euclidean distances following the same procedure used for the river features.

3.4.3.3. Extreme temperatures

Urbanization driven land use and land cover changes have resulted in the modification of , the urban micro-climate. This is manifested as an urban heat island (UHI), where the temperatures in urban areas are higher in comparison to their surroundings (Seto & Shepherd, 2009; Zhou, Zhao, Zhang, Sun, & Liu, 2015). Furthermore, extreme temperatures are predicted for East Africa due to climate change (Revi, David E., et al., 2014), and are already being experienced in Kenya (GoK, 2010). To capture this phenomenon, we use day time and night time LST data. LST is a measure of radiative emissivity of the earth's surface. It has been used to analyse the UHI phenomenon (e.g. Zhou et al., 2015), and in mapping deprived areas where J. Wang et al. (2019) found that slum locations are exposed to higher local temperatures. For the LST, we consider the maximum temperature since it is a recommended unit for assessing climate extremities (Gallina et al., 2016). Additionally, maximum values indicate the point at which the highest temperatures are reached within the city and are therefore a suitable indication of susceptibility to hazards. As aforementioned, these data are downloaded using GEE.

3.4.3.4. Transport accidents

The proximity to transport modes pose high risk of accidents which may cause injuries and fatalities.. They also expose the residents to air and noise pollution Literature on transport accidents, however, majorly focus on road related accidents, probably due to their infamous occurrence. In Kenya, reported causes of road accident fatalities in a recent study are attributed to driver related causes such as 'running over victims' (Muguro, Sasaki, Matsushita, & Njeri, 2020). From this we infer that secondary causes of accidents such as the lack of pedestrian walk ways that are often not reported, influence the occurrence of road accidents. Furthermore, no literature is found on rail and air accidents. Therefore, we use proximity to road, rail and air transport infrastructure as indicators to identify high-risk areas where transport accidents are likely to occur. We compute the Euclidean distances of the city from the three transport infrastructure. The transport infrastructure features are obtained through the filtering of OSM data in QGIS. Next, we rasterize the vector features and we compute the Euclidean distances of the three features.

3.4.3.5. Industrial accidents

Industrial accidents are caused by anthropogenic factors, including lack of monitoring and maintenance of industrial infrastructures and; are also triggered by natural hazards such as extreme weather events and physical hazards (UNECE, 2021). Particularly, chemical accidents related to the oil and gas refineries, storage and pipelines have resulted in significant deaths, serious injuries and economic losses at a large scale (e.g. Mutiga, 2011). Such cases affect deprived settlements that are often located close to industrial infrastructures and lack adequate disaster-coping mechanisms and infrastructures. Additionally, the nature of the settlements, typically crowded, only worsens the situation. To account for industrial accidents, this study considers the proximity to industries and the density of industries as indicators. Industries data is obtained from OSM and land use data. The Euclidean distance from industries is then computed, and using the Kernel Density Estimation tool in QGIS, the heatmap indicating densities of industries per km2 is calculated. Due to the unavailability of data on industrial infrastructures like oil pipelines, we only consider the building locations of industries in the study.

Further, we note that the triggering effect of natural disasters on industrial accidents and particularly chemical accidents (referred to as Natech hazards) are gaining recognition as an emergent risk globally (Krausmann, Cozzani, Salzano, & Renni, 2011; UNECE, 2021). These interrelations highlight the need for an integrated approach to investigating and managing hazards.

3.4.3.6. Biophysical hazards

I. Fire

Urban fires are frequent in deprived settlements within Nairobi. These fire outbreaks are attributed a number of factors including poor power connections, and drunkenness (Ngau & Boit, 2020). Besides these causes, fires due to explosions of oil pipelines were also highlighted by two experts. In spite of the causes, slum conditions intensify fire outbreak incidents due to their high density and compactness; lack of open spaces to provide safety; and combustible building materials (Ngau & Boit, 2020). The poor road connectivity further hinders responses to the fires within the settlements (Ngau & Boit, 2020). These secondary factors are what we consider for our index except for building material since spatial data or proxies couldn't be identified. Therefore, we consider three factors: (i)road density, (ii) building density and (iii)normalized difference vegetation index. Starting with the road density, we use OSM data where we first filter out footpaths, etc., from the data and remain with a selection of roads, which are wide enough to allow fire extinguisher tracks and services. Then, using the QGIS line density interpolation tool, we compute the density of roads per km^2 for the city. To obtain the building density measured as the number of buildings per km^2, we manipulate the building outline data using the Kernel Density Estimation tool in QGIS. Lastly, in QGIS, we also create the NDVI from the Sentinel 2 multispectral image using the red and NIR (near-infrared) bands by applying the formula:

$$NDVI = (NIR - Red)/\,(NIR + Red)$$

Equation 3: Normalized difference vegetation index

II. Air pollution

We only consider outdoor/ambient air pollution in compounding the index since we cannot capture indoor air pollution using geospatial data at a city-wide scale. Ambient air pollution is caused by excessive atmospheric gases and particulate matter (World Health Organization (WHO), 2018). Currently, air pollution is considered the most prominent environmental health risk factor, especially in cities, since it is a major cause of morbidity and respiratory diseases (WHO, 2016). The reported air pollutants with substantial evidence of causing health implications are Particulate Matter (e.g. PM2.5, P.M10), Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Sulphur Dioxide (SO_2), and Ozone (O_3) (WHO, 2021b). Therefore, we consider these pollutants, except PM – due to data unavailability, for our analysis. These data are obtained from sentinel 5P. They are measured as a ratio of pollutants to air mass factor. The data are obtained as the annual maximum for the city and resampled to 10m. Maximum values are considered using the same logic highlighted for extreme temperature.

3.5. Application of Multi-Hazard Index to Predict Deprivation

3.5.1. Statistical Discriminant Analysis

After constructing the multi-hazard index, we first statistically evaluate whether the multi-hazard datasets can be used to predict deprivation. Our dataset comprises the multi-hazards and the grouping variable (label data) containing four classes (section 3.2): (i) high to mid-density built area, (ii) low density built area, (iii) deprived area (type I), and (iv) deprived area (type II). Multi-hazard datasets from our susceptibility index are first analysed as the dependent variables using multivariate analysis of variance (MANOVA) performed using SPSS software.

Next, we carried out a discriminant function analysis to evaluate the performance of outcome variables in discriminating the classes. The multi-hazard dataset represents the outcome variables, and the grouping variables represent the dependent variables. For each outcome variable, there are underlying linear dimensions called variates used to determine the discriminant functions (Field, 2017). These linear variates are used for determining group membership predictions and can be described using a linear regression model function (Equation 4) (Field, 2017).

$$y_i = b_0 + b_1 x_1 + b_2 x_2 + \cdots b_n x_n$$

Equation 4: Linear regression equation

Where: X_1 = outcome variable 1 and b = weights indicative of each variable's contribution to the variates.

The *b* values are calculated from eigenvectors that assess the ratio of systematic variance to unsystematic variance (SS_M/ SS_R) for the underlying variates to the functions and reduce the dimensionality of the dataset (Field, 2017). From eigenvectors, we obtain eigenvalues that measure the degree of freedom of our model (similar to F-statistics)(Field, 2017). The larger the eigenvalue, the higher the variance between the linearly combined variables. On the other hand, the functions (e.g., Y_i) representing the ratio of variability within the outcome variables as explained by the model (SS_M) and error in prediction (SS_R) maximize group differences using a linear combination of outcome variables (Field, 2017). Lastly, the number of discriminant functions is determined by *K*-1 (where *K* is the number of categories/group memberships) or the number of variates which is equivalent to *p* (number of outcome variables) (Field, 2017).

Given that we have four groups, three statistically significant discriminant functions were obtained (Table 13). The first discriminant function has a maximized ratio of variability since it tests the model as a whole. As a result, the first function is the most powerful discriminating dimension. Subsequent functions control the preceding functions while following a similar approach. Canonical correlation analysis is also carried out to assess the strength of association between the multi-hazard datasets and the categorical groups. They help explain the nature of the variates. Lastly, predictive classification is undertaken, with the outputs given as an overall accuracy of the model and the cumulative accuracies.

3.5.2. Predicting Deprivation Using Random Forest Classifier (RFC)

The results of the discriminant analysis statistically prove that multiple hazards can be used to predict deprivation. First, we perform land cover classification to extract built-up areas that serve as input for the deprivation prediction. Next, we use the multi-hazard index as variable input for the RFC. To test the accuracy of multi-hazard predictors, we also use conventional variables (spectral-textual features) and compare the models' performance (fig.15).

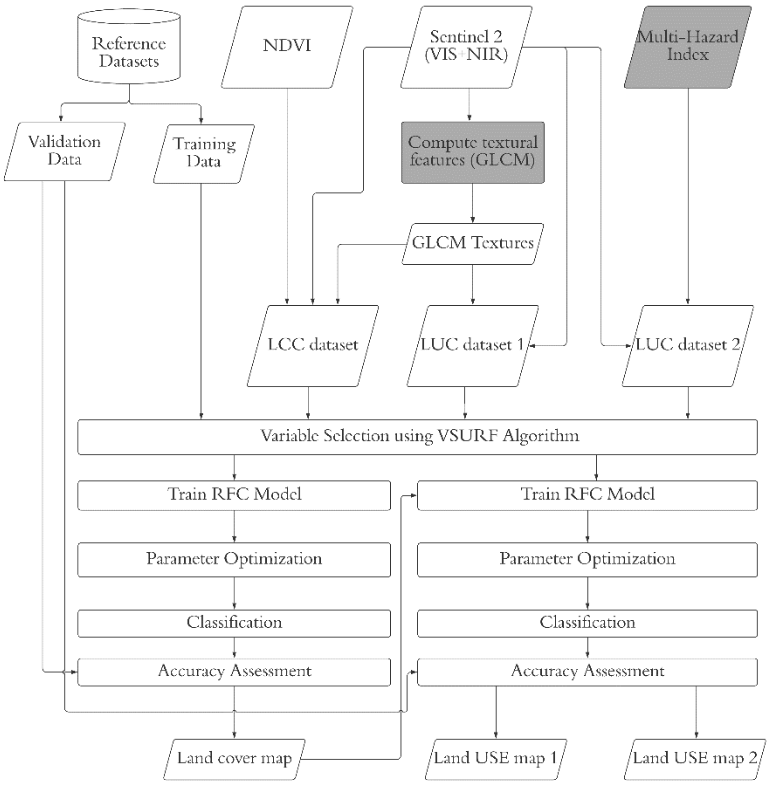


Figure 15: Land cover land use classification process

3.5.2.1. Extracting Features

Studies show that spectral, textural, and contextual features are important for image interpretation (Haralick, Dinstein, & Shanmugam, 1973; Kuffer, Pfeffer, & Sliuzas, 2016). Textural features, in particular, have proven to be effective complementary data to the more readily available data (spectral and contextual) for land use land cover classifications (LULC) (Engstrom, Harrison, Mann, & Fletcher, 2019; Kuffer et al., 2020). Textural features capture the grey tone variation of a surface and the spatial statistical distribution of the tones that reveal the structure of the surface and its relationship to its environment (Haralick et al., 1973). Therefore, we extract the Grey-Level Co-Occurrence Matrix (GLCM), a commonly used texture measure for LULC classification. Specifically, the use of GLCM has yielded good results for extracting built-up areas using Sentinel-2 (10m resolution) (Saini, Verma, & Gautam, 2021), as well as for deprivation mapping using random forest (Kuffer, Pfeffer, Sliuzas, & Baud, 2016).

Our study extracts eight common textural measures using 10m resolution Sentinel-2 red, green, blue (VIS-visible bands), and near-infrared (NIR) bands. These textural features are: Contrast, Entropy, Mean, Dissimilarity, Homogeneity, Angular Second Moment (ASM), Correlation, and Variance. These are generated in R Studio using the GLCM package. We set the user-defined parameters of window size (kernel) and shift. Kernels apply a function to the central pixel based on the neighbouring pixels. They, therefore, not only deal with noise in data but also influence the performance of models. Thus, it is important to select one that best fits the data characteristics. Using a scaling factor of two, we use varying kernel sizes ranging from 5X5 to 27X27 for each of the four bands and, applying a shift factor of 1. As a result, we have 416 textural bands collectively described as high dimensional data (or big data).

3.5.2.2. Variable Selection

To reduce the dimensionality of our data and by identifying the best kernel size for classification., we use the Variable Selection Using RF (VSURF) algorithm, implemented in R Studio. VSURF is a CART-based model that reduces data dimensionality through feature selection (Genuer, Poggi, & Tuleau-Malot, 2015) that has proven robust with geospatial data (Lou et al., 2020), especially for optimizing LULC classification (Georganos et al., 2018). The algorithm uses a step-wise approach for variable selection in three steps (Genuer et al., 2015). First, noisy variables - with the least importance are eliminated. The variables with the smallest OOB error are selected, and thirdly, variables for prediction are selected if the OOB error decrease is higher than the model's threshold (> than the mean variation with noisy variables) (Genuer et al., 2015).

3.5.2.3. Parameter Optimization

The optimal parameters in Random Forest (*ntree* and *mtry)* are determined using iterative tunning operations. The first parameter that we tune is *ntree*, which indicates the number of trees used to build the model. For both land cover and land use classification, the optimal value is determined by starting the value at 0-5000 and varying the intervals by 500 until the learning curves of each predictor class (including OOB samples) stabilizes. While optimizing *ntrees*, the *mtry* values are kept at default, i.e., *mtry*=√*number of variables*. *Mtry* represents the number of nodes to be split in each tree. After finding stable values for *ntrees*, *mtry* is optimized by varying the value starting with the number of predictor variables in each scenario.

3.5.2.4. Land Cover Classification

To perform land cover classification, we identify four classes of land cover in the study area, i.e. built-up, bare land, vegetation and water. Within each category, we further identify sub-classes to capture the diverse nature of our study area. The built-up sub-classes are mentioned in previous sections (high-mid density, low density, deprived type I, deprived type II, including non-residential buildings) (Vanhuysse et al., 2021). Non-built classes are generated from OSM data, and specifically, for bare land, unpaved roads, and vegetation cover, the process is supplemented by manual sampling. Given the diverse vegetation cover in the study area, we increase the sample data in this sub-class. To ensure correctness, we use visual assessment to validate the label samples. In total, we generate 1219 labelling samples (table.8). The labelling data is randomly split into 70% for training and 30% for validation. Since we use a random forest algorithm with an internal validation system, we don't generate testing samples but rely on the OOB samples to evaluate the model's classification performance.

Table 8: Label data sampling scheme for land cover classification

Land Cover	Sub-Class	No. of Samples	Total
Built-up	High-Mid Density	94	579
	Low Density	86	
	Industrial/Commercial/Administrative	100	
	Deprived Type I	100	
	Deprived Type II	100	
	Tarmac	99	
Bare land	Bare Soil	108	213
	Untarmacked Roads	105	
Vegetation	Trees	118	309
	Vegetation	191	
Water	Rivers, streams, dams, and ponds	118	118

The predictor data comprises two datasets. One that comprises GLCM textures of kernel size 3X3 in combination with the spectral bands and NDVI. These data have been successfully used to extract urban built-up areas using Sentinel 2 imagery (Saini et al., 2021). The second dataset comprises the entire GLCM textures in combination with spectral bands and NDVI. The classification is performed using RF model.

3.5.2.5. Predicting Deprivation

Next, we conduct a predictive classification of deprived settlements – a land-use classification problem. For the classification, we identify six thematic land uses: deprived type I, deprived type II, low-density residential, high-mid density residential, non-residential (commercial/industrial/administrative), and non-built areas (Table 9). The label data for the land uses are sampled using a random stratified strategy. Since the heterogeneous nature of the urban regions deters land use classification, we use polygons of 50X50m as label data ensuring that we maximize the number of pixels used to identify the thematic classes. We immitate the classification process used by polygon reference datasets and generate points for each pixel corresponding to the polygons. This approach is undertaken due to the long processing time to run the model using polygon reference data. Given that our raster data has a 10 m resolution, the total number of extracted sample points is 14600. However, these sample data result in overfitting of the data and incur a long processing time. Therefore we sample a third of the points from within the labelled dataset, resulting in 4865 samples. We split the data into 70% for training and 30% for validation.

Table 9: Sampling scheme for land deprivation prediction

Land Use	No. of Sample Polygons	No. of Sample Points
High-density residential	94	780
Low-mid density residential	86	730
Deprived type I residential	100	830
Deprived type II residential	100	830
Non-residential (Industrial/commercial/administrative)	100	830
Non-built up	104	865

To predict deprivation, we develop two datasets (Table 10). For the first dataset, we use GLCM textures and spectral features (VIS+NIR) to form the first dataset. Secondly, since our study aims to test the predictability of deprivation using multi-hazards, we combine the multi-hazard dataset with the VIS+NIR to create the second dataset. Then, for both datasets, we select predictor variables using VSURF. The selected predictor variables are in combination with the land cover classification map to perform land use classification.

Table 10: Datasets composition for land use classification (predicting deprivation)

Group	Variables	Dataset 1	Dataset 2
	Red (B1)	x	x
	Green (B2)	x	x
	Blue (B3)	x	x
	NIR (B8)	x	x
Texture features	Contrast	x	
	Entropy	x	
	Mean	x	
	Dissimilarity	x	
	Homogeneity	x	
	Angular Second Moment (ASM)	x	
	Correlation	x	
	Variance	x	
Land cover	Built-up vs non-built	x	x
Multi-hazards	Multi-hazards (Table 6)		x

3.5.2.6. Accuracy Assessment

To assess the model performance, perform accuracy assessments by comparing classified data to reference data. We compute the overall accuracy from the confusion matrix that provides the global accuracy assessment measure based on total correctly classified pixels. Additionally, we compute the F1 score- the weighted average function of precision and recall (Brownlee, 2014). Precision is calculated using the confusion matrix as a ratio of correctly predicted observations to the total predicted positive observations. On the other hand, recall is calculated as the ratio of correctly predicted positive observations to total observations by class.

$$F1\,Score = 2 * (Recall * Precision) \,/\, (Recall + Precision)$$

Equation 5: F1 score function

3.6. Validation of Multi-Hazard Index Using Household Survey

3.6.1. Design and Structure of Questionnaire

To cross-validate the outcome of the multi-hazard index, we conduct household surveys in two deprived settlements in Nairobi: Kibera and Kariobangi North. In developing the questionnaire, we focus on the hazards identified in the multi-hazard index. Furthermore, given that the purpose of the questionnaire is to understand the hazards experienced in deprived settlements and at the household level, we use a funnel approach for the survey design (from settlement to household level). The questions are closed-ended with an allowance for additional commentary by the respondents (annex 8.4). Due to COVID-19 regulations, a local community group (Community Mappers) was contacted to provide research assistant services.

3.6.2. Target Population and Number of Participants

The two settlements were selected following a recommendation by the contracted community group as they represent different types of deprived settlements in Nairobi. Surprisingly, they match the two types of deprived settlements captured in section 3.2. Kariobangi North, though not captured in our provided data of morphological slums, we find that its morphological characteristics match those classified as deprived settlements type II. The availability of research assistants in the two settlements was also a contributing factor in selecting the two settlements. Given time and resource constraints, the total number of target households was 100; 70 households in Kibera and 30 in Kariobangi North. This approach was taken given that Kibera (approx.2.2km^2) is double the size of Kariobangi North (approx.1.1km^2).

3.6.3. Sampling Technique and Data Collection

A random sampling strategy was employed as a measure to reduce bias in the data collection. Specifically, we created grids of 100mx100m over the settlements and using the random selection tool in QGIS, 70 grid cells were selected in Kibera and 30 in Kariobangi North (fig.16). In developing the questions, we focused on both settlement and household levels. The questions comprise basic household information, including household size. The main focus is, however, on the hazards present/ experienced by the respondents. The design of the questionnaire is a combination of closed and open-ended thus, leaving room for capturing additional comments. The questionnaire was designed and deployed using the KoBo toolbox, which was selected due to its compatibility with mobile devices, geo-location collection capabilities, and is a free and open-source application. Also, considering that the enumerators constituted members of different informal settlements within Nairobi, they were consulted to fine-tune the questionnaire. Some of the changes included adding options to our response choice lists and translating the key questions (and response choices) into Kiswahili.

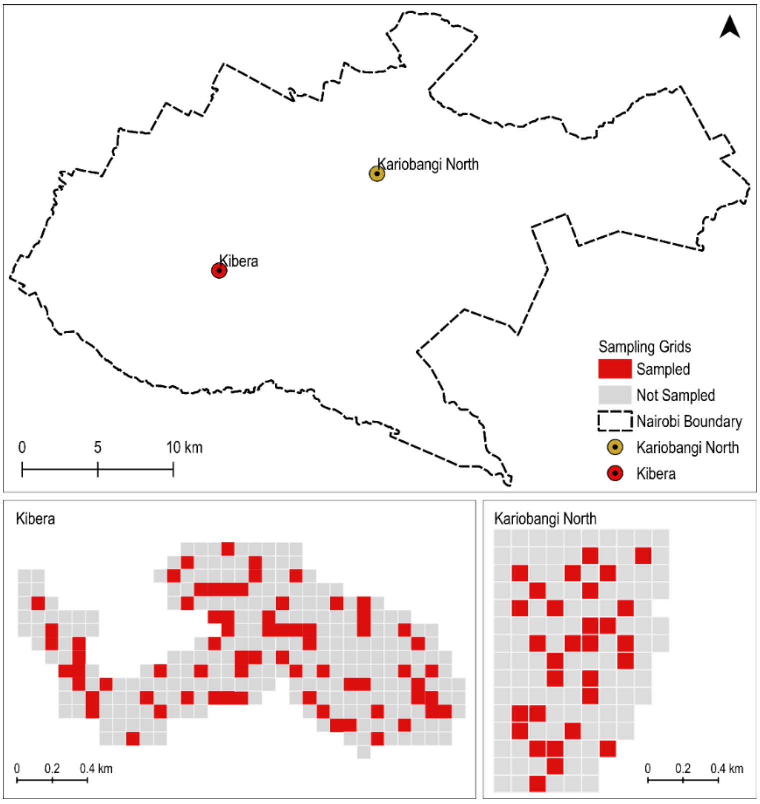


Figure 16: Study area with location of selected settlements for HH survey(top), and randomly selected grids within Kibera (bottom left) and Kariobangi North (bottom right).

3.6.4. Data Analysis

To analyse the household responses, we first encode the responses by stripping off the personal identification details (respondents names), then categorize the data into different topics to answer the research questions. Given the nature of the survey, we have multiple responses, which are first grouped by defining variable sets. All the responses are then analysed through descriptive statistics, which give a summary of the answers. Contingency tables are also produced to summarize the relationship between different variables. The analysis is conducted using SPSS and graphs generated using MS Excel. Also, using the grids from which households were selected for interviews, we compute the hazard scores based on the multi-hazard index. The operations are undertaken in QGIS employing the zonal statistics tool, where the mean values are calculated. For ease of comparison, the results are aggregated and presented per sub-hazard category (Table 6), with the aim of contrasting household survey results to those of the multi-hazard index.

3.7. Limitations of Data and Methods

In designing the multi-hazard index, we accommodate the inherent heterogeneous nature that characterises multiple hazards, which means that we use data generated using different measurements and spatial resolutions. For instance, we use data with coarse resolution (e.g., LST and air pollution data with 1000m resolution) that minimally show the intercity variations. Thus, we upsample these data to 10m resolution - our chosen analysis unit. We use bicubic interpolation techniques that alter the original data values by considering the surrounding pixels to highlight intra-city variations. Transition areas are most affected since interpolation techniques rely on surround cell values to compute new values for the resampled data.

The multi-hazard index also relies on proximity measures, specifically, the computation of euclidean distances. By using such measures, we assume that proximity to a hazard threat indicates high vulnerability. These assumptions simplify the complexity of hazards and are significantly influenced by a lack of data. For instance, in assessing transport accidents, the proximity to rail infrastructure is computed. However, we are aware that despite having rail infrastructure in most city areas, operations are only currently underway in the city's central and eastern regions. Thus only settlements in these areas would technically be threatened by rail accidents. To capture this complexity, we would require rail operations data to select better fitting data.

Additionally, since hazards are analysed as a factor of occurrence and magnitude, our multi-hazard index shows the susceptibility of an area to identified hazards. Primarily, this decision is influenced by insufficient and complete hazard data (occurrence and magnitude) matching our index. Further, our goal is to estimate the relative severity of multi-hazard risk using justifiable spatial proxies. In compiling the index, we acknowledge the interrelation among the different hazards. However, we neither assess the causal interactions among the hazards nor explore causal relationships between the hazards and the presence/conditions of deprived settlements. Nonetheless, we highlight these relationships with evidence from literature and expert interviews.

Further, our attempt to predict deprivation using multi-hazards acknowledges that some data such as building and road characteristics and their derivatives, e.g. density, have been used to inform some conventional textural features used in image classification (Kohli et al., 2012). To mitigate this conflict, we generate two datasets and compare their predictive power. One dataset is based on spectral data with textural features, while the second combines spectral data and multi-hazards.

Lastly, in designing the household questionnaire, the main limitation is the failure to capture all the hazards from the index. A challenge that was encountered since both processes were being conducted concurrently. Additionally, the responses rely on the respondents' perceptions, for instance, in assessing the degree of extreme temperatures. The responses are based on the households experience and sensitivity to weather elements. As we know, these vary from person to person due to different factors, including biological 'make-up'.

4. RESULTS

In this chapter, we present the results of the study. First, we present the outcomes of the expert interviews on hazards affecting Nairobi. These were used to refine our theoretical multi-hazard index that is presented next. As aforementioned, the multi-hazard index is used to contrast the degree of hazardousness between deprived and non-deprived settlements presented in section 3.2. We then present the discriminant analysis results that were used to evaluate whether the multi-hazard index can predict deprivation. In addition to the statistical evaluation, the discriminant analysis was used to identify discriminant variables for each settlement. These results are contrasted to conventionally used datasets for predicting deprivation. Next, we present the household survey outcomes that assessed settlement and household level disbursal of hazards. Lastly, since we discuss different subjects with the experts, these results are infused in the different sections as deemed appropriate and cover the topics on 'slum' definitions, the data used for deprivation mapping and geo-ethics issues related to studying deprivation.

4.1. Multi-Hazard Susceptabillity Index

The multi-hazard index developed in this study covers the entire city of Nairobi and is informed by extensive literature search and expert opinion. The index approximates hotspots indicating areas most susceptible to hazards and the regions of hazard overlaps. To understand intra-city disbursal of hazards, the degree of hazardousness among different residential settlements is contrasted.

4.1.1. City-wide vs Deprived Settlements Hazards in Nairobi

The interviewed experts identified hazards affecting our study area at the city level and deprived settlement level. Two significant hazards identified by all experts were flooding and fire (Table 11). For flooding, distinctions were made between runoff and riverine flooding, as these were more prominent in non-deprived and deprived settlements, respectively. Runoff was attributed to insufficient and blocked drainage channels, while riverine floods were attributed to the settlements encroaching on riparian reserves. However, the experts distinguish that runoff is the primary flooding hazard at the city level while riverine floods are localized in deprived settlements. Fire hazard was specifically considered more prone in deprived settlements. Further, deprived settlements were exposed to more hazards than the rest of the city (Table 11).

Table 11: Frequency count of expert opinion on hazards at city level and deprived settlement level in Nairobi

City-wide	Hazard	Deprived Settlements
7	**Floods**	7
0	**Fire**	7
0	**Crowding**	1
1	**Transport accidents**	1
0	**Industrial accidents**	1
1	**Unfavourable micro-climate**	1
0	**Contamination (air/water/land pollution)**	2
0	**Political Conflict**	1
0	**Building collapse**	1
0	**Garbage accumulation**	1
2	**Infectious diseases**	2

4.1.2. Spatial Analysis of the Multi-Hazard Index

To test the assumption that deprived settlements are located in hazardous areas, we use the identified four types of residential settlements within our study area in combination with the multi-hazard index. Using the weighted values per sub-hazard category, we standardize the values by dividing the values using the maximum for ease of comparing the susceptibility of the settlements. We select an equal number of samples (n=86) spread throughout the city for each settlement class. Lastly, we present the results of the multi-hazard index entailing all identified sub-hazards and, in a similar fashion, compare the inter-settlement performance.

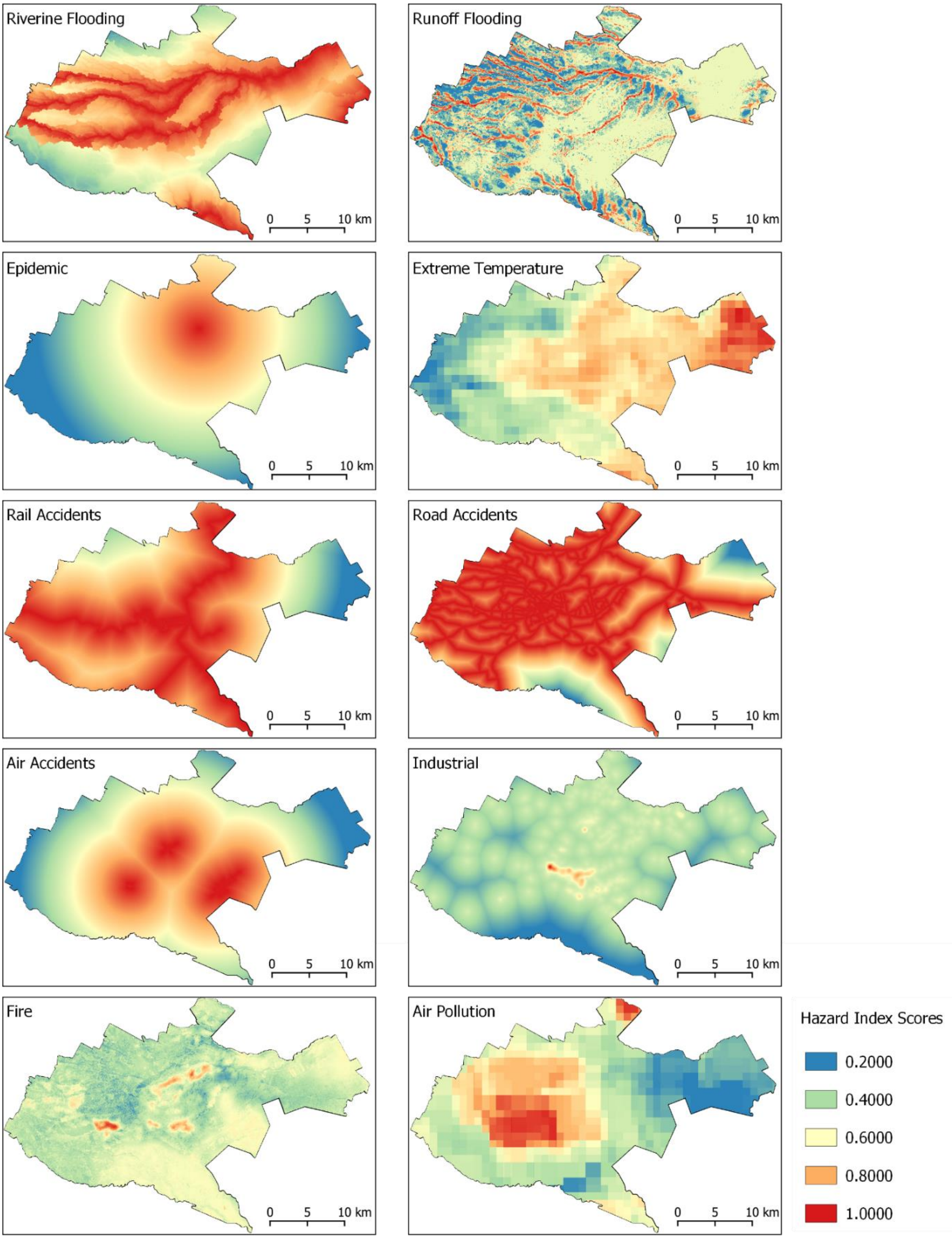


Figure 17: City-wide analysis of degree of hazardousness by the ten sub-hazard categories of the multi-hazard index.

4.1.2.1. Flooding

Notably, most of the city has index scores > 0.6 as explained by two factors. First, the city has a relatively flat terrain. The difference between the highest and lowest point, computed from the DEM, is approx. 700m. Secondly, the city has three major tributaries (Nairobi, Mathare and Ngong rivers) of the Nairobi Drainage basin system cutting through the city. The combination of these two factors makes the city highly susceptible to river inundation. Despite the seemingly high susceptibility to riverine flooding, the experts inform us that the river tributaries do not pose a significant threat to the city because the tributaries don't have a big extend and are in valley terrains. In contrast, the drainage system threatens many deprived settlements since they are located on riparian reserves, as highlighted by the outcomes of our analysis (fig. 17).

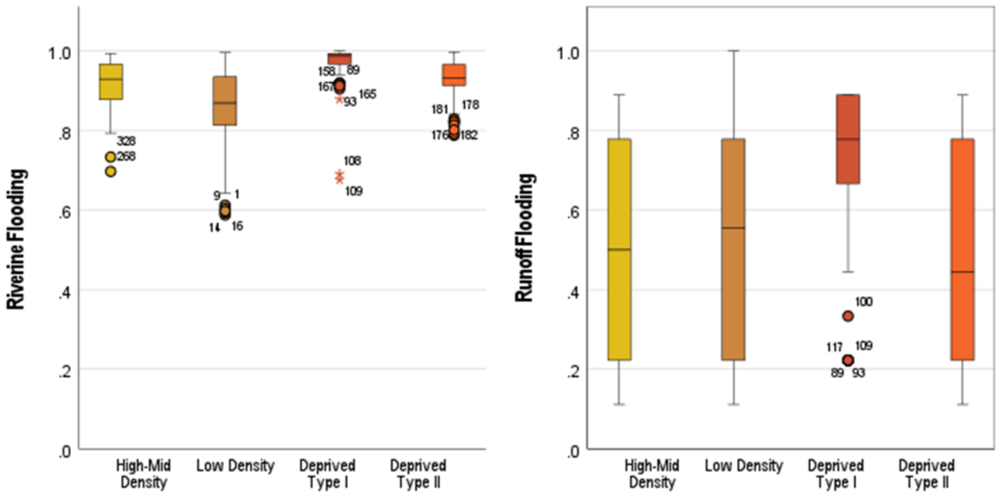


Figure 18: Variability of residential settlements per flooding hazard sub-categories; riverine and runoff flooding.

On the other hand, the experts indicate runoff flooding as a more significant threat at a city-wide scale. We find that the central and eastern regions of the city are at high risk of runoff flooding with index scores of approx.0.5. The terrains of these regions are characterized by foot slopes that transition into flat terrain. Additionally, they lie downstream of the Nairobi drainage system. In addition to terrain form, blockage of drainage systems by garbage and lack of adequate drainage systems are mentioned as causes of runoff floods. Regardless of the anthropogenic causes of runoff, deprived settlements type I are found most susceptible (fig.17b). The other residential settlements show high variability indicating that they are located in different city areas with varying degrees of hazardousness. Both high-mid density settlements and deprived areas type II have a similar trend for both types of flooding. However, high-mid density settlements are more susceptible to runoff flooding than deprived type II settlements. In both instances, low-density settlements are least vulnerable to flooding hazards.

4.1.2.2. Epidemic

Susceptibility to epidemics is based on the proximity of settlements to the city's sole landfill, located in the northeastern region. We find that all the different residential settlements are located near the landfill (fig.18a). However, both deprived settlements type I and high-mid density settlements are most susceptible, given that 75% of the sampled cases are found in areas with scores >0.7 (fig.18). In contrast to low density and deprived type II settlements, these settlements are primarily located in the inner city (fig.9).

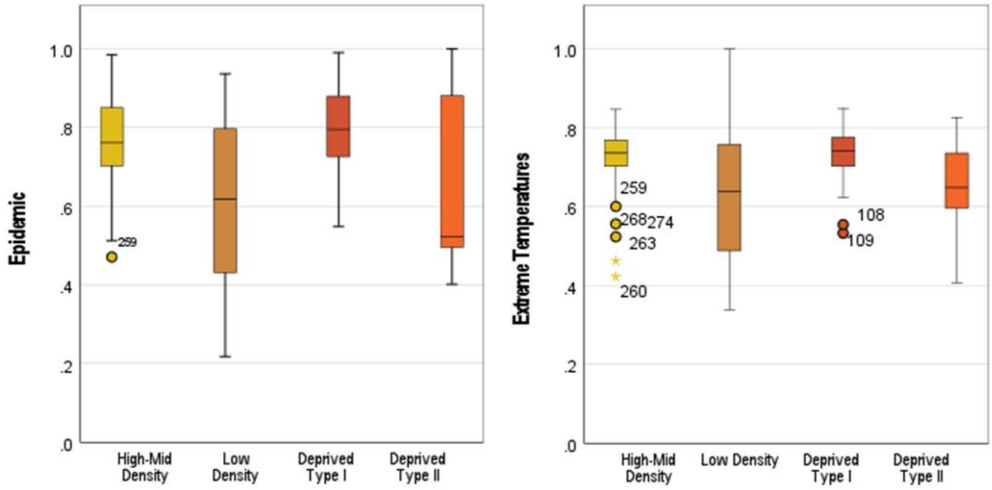


Figure 19: Variability of residential settlements per epidemic and extreme temperature hazards.

4.1.2.3. Extreme Temperatures

Extreme temperatures gradually transition from high temperatures to cool temperatures from the eastern to western regions of the city, with the northwestern areas being the coolest (fig.16). Climatic factors can explain the sharp contrast between the west and east parts of the city. The western regions are closer to the highland areas (Kiambu), while the eastern and southern regions are towards the semi-arid climatic zones (Machakos and Kajiado). Interestingly, we find that high-mid density settlements and deprived type I settlements have similar trend and tend to be located in areas with high temperatures. Low-density settlements have the highest variability, followed by deprived settlements type II since they are mainly located in the periphery of the urban core (fig.9). Meaning that these types of settlements can be found in both the cooler western regions and the hotter eastern areas.

4.1.2.4. Transport Accidents

Nairobi is well served by road, rail and air transport infrastructure (fig.16). Given that we use proximity measures to these infrastructures, most of the city is susceptible to transport accidents. Due to the high connectivity of the road and rail infrastructure, most of the city except the far eastern region is hazardous (fig.19a&b). As a result, we find that all settlements are susceptible to road accidents, having hazardous scores >0.8. Despite high city-wide connectivity, the distribution of low-density settlements is negatively skewed, indicating lower susceptibility to road accidents. On the other hand, high-mid density settlements rank highest, explained by their high road connectivity as depicted by the neighbourhood layout (section 3.2). Deprived settlements (both type I and II) are generally found near rail infrastructure. In terms of air transport accidents, generally, the wider central region of the city is found hazardous (fig.16). We, however, find that deprived type I are the most susceptible, followed by high-mid density settlements (fig.19).

4.1.2.5. Industrial Accidents

The industrial hazards are distributed throughout the city (fig.16). As a result, the risk of industrial hazards isn't as high compared to the other assessed hazards. The presence, however, of an industrial area in the city's core, as zoned in the 1948 master plan (fig. 6), is captured in our analysis as a highly hazardous hotspot area. Similarly, the assessment of the residential settlements reflects this (fig.19). Despite the low variation among the settlements, deprived settlements type I are most susceptible given their positively skewed distribution, including outliers. Additionally, despite having the highest variance, over 50% of the sampled deprived type I cases are in areas with hazard scores >0.6, contrasting the other settlements with approximately 75% of the cases in areas with scores <0.6. High-mid density settlements rank second most hazardous due to the location of these two kinds of settlements in the inner city, where the city's industrial area is located.

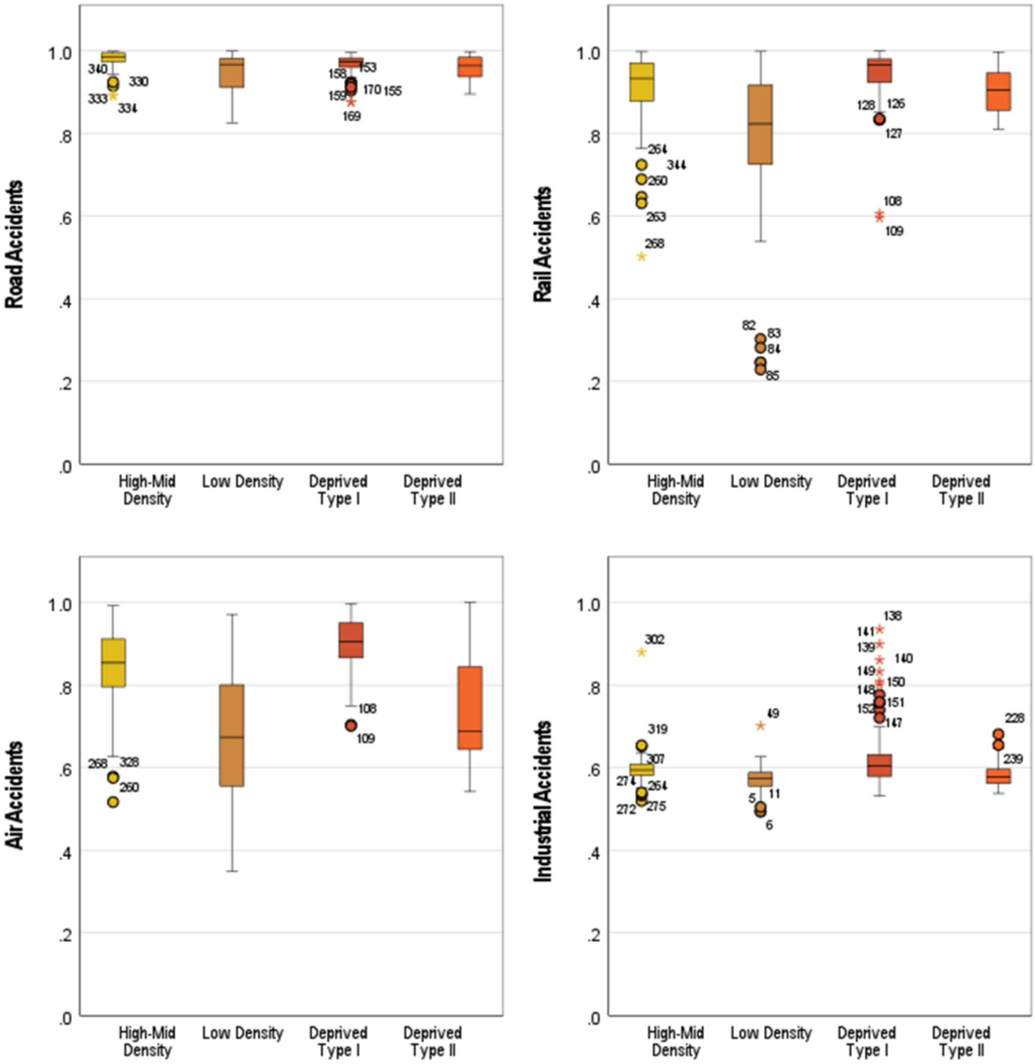


Figure 20: Variability of residential settlements per road, rail, air transport and industrial accidents.

4.1.2.6. Biophysical Hazards

I. Fire

Distinct hotspot areas highlight fires hazards in the city (fig.16). Interestingly, the hotspot patterns distinctively outline some commonly known deprived settlements such as Kibera, classified as deprived type I settlement (fig.15). Unsurprisingly, the comparison between settlements reveals that deprived areas type I are most susceptible to fire hazards, followed by deprived areas type II (fig.20). These results match the outcomes from the expert interviews (Table 11). Attribution is made to the district characteristics of deprived settlements, i.e. densely built, lacking adequate road infrastructure and green spaces (all features used to describe the susceptibility to fire hazards).

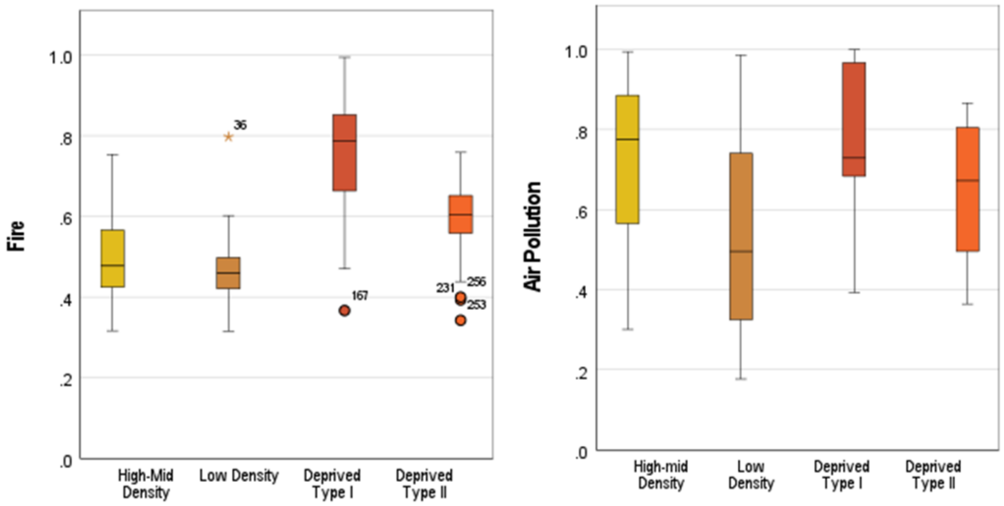


Figure 21: Variability of residential settlements per biophysical hazards (fire and air pollution hazards).

II. Air Pollution

The urban core is the most affected by air pollution, spreading to the western, northern and southern regions (fig.16). The possible reason for this is that by looking at the annual wind directions, the most predominant winds blow from the northeast direction and hardly any from the west (Windfinder, 2021). In the inter-settlement comparison (fig.20), all settlements are located in areas with hazard scores of approx. 0.44. and over 75% of all settlement types, with the exemption of low-density settlements, are located in areas with hazard scores (>0.7). However, high-mid density settlements are most susceptible to air pollution hazards, followed by deprived settlements type I. Additionally, we find that despite the high variance, low-density settlements have a positive skew, indicating that most of these settlements are located in hazardous areas.

4.1.2.7. City-wide Hazard Susceptibility

To compute the overall hazard index, we sum the weighted indicators. As a result, we find that the urban core of Nairobi is the most hazardous while the western region is the least. Comparing the settlements' scores, deprived type I are the most hazardous, followed by high-mid density settlements; their location partly explains since they are mostly found in the inner city, whereas deprived type II and low density are located further from the urban core.

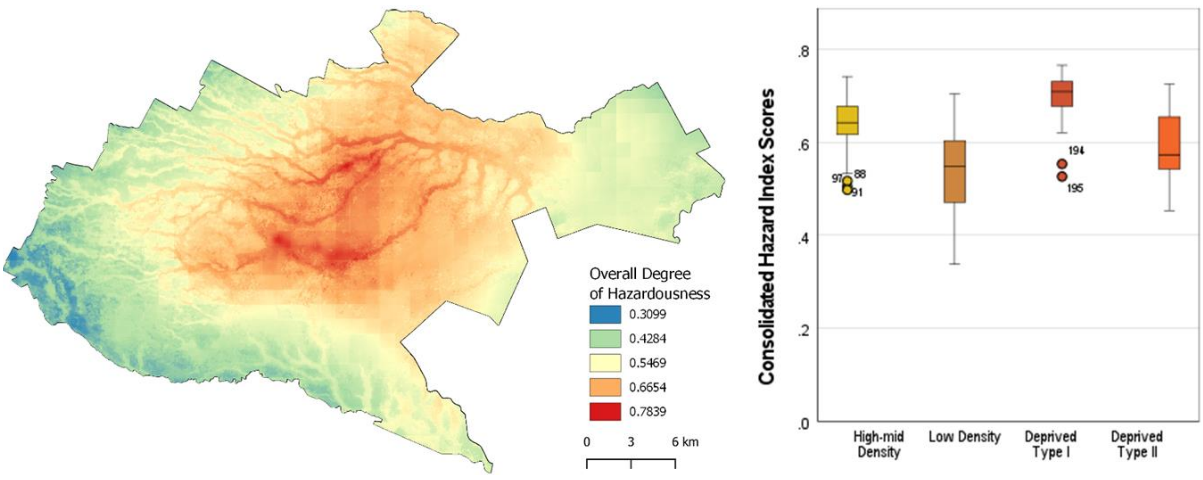


Figure 22: Spatial distribution of hazards in Nairobi. Categories indicate the degree of hazardousness computed from the summation of weighted hazard indicators.

4.2. Predicting Deprivation Using MultiHazards

4.2.1. Discriminant Analysis

By first running MANOVA, we find that our model based on the multi-hazard dataset and four types of residential settlements (section 3.2.4) is statistically significant. We establish this by evaluating all four multivariate test statistics (Pillai's Trace, Wilk's Lambda, Hotelling's Trace and Roy's Largest Root) that obtain $p<0.0001$, indicating that the classes differ significantly (Table 12). Additionally, a Wilk's lambda of $V=0.062$ indicates a high ability of class separation within our data.

Table 12: Multivariate analysis of variance for multi-hazards dataset

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	1.000	69404.007[b]	18.000	359.000	.000
	Wilks' Lambda	.000	69404.007[b]	18.000	359.000	.000
	Hotelling's Trace	3479.867	69404.007[b]	18.000	359.000	.000
	Roy's Largest Root	3479.867	69404.007[b]	18.000	359.000	.000
Classes	Pillai's Trace	1.701	26.267	54.000	1083.000	.000
	Wilks' Lambda	.062	30.647	54.000	1070.495	.000
	Hotelling's Trace	5.474	36.258	54.000	1073.000	.000
	Roy's Largest Root	3.789	75.985[c]	18.000	361.000	.000

Next, we carried out a discriminant function analysis to evaluate the performance of outcome variables in discriminating the classes. Given that we have four groups, three statistically significant discriminant functions were obtained (Table 13). The first discriminant function tests the model as a whole, explaining 69.2% of the variance, with canonical $R^2= 0.79$; the second function explained only 17.6%, with canonical $R^2=0.49$, and the third 13.2% with canonical $R^2=0.42$.

Table 13: Eigenvalues of the discriminant functions

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3.789[a]	69.2	69.2	.889
2	.964[a]	17.6	86.8	.701
3	.722[a]	13.2	100.0	.647

To understand the discriminant functions, we explore the model's canonical structure matrix (fig.23). The structure matrix indicates variables with high contribution to class separation. The higher the canonical variate correlations value (R), the higher the contribution of outcome variables to group separation

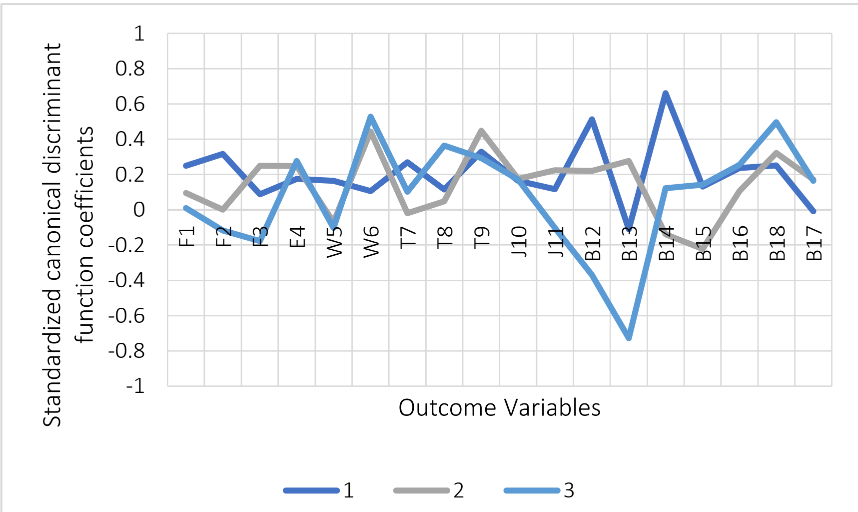


Figure 23: Relative contribution of variables to the discriminant functions

(Bargman, 1970, as cited in Field, 2017). Typically, variables with R>3 are considered to have a significant effect (Field, 2017).

In addition to the canonical structure matrix (fig.23), the discriminant analysis gives the functions at group centroids (Table 14). These show the functions used for discriminating the classes. The group centroids represent the group/class mean function scores (Field, 2017). The class with an opposite sign indicates that the function discriminates that specific class (Field, 2017). Therefore, the first function differentiated low-density areas (mean = -3.29); the second function differentiated type II deprived areas (mean = -1.515); and the third differentiated high to mid-density built areas (mean = 1.473). Also, by observing the differences in mean values for the classes, we understand the relationship among the classes. We find that deprived areas - type I have positive high mean scores in both the first and second functions (Table 14). Thus, highlighting the high contrast between deprivation type I and the classes discriminated by these functions, i.e. low-density built-up areas and deprived areas type I (Table 14). Similarly, deprivation type II also has high values (mean = 0.705), though not as high as deprivation I. Additionally, these observations can be observed from the plotted distribution of the class samples (including the group centroids) against the functions (fig.24).

Table 14: Functions at group centroids

	Function		
Classes	1	2	3
High-Mid Density	.019	.070	1.473
Low Density	-3.290	.428	-.492
Deprived Type I	2.107	1.081	-.528
Deprived Type II	.705	-1.515	-.433

Using the observations from the canonical structure matrix and group centroid functions, we can further understand the discriminating factors (variables) between classes. Importantly, we note that positive relationships show the functions' discriminating classes based on dimensions that similarly affect the variables, whereas negative ones show the inverse (Field, 2017). Looking at the outcome variables that are significant in the first function (fig.23), we deduce that NDVI (B14) (R=0.661), the density of buildings (B12) (R=0.512), proximity to airports (T9) (R=0.329) and proximity to rivers (F2) (R=0.317) are significant discriminators of low density built areas from the rest of the classes, and especially type I deprivation (further captured by the significant distance along the horizontal axis)) (fig.24). In terms of class separability, however, the low-density areas group centroid sign is negative, highlighting it has an opposite relationship to the variables.

Considering NDVI and building density as examples, and taking into account that the variable values were normalized (section 3.4.2); where typically high NDVI values signify high vegetation cover, the NDVI values were inversely transformed. Therefore, areas with low vegetation cover have higher values to indicate higher hazard susceptibility. In contrast, building density values remained unaffected through data normalization and, areas with high building density have higher values and represent high hazard susceptibility. The canonical values for these variables are positive (fig.23), whereas the group function for low-density settlements is negative (Table 14). Hence, low-density settlements have higher vegetation cover and low building densities. A sharp contrast is observed for deprivation type I, with a high group centroid value with a positive sign (mean = 2.107), indicating lower vegetation cover and high building densities.

The second function discriminates type II deprivation from the other classes and especially from type I deprivation (Table 14) (also see fig. 24 vertical axis distance). The significant variables within the function were found to be proximity to airports (T9) (R=0.448), night LST (W6) (R=0.444), and CO (B18) (R=0.323). The variables were also found to have a positive relationship with the discriminant function, while the type II deprivation group centroid was negative (mean = -1.515). Therefore, we find type II deprivation settlements farthest from airports and have low night LST and CO exposure of all the settlements. However, the inverse applies for deprivation type I having a significant positive group mean value of 1.081 compared to deprivation type II.

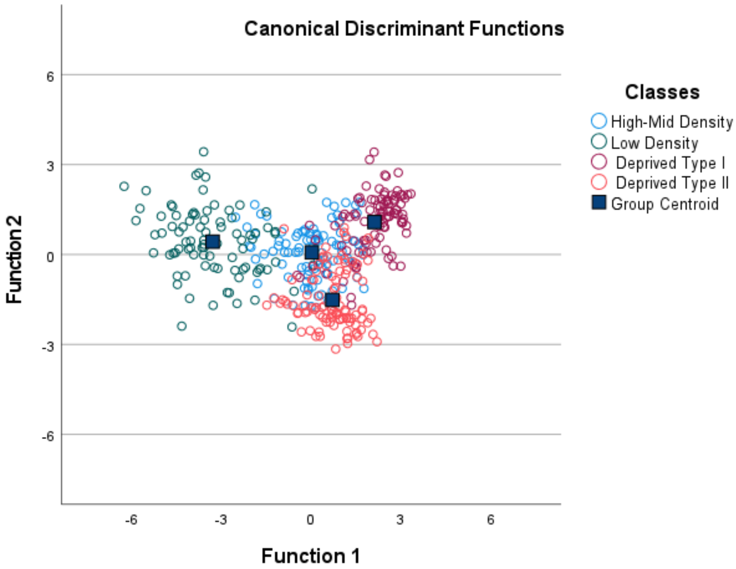


Figure 24: Plotted samples and respective group centroids against the first and second canonical discriminant functions

The third function discriminates high-mid density built-up areas with a positive mean of 1.473 (Table 14) and, using the variables road density (B13) (R= -0.728), night LST (W6) (R=0.528), CO (B18) (R=0.497), building density (B12) (R= -0.370) and proximity to major roads (T8) (R=0.364) (fig.23). For the variables night LST, CO and proximity to major roads, the canonical values are positive, as is the case for the function at group centroid for high-mid density settlements. These positive values indicate that high-mid density settlements have high night LST, CO, and proximity to major roads. On the other hand, road density (transformed similarly to NDVI) and building density have negative values. Thus, high-mid density settlements have high road and building density. However, these variables can discriminate other classes better despite their significance in discriminating high-mid density settlements. And as seen from the first and second functions, building density and road density contribute significantly to the discrimination of deprivation type I.

In summary, the first function demonstrated that deprived type I settlements are most susceptible to fire hazards (building density and NDVI) and riverine flooding. Both the first and second functions further explain that deprivation type I is susceptible to air transport accidents. On the other hand, the second and third functions show both deprivation type I and high-mid density settlements experience high night LST and CO exposure. However, the canonical values of the third function highlight high-mid density settlements to be more susceptible to these hazards. The high nigh time LST values can be explained by the highly built nature of these settlements (J. Wang et al., 2019). High CO values can be attributed to fossil fuel combustion due to road traffic related to the high road density (WHO, 2016). Further, the third function shows that high-mid density settlements are most prone to road transport accidents. Therefore, our assessment finds deprivation type I more susceptible to hazards, followed by high-mid density settlements than the other settlements. These findings are similar to those of the multi-hazard index and expert interviews. Lastly, by testing classification performance, the discriminant functions yield overall classification accuracy of 79.7%.

4.2.2. Random Forest Classifier

4.2.2.1. Land Cover Classification

We undertake supervised multi-class prediction. Similar to the study by (Saini et al., 2021), we first use GLCM textural features of kernel size 3X3 in combination with VIS+NIR, and NDVI bands, resulting in a total of 37 features (Dataset1). The features were then selected using the VSURF algorithm, where ten predictor features are obtained. These variables are used in RF with the parameters set at *ntrees* = 5000 and *mtry* at default. The overall accuracy of 68.4% at a 95% confidence level is obtained. We also conduct a second test using all GLCM features (416 bands) combined with VIS+NIR and NDVI bands totalling 421 (Dataset2). Following a similar approach, seven predictor variables are selected using VSURF, and an accuracy of 70% at a 95% confidence level is obtained. The kappa coefficients of both tests only differed by 0.1, indicating moderate agreement that can be expected at random chance. The results summary are in table 15. Also, an experiment of binary classification, i.e., built vs non-built, achieved a higher overall accuracy of over 80% using both datasets (Dataset1 and Dataset2). From the first level, therefore, the errors are propagated into the subclasses.

Table 15: A summary of land cover classification using RF model with ntree=5000 and mtry=√number of variables.

Test	VSURF Computation Time	Overall Processing Time	Total No. of Variables	No. of Pred. Variables	OA	Kappa
Test 1	0.2 hrs	1.2 hrs	36	10	68.4%	0.52
Test 2	6.6 hrs	7.7 hrs	421	7	70%	0.53

From the tests, the use of only 3X3 kernel textures results in noisier land cover classification results (fig.25&26). Through visual inspection, both datasets overestimate bare land by confusing the areas to built-up. On the contrary, the second test requires significantly higher processing time despite obtaining higher accuracy and less noisy classification maps (fig.25&26) - additionally, the second dataset yields a higher overall accuracy and kappa coefficient value. Therefore, we consider the second dataset for classification and test the parameters by tunning the *ntree* and *mtry* values. We find that the combination of *ntree* =2500 and *mtry* =2 yields a classification accuracy of 73% at a 95% confidence level and a kappa coefficient value of 0.59

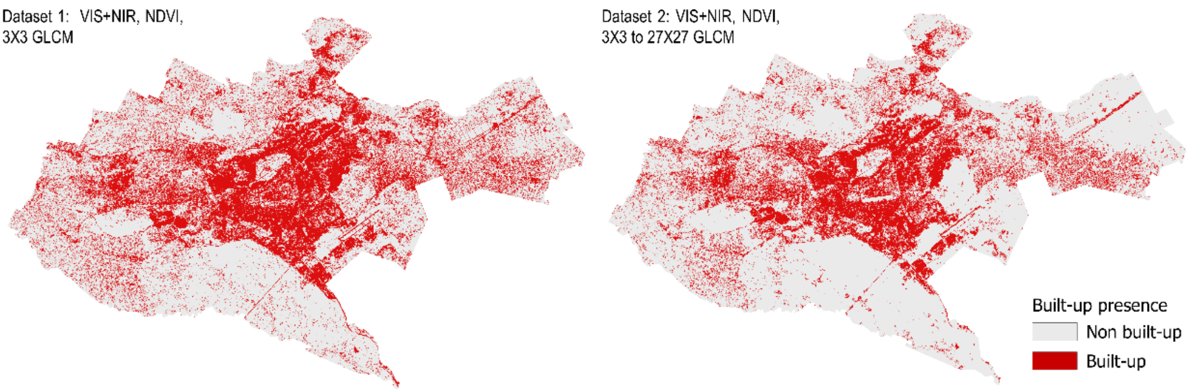


Figure 25: Land cover classification comparison based on two datasets.

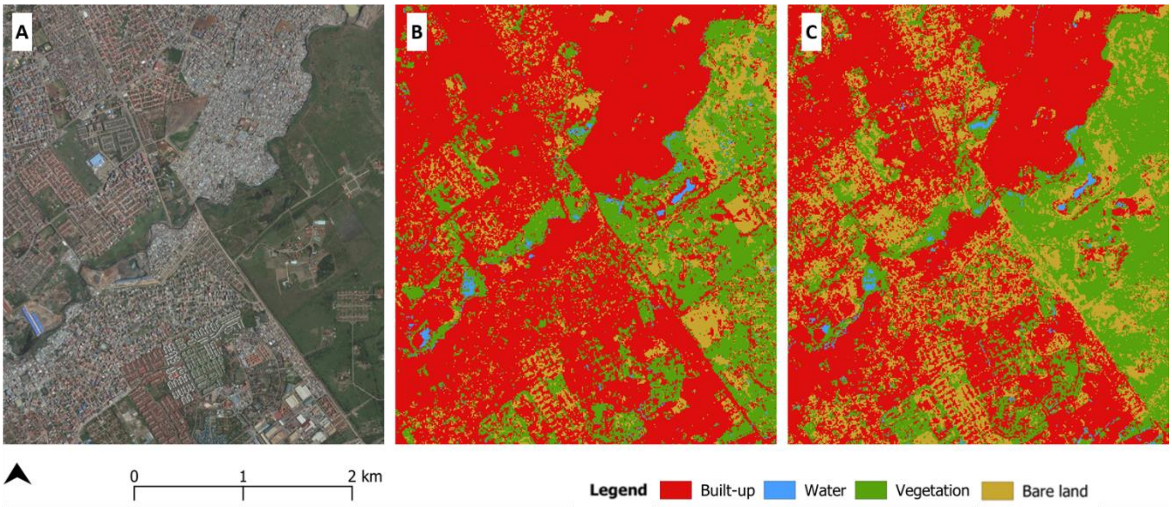


Figure 26: Subset region's land cover classification results: a) reference image and b) dataset one and c) dataset two classifications.

Additionally, our presented outcomes indicate that a kernel size of 3X3 performs well for extracting textural features for land cover classification using HR imagery (10m resolution) and; of the eight textural feature types, variance, entropy, and mean are significant since they are selected using VSURF (fig.27). These outcomes are, however, inconclusive since, through several tests, we make a notable observation that descriptive statistics texture measures (variance and mean) and orderly measures (i.e., entropy) are generally significant for classification when generated by smaller kernel sizes 3X3 to approx.9X9. In contrast, textural measures (homogeneity, dissimilarity and contrast) generated by larger kernel sizes, e.g. 15X15, are significant for land cover classification. These findings can be explained by the type of texture measure and the possible amount of texture information generated per region defined by the kernel size in relation to the image resolution (Haralick et al., 1973).

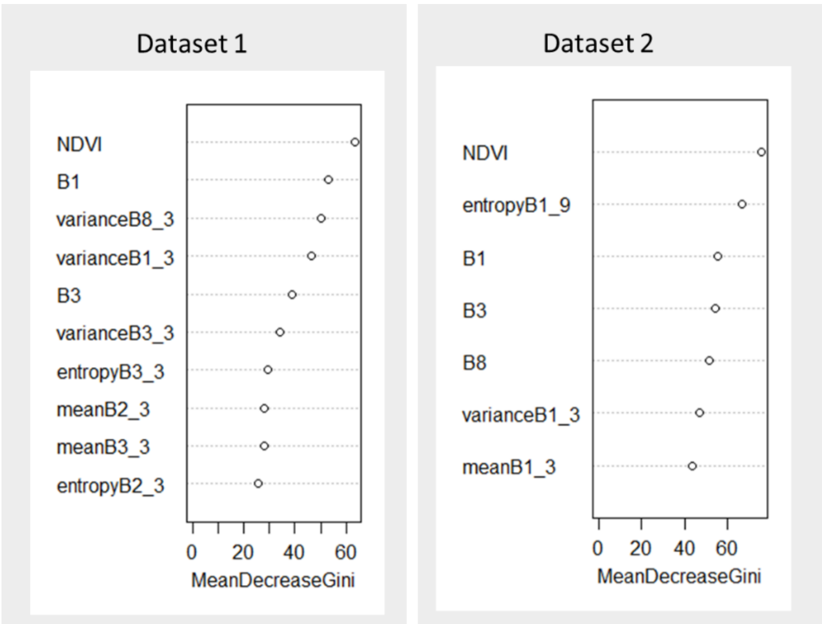


Figure 27: Random Forest generated ranking of variable imporance.

Moreover, there are fluctuations in the selection for predictions of textures generated using different spectral bands using the VSURF algorithm. High correlations among the features can explain the variations observed in various iterations (Haralick et al., 1973). Therefore, no explicit contrasts are observed that distinguish kernel sizes or spectral bands of textures for land cover classification using HR imagery. In all cases, however, NDVI is selected and ranked as a significant variable. Despite having high significance for image classification, spectral data yields better results in combination with additional features.

4.2.2.2. Land Use Classification: Deprivation Prediction

Following the land cover classification, we conduct land use classification. One of the aims of our study is to test the ability of multi-hazards to predict deprivation. A comparison between textural features and multi-hazards, when combined with land cover data, is made. We use the VSURF algorithm to select the best features for predicting deprivation. 35 features of 420 are obtained from the texture-based dataset, while only seven are selected from the multi-hazard dataset. Both datasets obtain overall accuracy of above 70% at 95% confidence (Table 16). The multi-hazard dataset, however, performs slightly better by 2% OA. These are results obtained with the tuned parameters of *ntree*=3000 and *mtry*=2 for the multi-hazard dataset and *ntree*= 2500 and *mtry*=6 for the texture dataset.

Table 16: Land use classification model summary.

Datasets	Total No. of Variables	No. of Pred. Variables	OA	Kappa
Multi-hazards + Spectral Features	22	7	74%	0.69
Textural + Spectral Features	420	35	72%	0.66

For the multi-hazard datasets, the confusion matrix shows the confusion of class prediction between low-density settlements and non-built areas. The lower recall values (0.58) of low-density settlements followed by non-built (0.63) reflect this (Table 17). Also, low-density settlements have the lowest F1 score (53%), indicating that they were under classified. We find this to occur due to the high vegetation cover that lew-density settlements have similar to non-built areas in many parts of the city. Additionally, the data we use has a resolution of 10m; thus, some houses in low-density neighbourhoods may be undetected.

As for the textures dataset, a similar observation is made where the low-density settlements are confused with non-built areas. Also, we find high-mid density settlements to be confused with both types of deprived settlements. As a result, we see that high-mid density settlements have the lowest recall (0.66) (Table 17). However, the F1 scores show that deprived type II areas have the lowest value of 63%, indicating the model's underperformance in classifying these areas and classifying high-mid density settlements (Table 17).

Table 17: A comparison of precision, recall and F1 score per class for multi-hazard and texture-based datasets

	Multi-Hazard + LCC			Textures + LCC		
	Precision	Recall	F1	Precision	Recall	F1
High-Mid Density	0.73	0.64	0.68	0.74	0.66	0.71
Low Density	0.50	0.58	0.53	0.53	0.83	0.65
Non-Residential	0.99	0.89	0.94	0.63	0.73	0.68
Deprivation Type I	0.85	0.83	0.84	0.84	0.73	0.78
Deprivation Type II	0.68	0.87	0.76	0.59	0.69	0.63
Non-Built	0.71	0.63	0.67	0.93	0.67	0.68

Despite having a high OA, we see from the visual interpretation of the classification that the multi-hazards dataset generalizes entire regions and appears to overfit the model. On the other hand, the results of the textures though noisy, capture a more realistic scenario (fig.28).

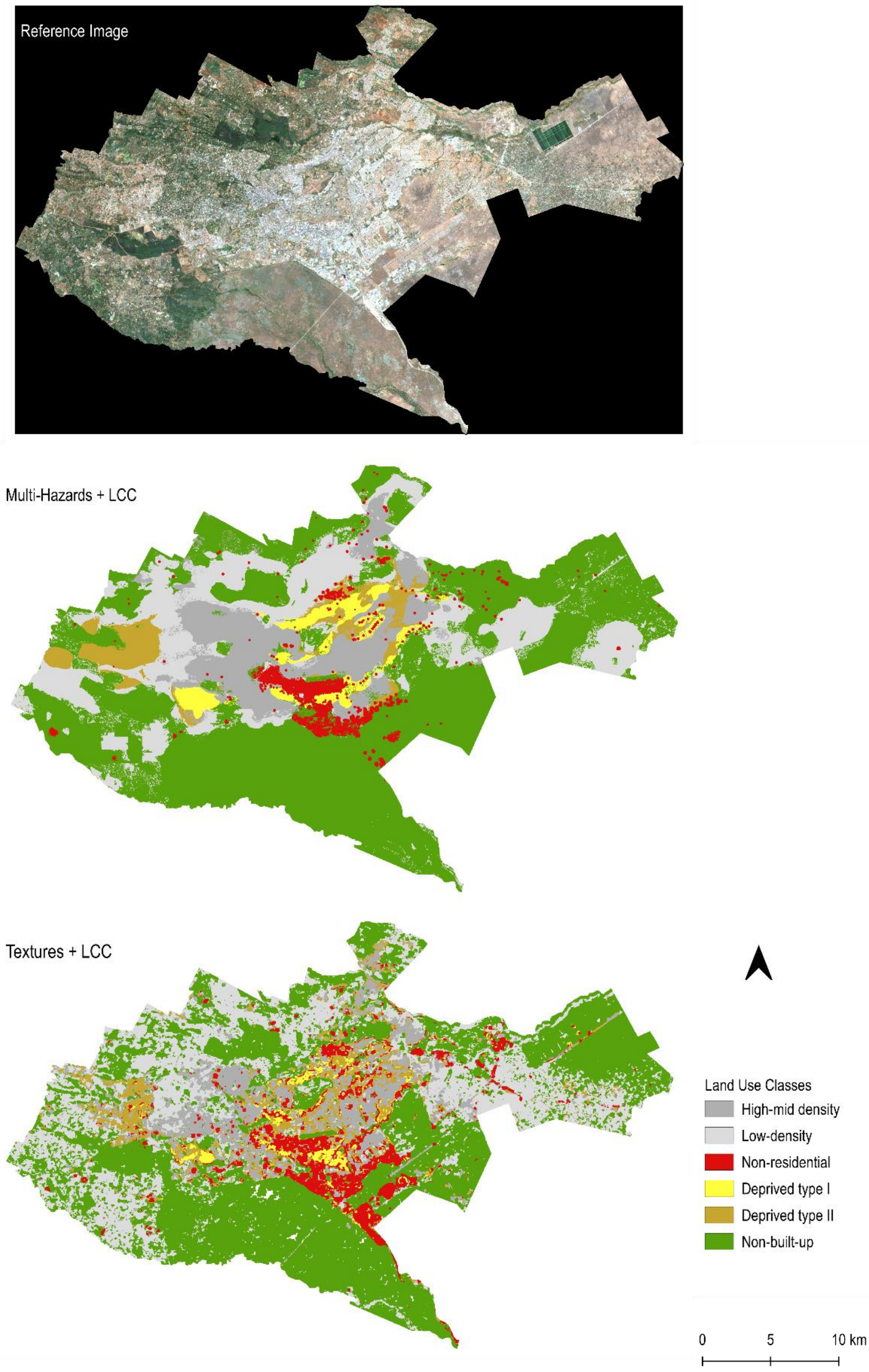


Figure 28: Reference image, and land use maps generated from multi-hazard dataset and texture features (top-bottom).

Lastly, the investigation of variables used for mapping deprivation from the multi-hazard dataset reveals that proximity to industries, building density and proximity to rivers are the leading features for predicting deprivation (Table 18). When using texture features, correlation, ASM, and contrast are the most significant. Notably, most of the texture features with high importance are generated by large kernels (Table 18). This is possibly due to the need to increase the scope of focus by the model, given the highly heterogenous nature of urban land use.

Table 18: Variables with leading feature importance as evaluated by Randon Forest (mean decrease accuracy)

Multi-Hazards + LCC		**Textures + LCC**	
Features	**Var. Importance**	**Features**	**Var. Importance**
Industry_Proximity	263.94	correlationB8_27	95.86839
Building_Density	194.67	second_momentB8_11	88.13104
River_Proximity	182.87	contrastB3_27	85.8445
Road_Density	162.55	correlationB8_23	84.54208
Rail_Proximity	155.49	contrastB8_27	82.7192
Dumpsite_Proximity	132.82	homogeneityB8_27	82.27968
Nighttime_LST	121.72	contrastB2_9	81.71509

4.2.3. Discriminant Functions and VSURF Algorithm Hazard Predictors

First, by looking at the outcomes of the discriminant analysis and VSURF procedure undertaken as part of the land use classification process, significant hazard indicators (variables) for differentiating residential settlements are identified (Table 19). This assessment highlights the advantages of discriminant analysis in determining the variables used for differentiating specific classes. Both analyses find proximity to rivers, night LST, building density, and road density significant. These variables are essential for distinguishing settlements and highlight the hazards that the settlements are exposed. While nearness to rivers highlights susceptibility to riverine flooding, high building density and low road density – indicate fire susceptibility; night LST represents modified micro-climates resulting from urban land transformations (section 3.4.3).

In addition to identifying the more prone hazards in settlements, our use of hazard proxies that characterize settlements enables us to contrast the morphological characteristics of deprived and non-deprived settlements empirically. For instance, theoretically, deprived settlements are said to be located in flood-prone areas. From our discriminant analysis, we find that distance to rivers significantly discriminates deprived settlements. This is also the case for high building density and low road density measures. High night-time LST (J. Wang et al., 2019) and high road density discriminate high-density settlements; thus, we infer that these areas are highly built-up.

Table 19: A comparison of the significant hazard indicators selected by the canonical discriminant functions and VSURF algorithm.

Sub-Hazards	Hazard indicators	Discriminant Analysis	Random Forest (VSURF)
Riverine Flooding	Height Above Nearest Drainage (H.A.N.D)		
	Proximity to Rivers	×	×
Runoff	Geomorphons		
Epidemic	Proximity to Garbage dumpsites		×
Extreme Temperatures	Day Land Surface Temperature (LST)		
	Night Land Surface Temperature (LST)	×	×
Transport Accidents	Proximity to Railway lines		×
	Proximity to Major roads	×	
	Proximity to Airports	×	
Industrial Accidents	Proximity to Industries		×
	Density of industries		
Fire	Density of buildings	×	×
	Road density	×	×
	NDVI	×	
Air Pollution	Sulphur Dioxide (SO_2)		
	Nitrogen Dioxide (NO_2)		
	Ozone (O_3)		
	Carbon Monoxide (CO)	×	

4.3. Inter and Intra-Settlement Disbursal Of Hazards

We investigate the household survey results conducted in Kibera and Kariobangi North to understand the hazards faced at the settlement and household level. Analysis of the performance of the two settlements based on the multi-hazard index is also undertaken to compare results to the settlement level hazards as reported by households.

4.3.1. Settlement Level Assessment

The highest reported hazard is garbage accumulation in both settlements, at 26% in Kariobangi North and 19% in Kibera at the settlement level (fig.30). Garbage accumulation in both settlements can be attributed to the city's lack of adequate garbage disposal services. For many years, Nairobi has relied on the Dandora landfill, which was declared full 25years ago (UNEP, 2018) and only recently was sanctioned for closure through a court ruling (Kiplagat, 2021).

Figure 29: Garbage accumulation in the Nairobi River in Kibera, 2019.

In Kariobangi North, disease outbreaks (22%) is the second most reported hazard, followed by air pollution (12%) and fire (11%) (fig.30). The three leading causes of disease outbreaks are attributed to inadequate water drainage systems (23%), poor environmental conditions (20%) and burst sewerage pipes (20%). Air pollution in Kariobangi North is reportedly caused by burning garbage (44%) and industries (35%), while fire is mainly caused by poor electricity connections (53%) and industrial accidents (24%). In Kibera, air pollution (14%) and fire (14%) are the second-highest reported hazards, followed by floods (10%) and disease outbreaks (10%) (). Similar to Kariobangi North, burning garbage is the highest cause of air pollution, and poor power connections (60%) is the leading cause of fires in Kibera. The reported causes of flooding are blocked drainage channels (33%), insufficient drainage channels (33%), and proximity to river channels (31%). Disease outbreaks, on the other hand, are linked to several factors, including the reliance on unprotected toilets (19%), bursting of sewerage pipes (18%), poor water drainage systems and poor sanitation and hygiene by households (17%).

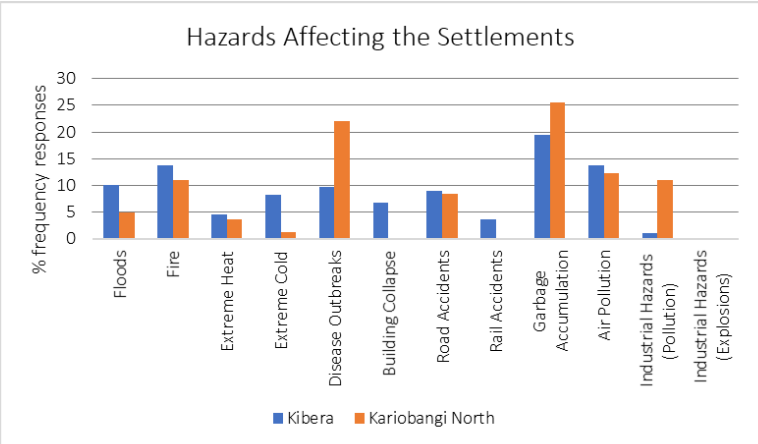


Figure 30: Comparison of hazards affecting Kibera and Kariobangi North

We also explore the two settlements' degree of hazardousness as captured by the multi-hazard (fig.31). Generally, both settlements have a degree of hazardousness higher than the city's average and, when contrasted, Kariobangi North has a higher overall degree of hazardousness than Kibera. Specifically, we find that all sub-hazards afflict Kariobangi North at a higher degree than the city's average with the exemption of runoff flooding and industrial accidents.

To compare the settlement level highly reported hazards to the multi-hazard index, we find that Kariobangi North faces a high threat of epidemics at a significantly high degree compared to the city's average and Kibera. Similar findings are reported at the settlement level. While the index outcome is due to the settlement's proximity to the city's landfill, the challenges are more localised at the settlement level; the causes are linked to the high accumulation of garbage and infrastructure failures. Further, looking at Kibera, the second-highest reported hazard from the household survey is fire. The index reflects this, showing that the scores are significantly higher than the city's average and Kariobangi North. On the other hand, air pollution was reported among the top-three hazard threats in both settlements. We find both settlements to have higher than average scores from the index, with Kibera's being significantly higher (>0.9). Similarly, the household reporting of air pollution hazards was higher in Kibera at 14% to 12% in Kariobangi North.

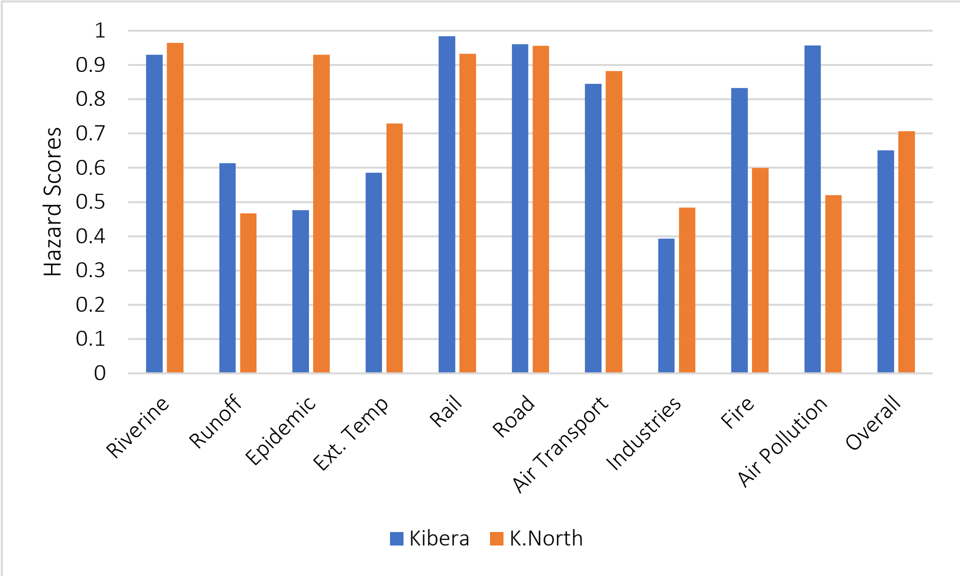


Figure 31: Multi-hazard index scores of Kibera and Kariobangi North

Furthermore, both settlements show a high degree of hazardousness for riverine flooding and transport accidents from the index. On the contrary, these hazards were not highly reported by the household survey. Nonetheless, we found the main causes of flooding in the settlements are blocked drainage channels at 57% in Kariobangi North and 33% in Kibera. Proximity to rivers (29%) and proximity to sewerage plant inlets (14%) was also reported in Kariobangi North, and insufficient drainage channels (33%) and proximity to rivers (31%) in Kibera. Interestingly, however, these findings reflect the experts' opinion whereby both runoff flooding due to blocked drainage channels and inadequate channels was reported and regarded as a city-wide threat, including in deprived settlements. However, both expert opinion and the index reflect that riverine flooding is a more prominent threat than runoff. These findings are contrary to household interviews, where runoff flooding (as implied by the causes of flooding) poses a higher threat in deprived settlements. The low extends of river inundation by the drainage channels in Nairobi can explain these findings. Additionally, we find that Kariobangi North is also affected by point source runoff due to its high proximity to the city's sewerage treatment plant.

Additionally, road and rail transport accidents are high in both settlements, as captured by the multi-hazard index. However, these findings are contrary to the household survey since only 9% reported road accidents in both settlements. Additionally, rail accidents (4%) were only reported in Kibera. Despite this, it is worth noting that the index captures Kibera as having high susceptibility to rail accidents compared to Kariobangi North. These differences can be related to the scale of analysis. The multi-hazard index is a city-wide assessment, thus when contrasted to other types of settlements, Kariobangi North is closer to rail infrastructure. However, making an inter-settlement comparison reveals that Kibera has higher rail accidents scores from the index than Kariobangi North. The threat to rail accidents can be related to operational railway lines cutting through Kibera and not Kariobangi North. This is is reflected in the reported causes of rail accidents, i.e. proximity to the railway lines (56%) and insufficient/lack of pedestrian crossings (44%) by Kibera residents. Similar to rail accidents, the causes of road accidents are attributed to close distance to transport lines (43%) and insufficient/lack of pedestrian crossings (43%) in Kariobangi North. 14% also reported that poorly trained/untrained motorcycle riders cause road accidents. In Kibera, the reported causes of road accidents are mainly related to insufficient pedestrian crossings (44%), proximity to roads (42%) and inadequate road networks (14%). We were not able to capture the air transport hazard threats using the household survey.

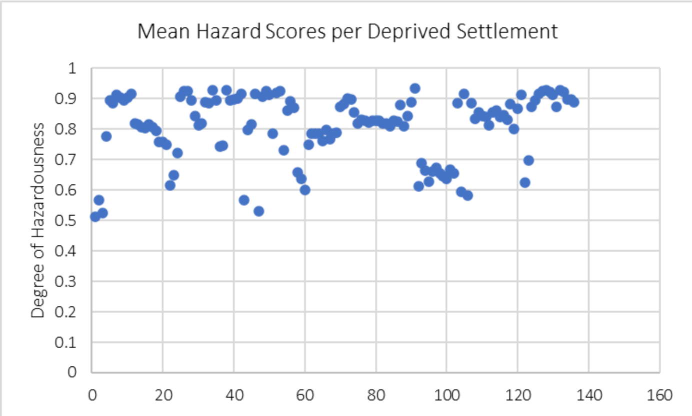


Figure 32: Distribution of deprived settlements by degree of hazardousness

Lastly, based on the morphological deprivation dataset (having 136 deprived settlement boundaries); computation of the degree of hazardousness based on the overall multi-hazard index reveals that all morphologically deprived settlements are located in areas with hazard scores >0.5, with the majority having scores >0.8 (fig.32).

4.3.2. Household Level Assessment

Following a similar fashion, garbage accumulation is the highest reported hazard at the household level in both Kibera (32%) and Kariobangi North (35%) (fig.33). In Kariobangi North, this is followed by disease outbreaks (26%), fire (12%) and industrial pollution (12%), following a similar trend with the settlement level reported hazards. In Kibera, extreme cold (21%) is the second-highest hazard, followed by fire (11%) and extreme heat (10%). Additionally, building collapse is also only reported in Kibera at a low percentage (7%).

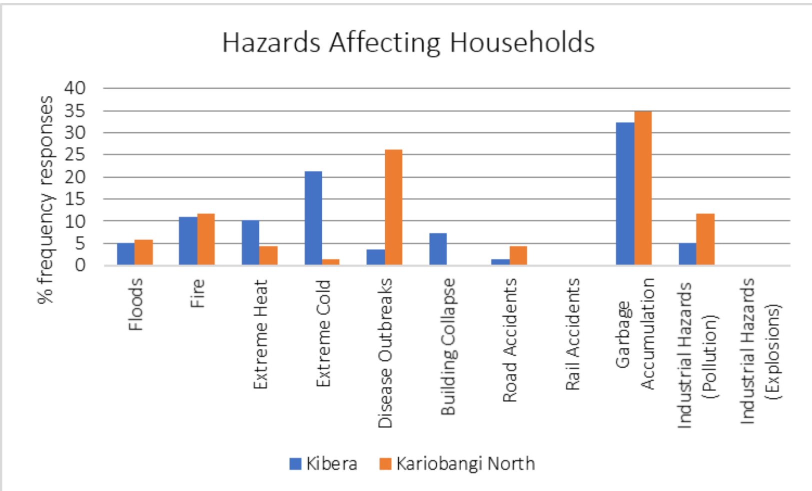


Figure 33: Hazards reported at household level.

To understand the interaction of the hazards captured at the household level, we asked the respondents the reason as to why they were affected by reported hazards based on four household dwelling characteristics: roofing, walls, floor, and geographic location. However, in the household survey design, we acknowledge that not all hazards affect the household based on the four aforementioned characteristics. Therefore, for the reported hazards of disease outbreaks and industrial pollution, we infer that the causes are similar to those reported at the settlement level.

For the hazards reported at the household level but not at the settlement level (extreme heat, extreme cold and building collapse), we found the type of flooring (34%) as the highest cause of extreme cold, closely followed by the type of walls (30%). Upon investigating the type of floor and wall material of by reporting household, we find that 53% have concrete plastering as they type of floor material and similarly 53% have iron sheets for walls. Also, all households reporting extreme heat had iron sheets for walls. On the other hand, the majority (50%) of households reporting building collapse had mud walls (without modifications) followed by mud walls with concrete plastering 38%. In Kariobangi North, only 4% and 1% reported experiencing extreme heat and cold. Of those experiencing extreme heat, 75% reported the type of roof as the cause. All the households had iron sheets for roofing material.

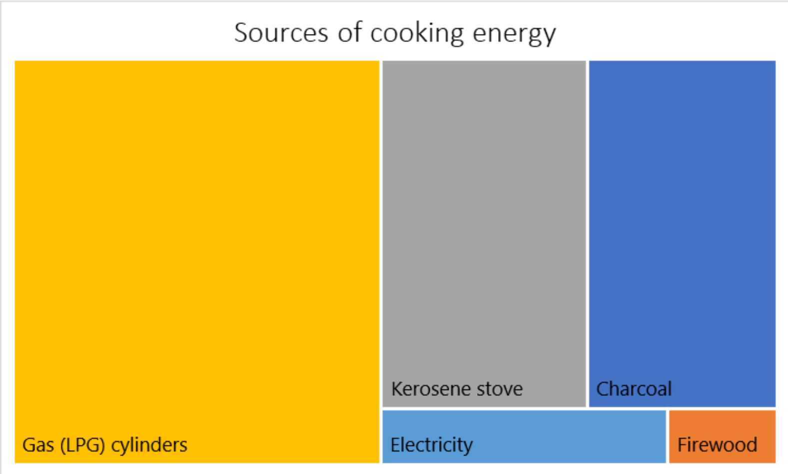


Figure 34: Reported sources of cooking energy in surveyed households.

Next, since fire is considered a significant hazard in deprived settlements, we deem it necessary to investigate additional household characteristics. From literature, we also found fire hazards are also attributed to the presence of combustible material mong other household characteristics and behavioural mannerisms (Ngau & Boit, 2020). These sentiments were also shared by one of the interviewed experts, a resident in a deprived settlement. The expert also highlighted an anticipated cause of fire within deprived areas: liquefied petroleum gas (LPG) cylinder explosions. Advancement in technologies has led to a shift in the sources of cooking energy from more traditional means such as charcoal and firewood to reliance on LPG as reflected in the household data collected, where 48% of all surveyed households rely on LPG, followed by kerosene (23%) and charcoal (20%) (fig.34).

In comparison, a study undertaken in 2014 found that charcoal and kerosene were the primary sources of cooking energy in Kibera, whereas only 0.5% of their respondents used LPG (Kamengere, 2014). Additionally, a majority of the households rely on electric power (89%) for lighting, and only 11% rely on kerosene lamps. These electricity connections are often illegally supplied by cartels who fill a service and utility void left by local governments (Langat & DiCampo, 2019). These findings are reflected in the settlement level assessment of fire causes. Additionally, the combustible building material is also considered a significant enhancer of fires (Ngau & Boit, 2020). Therefore, we also look into internal housing modifications made by households, such as polythene-based rugs/carpets (fig.35). Generally, synthetic materials are considered highly flammable in comparison to natural materials (Michalovič, 2014). We find that 26% of the respondents have bare/earth floors with polythene carpets, and 19% have polythene ceilings. Despite the low figures, their presence in densely populated areas poses a high risk of combustion and fire spreading.

Figure 35: Cemented floor with polythene-based carpet
Source: Household survey, Nairobi, 2021

Furthermore, since riverine flooding was considered a significant threat for deprived settlements by the experts, we investigated the causes of flooding at the household level. Important to note that both settlements had low reporting of flooding at the household level compared to other hazards, with only 6% in Kariobangi North and 5% in Kibera. In both settlements, the dwelling location is the main reason behind the households being affected by flooding (33%- Kibera, 40%-Kariobangi North). Interestingly, in Kibera, we found the type of roofing (29%) as the second most reported cause of flooding at the household level. In Kariobangi North, other reasons, including poor drainage infrastructure was the second largest cause of flooding, indicative of runoff flooding. In terms of household characteristics, the type of walls (20%) was the third reported cause of flooding at the household level in Kariobangi North.

Since the location of dwellings was the highest reported cause of flooding, and following the expert opinions that riverine flooding is a major hazard in deprived settlements, we investigate the distance of households from rivers. All households that reportedly experienced flooding due to their location had a collective average distance from rivers of 21meters, a significantly low value compared to the total average of 41meters for all surveyed households. In Kibera, the average distance from rivers of households experiencing flooding was 20meters, while in Kariobangi North, it was 27meters. Furthermore, the distances from rivers varied most in Kibera, with 71% of households reporting flooding being less than 10meters from the rivers. The remaining 29% were over 50meters from the rivers. This variation is possibly due to other flooding causes such as insufficient drainage systems and blocked drainage systems (runoff flooding), as captured in the above reporting, expert interviews and settlement level analysis.

4.3.3. Location And Household Characteristics: What Do They Tell Us About Hazards In Deprived Settlements?

From the multi-hazard analysis, we find that deprived settlements are generally located in hazardous areas. Additionally, we find that their exposure to hazards goes beyond the settlement's location and that there are intra-settlement and household variations to hazard exposure. Therefore, in this section, we first explore the understanding by experts of durable building materials and how those translate into different types of settlements in the urban space. Next, we investigate household characteristics, including reasons for living in the surveyed settlements, type of building material, and rental income. Further, we analyse the relationship between these household characteristics.

4.3.3.1. Discourse on Durable Housing

A discussion with the experts involved ranking from most to least durable different types of settlements (and housing) found in urban areas (tents, pavement dwellings, slums, informal settlements, formally planned settlements, shacks, old Swahili housing). The list, by large, captures different types of settlements and housing options within the study area, including non-conventional shelter options such as tents and pavement dwellings.

Among our first findings, it's highlighted that tents are no longer present in deprived areas within Nairobi and remain an emergency/transition shelter option found in IDP and refugee camps. Acknowledgement is, however, made that this reality could vary in different regions globally. On the other hand, Pavement dwellings were highlighted as a housing option that often gets left out in the discourse on informality. These findings are also captured by our analysis of household dwelling characteristics discussed below.

Next, we find that some clustering is possible for some of the housing/settlement options listed. For example, shacks were considered a housing option captured under slum settlements. Both were, however, regarded as better classified under 'informal settlements'. The discussions highlighted the overlap in our presented options, which we acknowledge as the complexity of socio-spatial marginalization. Further, there is the diverse coexistence of multiple housing types in what we refer to as deprived settlements in this study (see Mwau & Sverdlik, 2020). Nonetheless, the rankings result in interesting results, where surprisingly, mud-walled houses are considered most durable after stone (quarry stone) and brick-walled houses (see fig.36).

Figure 36: Common types of building material used in deprived settlement dwellings in Nairobi
Source: Household survey, Nairobi, 2021

Further, to unpack the complexity of housing/settlement options presented, we look into the reasoning given for the rankings by the experts. Reasons given for the rankings are based on the characteristics of building material and structural integrity, such as their robustness and adherence to building standards. Interestingly, the 'number of parties involved in putting up the structure' was also highlighted since it is closely linked to building standards. Also, the effects of tenure security on the pattern of development and living standards (quality of life-QoL) are emphasized. Unanimously, the experts highlight that the quality of building material is highly influenced by the level of investment - a factor of both willingness and financial ability, thus stressing the socio-economic aspect of housing.

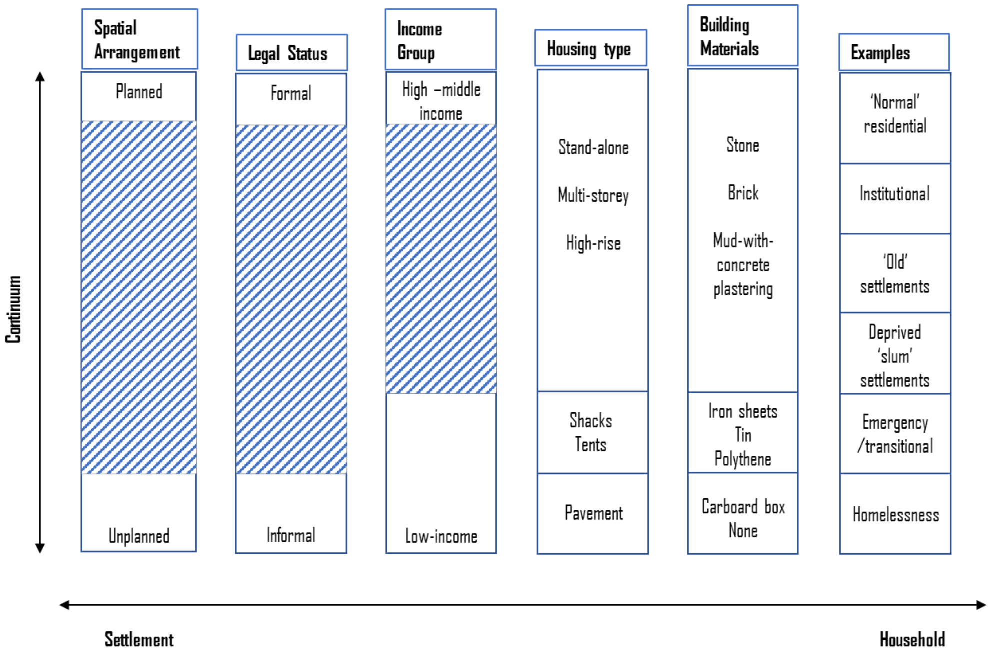


Figure 37: Discourse on durable housing as represented by five primary elements characterizing urban settlements

Consequently, we identify five primary elements: spatial arrangement, legal status, income group, type of housing and building material that summarize the discussion. Collectively, the elements form a conceptual model for the discourse on durable housing (fig.37). For ease of interpretation, we include examples presented to the experts for ranking. The x-axis is represented as a continuum given that there are no explicit 'cut-offs' for the elements. Also, by considering the elements to indicate likelihood, e.g. stone-walled houses are more likely to belong to high-middle income households with formal tenure and, in a planned neighbourhood, we better understand the different urban settlements/types of housing. We, however, acknowledge that this is a simplification of the very complex reality, especially in deprived settlements, especially when it comes to the type of housing (see fig.38). The y-axis shows the interaction among the elements, moving from the settlement to the household level and vice versa.

From our synthesis of the discourse on durable housing, the spatial arrangement of settlements is seen as an important 'entry point' in analysing urban settlements. Meaning, spatial patterns of settlements indicate their legality/tenure status and socio-economic conditions and status of the inhabitants. These give insight into the probable type of housing and the building materials used. Notably, this presents a top-down approach to understanding settlements and, by large, describes geospatial data and methods of analysing settlements. However, the inverse (a bottom-up approach) is also applicable whereby the type of building material informs the kind of housing, from which inference on the socio-economic and tenure status can be made. The cumulative deductions are revealed through the spatial arrangement of the settlement.

Figure 38: Image depicting three types of building material in one area. On the right is a stone-walled house with a second storey for an iron shack; on the left, a mud with concrete plastering house.

Source: Household survey, Nairobi, 2021

4.3.3.2. Why are Deprived Settlements Located in Hazardous Areas?

According to the experts, deprived areas are affected by hazards due to two main reasons: their location and the quality of the structure. As highlighted by the experts (also in literature see Ramin, 2009; Wekesa, Steyn, & Otieno, 2011), deprived settlements are located in non-conventional land use areas, including protected areas such as riparian reserves, abandoned quarries and generally land with low value. These, however, don't apply to all deprived settlements since some are found next to high-income neighbourhoods. Additionally, deprived households also exist in more formal neighbourhoods through unplanned or illegal in-fill practices. Further, the nature of deprivation in Nairobi is highlighted to have evolved over the years. Presently, the development of high-rise tenements is highlighted as a popular low-cost housing solution that accommodates the lower-middle-class (or upper-lower class) residents (see Mwau & Sverdlik, 2020). Nonetheless, the experts highlight a trade-off between hazard exposure to opportunities such as wealth, industries, markets, and mobility systems for the urban poor.

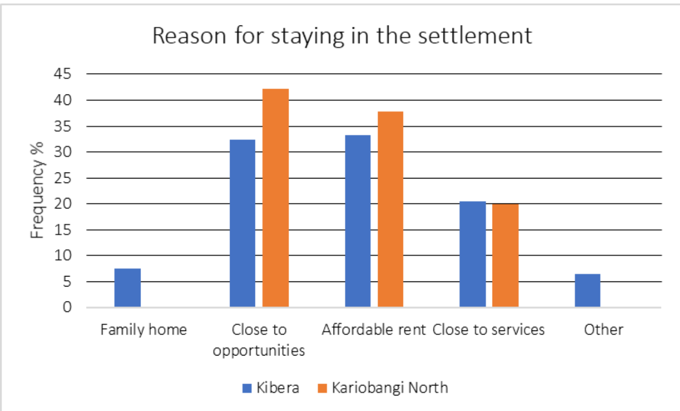


Figure 39: Respondents reasons for selecting the settlement they live in.

From the household survey, the interviewed households in Kariobangi North (42.2%) reflected the experts' sentiments by stating that their main reason for living in the settlement was the closeness to opportunities followed by rent affordability (37.8%) (fig.39). The main reason for the Kibera residents was the inverse, with the affordability of rent (33.3%) leading, followed by closeness to opportunities (32.3%) (fig.39). Closeness to services was the third leading reason for residents in both settlements. Further, the household survey shows that the paid rent per month by residents in Kibera is lower than in Kariobangi North. The majority of the Kibera residents pay approx. 20-30 USD in rent per month whereas Kariobangi North residents pay approx. 40-50 USD (fig.40). These values are significantly low compared to the reported median urban rental expenditures of approx.300USD in Kenya (KNBS, 2018). Further, only Kibera reported respondents who didn't pay rent, mainly because they/their families owned the structures.

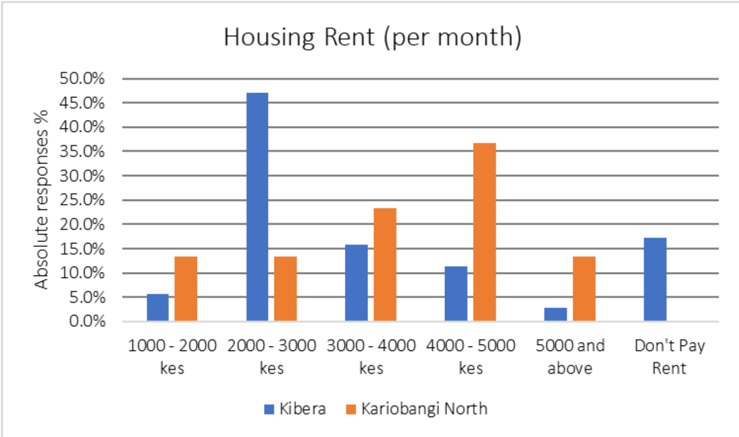


Figure 40: Comparison of house rent paid per month in Kibera and Kariobangi North

Discussions with the experts further reveal that rent affordability is linked to structure quality determined by the type of building materials used. In terms of building characteristics, we find that the type of walls is the important determinant for the rent. This was done by comparing the material used for roofing, walls, and floors in the surveyed settlements (fig.41). We find similarities in the materials used for roofing and flooring in the two settlements. Iron sheets are the most commonly used material for roofing and concrete plastering for flooring. On the other hand, significant differences are recorded for wall construction materials. In Kibera, mud with concrete plastering (51%) is the most common type of construction material, followed by iron sheets (23%) and mud (without any modifications) (19%). In Kariobangi North, stone walls (50%) are the most common, followed by mud with concrete plastering (27%) and iron sheets (17%).

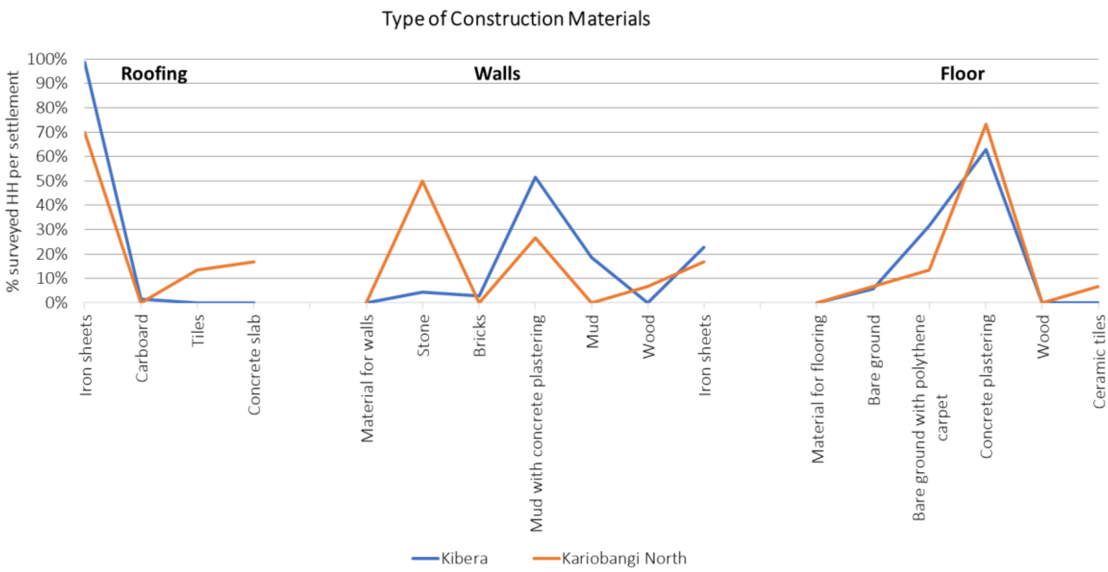


Figure 41: Comparison of building materials used for roofs, walls and flooring in the surveyed settlements.

Folowing these interesting results and considering the experts opinions, we analyse the relationship between building material and rent using the household dwelling characteristic with the most variance – walls. Since bricks and wood have low reporting in both settlements, we exempt them from the analysis. We find that housing with mud walls and concrete plastering, iron sheets and mud mainly cost approx. 20-30 USD whereas those made of stone walls are, however, more highly-priced at approx. 40-50USD (fig.42).

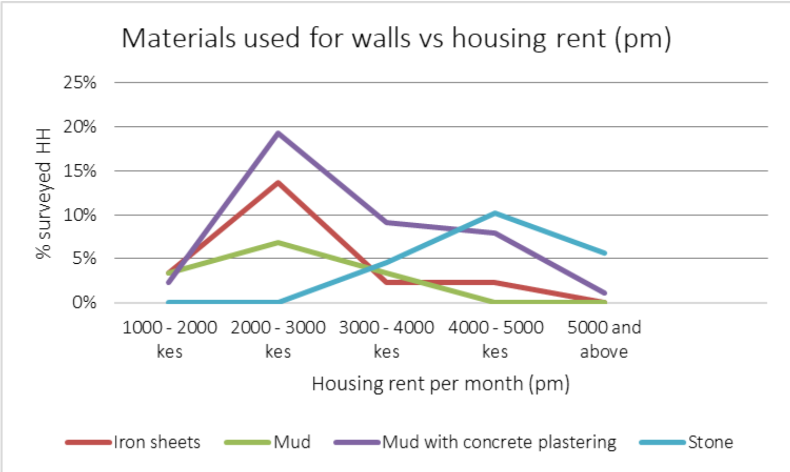


Figure 42: Comparison between building material and rent.

Lastly, as demonstrated by our analysis (section 4.3.2), literature and the expert interviews, the location of a settlement hence the dwelling, is also influenced by tenure security. Ofter, where deprived areas develop, they tend to lack tenure security, leading to structures with poor quality material. Two main reasons given for this relationship are (i) lack of access to financial mechanisms (Meinzen-Dick, 2009) and (ii) fear of demolitions as a form of forced evictions. Additionally, the building material used for housing influences the household's exposure to hazards. Descriptions given by two of the interviewed experts, who are residents of deprived settlements, highlighted these intra-settlement variations and further exposed the intra-settlement socio-spatial marginalization that exists in deprived settlements as described below:

- Houses near the river are often constructed using temporary materials due to their exposure to hazards. This is the preferred approach by slumlords to incur less financial losses in the advent of flooding.
- Housing near the bus stops, main entries into the settlement, and main settlement roads constitute more permanent housing due to the economic lucrativeness. These locations allow for more commercial use of the dwellings.

Our study highlights the complex nature of deprivation and the interplay of deprived settlement characteristics comprising tenure security and income that influence the location and building characteristics, which further determine the exposure to hazards of households. Using Kibera as an example where 7% of the households reported to be affected by building collapse (there were no reports of building collapse in Kariobangi North) (fig.33). The main reason for building collapse was found to be due to poor construction material (32%), followed by poor construction techniques (27%) and malicious demolitions (24%). Additionally, the dwelling walls were the most affected by building collapse, with a reporting of 53%, followed by the type of roof at 33%. Upon investigation of the type of building materials used for walls for households reported to be affected by building collapse, 50% had houses made of mud walls and 37.5% of mud with concrete plastering.

Despite mud being ranked third highest in durability after quarry stone and bricks, especially when modified by coating the surface with concrete plastering, these results aren't surprising. First, because the construction techniques are often poor (see fig.36). Also, since most of the residents are tenants, the responsibility of renovating the house belongs to the landlord. And, as demonstrated above, the landlords tend to lack a moral obligation to provide adequate and safe shelter for the tenants, an issue propagated by the privatization of housing that has resulted in the greed of maximizing profits at the expense of the lives of the urban poor.

4.4. 'Slums', Data, GEO-Ethics and Scientific Communication

4.4.1. Alternatives to the Term Slum

Literature highlights that the term 'slum' bears a negative connotation, resulting in alternative terms such as deprived settlements as applied in this study. We also find that deprived settlements have been misrepresented for decades under the term 'slum', globally adopted to refer to the urban residents living in squalor (Mayne, 2017). As further captured in his book, "Slums: The History of Global Injustice," Alan Mayne highlights the detrimental effects of this misrepresentation that goes beyond the terminology into influencing the types of interventions undertaken globally in attempts to eradicate slums.

We, therefore, find it essential to find out the terms used by our interviewed experts in their work. We find that all the experts use different words, with "informal settlements" being the most popular. Alternative terms such as "deprived settlements", "spontaneous settlements", and "home/community" are also mentioned (Table 20). The main reason behind the use of alternative terms is the need to *humanize* the residents of the said settlements, which enables the experts to shape and restructure their approaches in "slum" interventions and creates a sense of positive urgency in developing solutions for the residents of deprived settlements. The urban systems officer (expert) highlights this (paraphrased) by stating that the use of different terms other than 'slums':

"…*it gives the settlements a human aspect and allows one to view the people settled there as* ***clients****. This allows for one to understand the settlement better, including why the people choose to live there….*"

~Urban systems officer

Table 20: Summary of alternative terms to 'slum' used by interviewed experts.

Alternative Terms	Urban Policy Analyst	Urban Systems Officer	Human Settlements Program Officer	Spatial Data Expert	Professor in Geography	Deprives settlement resident (1)	Deprived settlement resident (2)
Informal	×	×			×		×
Spontaneous		×					
Deprived			×	×			
Low income					×		
Home/Community			×			×	
Ghetto							×

Of the responses, the term 'Home/community' emphasizes the need to adopt more applicable terms that are dignifying and how such seemingly minor changes help reshape how deprived settlements are viewed and studied/analysed. Additionally, the description of what a 'slum' is (paraphrased) by a deprived settlement highlights that the residents of these settlements are not blind to the challenges they face.

" Slums are home to a community of people living together but faced with many challenges that are beyond their control."

~Informal settlements resident I

Further, according to the programme officer (human settlements), a communal approach to defining the settlement can easily be translated into area-based approaches in defining settlements. Thus, presenting the opportunity to capture the totality of the challenges faced by the community and, in turn, provide opportunities for finding solutions.

"…there's no personal space. The community is interconnected, and individuals cannot detach themselves from the settlement…they share spaces and utilities such as communal bathrooms and toilets…."

~Human settlements program officer

4.4.2. Recommended Level of Aggregation/Disaggregation

As part of our study, we also aimed to investigate the recommended scale for mapping deprivation. The expert interviews reveal that the aim of the research determines the scale of use. However, there is a lack of a unified 'format' for data disaggregation and aggregation. Having identified three disaggregations/aggregation scales for deprivation mapping (ward level – lowest administrative unit in Nairobi, grids, and neighbourhood level -street block), they are presented to the experts.

Generally, street-blocks are found the most favourable unit for analysing deprivation. The main reason invoked from Tobler's first law of Geography is that "everything is related to everything else, but near things are more related than distant things" (p.236, Tobler, 1970). Furthermore, street blocks are helpful for localized reporting of findings, e.g., to the community. On the other hand, administrative units are reportedly most favourable for policy recommendation and implementation of interventions. At the same time, grids are the least prefered/recommended unit for analysis, except for scientific research purposes. Further, a distinction is made between two steps of research, i.e., analysis and reporting findings. Specifically, the reporting of results is seen as a more critical step since the findings of studies are intended for wider audiences than the analysis description.

4.4.3. Slum Data and Actors

Our study also aimed at identifying data sources and their use on deprived settlements. From the expert interviews, we find that there is a high reliance on non-spatial data. All interviewed experts rely on household interview generated socio-economic data and, only three of the experts rely on spatial data for their analysis. The main source of data being national census data. Besides the national census, the experts highlight that data on deprived settlements is hard to come by. Most of the data is generated by CBOs or NGOs working locally with the communities. Additionally, academic institutions and active local governments were identified as deprived settlement data producers. On the other hand, more actors were identified as interested groups in accessing data on deprived settlements. They comprise international development partners, donors and funders, individuals, and government bodies. This highlight's an imbalance between data production and demand.

Additionally, the experts highlight two main challenges facing data generation (and access) on deprived settlements: lack of collaboration between actors and a needs-driven approach to data generation. The experts state that actors often work in silos, resulting in unstandardized data at the local level, which hinders interoperability. Also, generally, there are no standardized frameworks for household data collection. A needs-driven approach, on the other hand, results in the generation of project-specific data. Meaning, the spatial coverage is limited since city-wide projects are rare. Also, only data on specific topic areas are captured. Thus, the data production process isn't dynamic or continuous. Additionally, the lack of centralized open-data repositories and coordination bodies within cities hinders data access and results in 'unnecessary' data reproduction (an expensive affair). In addition to the challenges mentioned above, data on deprived settlements often lack corresponding metadata and attribution.

4.4.4. Geo-Ethical Concerns

Regarding the use and sharing of data on deprived settlements, we focused on issues around the household data, and EO and GIS generated data capturing deprivation. The experts were interviewed on privacy issues and foreseeable concerns with advancements in the production of VHR data and the use of artificial intelligence methods in the analysis deprivation. The results show that when it comes to household data collection, personal information that includes personal identification details, health records data, and tribal related information is sensitive and should be protected. The development of open data repositories for data on deprived settlements was encouraged, with recommendations that the communities be the primary data custodians.

Regarding VHR data, the only concern is the residents' privacy, including exposing their locations, as this might subject them to evictions (due to their illegality). Notwithstanding, VHR imagery was prefered to HR imagery since deprived settlements are often crowded, making it hard to analyse using coarse imagery. A severe concern raised was the probable threat of combining VHR and artificial intelligence - since there are no known and set limitations to what AI algorithms extract from data. Furthermore, despite the advantages presented by EO in the mapping of deprived settlements, the accuracy of the mapping outcomes were questioned, including the validation processes. Especially since the methods are primarily top-down and most experts, including technocrats and organizations working in and with deprived communities, have little to no knowledge of EO based methods. Top-down approaches were further criticized for not integrating – especially the social dynamism of deprived communities. Despite the critical concerns, no expert had objections to using EO and AI methods. The principal recommendation is that scientists/researchers develop transferable knowledge that can be synthesized and used by practitioners and communities for local operations.

4.4.5. Democratization of Science

With advancements in science and technology, the participation of citizens is crucial. Dijkstra, Bakker, Dam, & Jensen (2020) highlight that communication about science goes beyond communicating scientific findings, and now more than ever, there's the need to democratize science since its implications affect everyone. Further, they highlight the need for inclusive practices and acknowledge the importance of citizen participation in identifying societal challenges and solutions. It is for this reason that we ask local experts for their opinions on this matter. We recognise the following themes that highlight current challenges, causes of the challenges, and some possible solutions from the discussion.

(i) Patch-Work Research

Due to the previously mentioned challenge of project-based approaches to interventions, most research ends up being "*patchwork*" since they lack continuity or connectivity to existing and previous works. Thus, research projects gradually lose their potential to impact societies positively. This highlights the dire lack of collaboration among actors in different fields of study. Additionally, we find that the lack of cooperation among actors working with deprived settlements worsens the situation. Further, we find that lack of collaboration is also an identified challenge to developing multi-hazard approaches (Melanie S. Kappes et al., 2012).

(ii) Extractive Research

The possible causes for the challenges highlighted above seem to arise from the lack of sufficient engagement of communities in research practices. We find that researchers often fail to engage the very communities they are researching through simple procedures such as data collection processes, where often non-community members are contracted instead of the community members. Also, collected data and study findings are hardly communicated back to communities. Generally, it is seen that researchers lack the flexibility to adapt their research interests to fit community needs or priorities. Therefore, from a community member's standpoint, researchers approach communities with pre-defined problems. As a result, communities are treated as subjects of science and not a part of it. Communities are locked out of the process of formulating their narratives and understanding their challenges despite having an interest. For instance, from the household survey, 9 out of 10 households were interested in our research findings with their preferred language of communication being both English and Kiswahili (the official and national languages in Kenya) and, the top three preferred modes of communication being through Social Media platforms, website links and posters.

5. DISCUSSION

The main aim of this study was to analyse the relationship between deprivation and hazards in cities. The aim also formed the title of this study, from which three sub-objectives were formulated. The first is to identify hazards within our study area and spatial indicators for constructing a multi-hazard index. Secondly, we aimed to test the predictability of deprivation from multi-hazards. This was contrasted to the use of conventional data. Also, the multi-hazard index was tested by comparing the degree of hazardousness among different types of residential settlements within Nairobi. Thirdly, to test the realizations of the multi-hazard index, household interviews were conducted in two types of deprived settlements within Nairobi.

In this chapter, the research findings are discussed, and the aim of the study is evaluated. Challenges that were encountered are also discussed, and possible solutions presented.

5.1. Identification of Hazards and Hazard Indicators

Identifying hazards to construct the multi-hazard index was a challenge for this study because very few studies have conducted interdisciplinary multi-hazard analyses, especially in urban areas (Greiving, 2006; Melanie S. Kappes et al., 2012). The challenge extends to identifying spatial data that serve as proxies of hazards due to unavailability of data, including coherent frequency, magnitude or occurrence of the hazards data. Further, analysing multi-hazards requires multi-source data with different measurement units (Melanie Simone Kappes, 2011). Thus, we implement a simple method for multi-hazard assessment –a susceptibility index implemented at the city scale.

Generally, we find that the inner city is highly hazardous in comparison to the periphery. And that riverine flooding, road and rail accidents are spread throughout the city, whereas fire and industrial accidents have distinct hotspot areas. Interestingly, fire hazard hotspot regions shape major deprived settlements such as Kibera and Mukuru. Runoff flooding is moderate through the city, with the river channels having high susceptibility. Despite having lesser development, the city’s eastern region has a higher degree of hazardousness due to extreme temperatures, which can be due to the climatic zones of the country. Air pollution, on the other hand, follows the opposite trend. Where the central and western regions are more hazardous, we found that the winds in Nairobi typically blow from the city's eastern region. Epidemics and air transport accidents have hotspots regions innately influenced by the features used to assess the hazards, i.e., the location of dumpsites and airports, respectively.

5.2. Relationship between Hazards and Deprivation

Our analysis revealed that the inner city is the most hazardous region. As a result, in comparing the exposure to hazards of different types of deprived and non-deprived settlements, we find that deprived settlements type I and high-mid density settlements are prone to more hazards than deprived settlements type II and low-density settlements. Generally, deprived settlements type I and high-mid density settlements have a similar trend for all analysed sub-hazard categories with exemption to fire hazards. Since the assessment of fire hazards is based on indicators that capture the physical traits of settlements, we find that similarly, deprived type II settlements, despite their primary location being farther from the urban core, have high susceptibility to fire hazards. Deprived settlements are often crowded and contiguous, thus highly dense and lacking adequate road infrastructure. These findings are in agreement with the expert opinions that fire hazards are more prone in deprived settlements.

Also, another notable difference between deprived settlements type I and high-mid density settlements is the variance in the location of the samples. High-mid density settlements have a higher variance and negatively skewed distribution for most hazards (riverine flooding, transport accidents, industrial accidents, extreme temperatures and air pollution), indicating that they are not typical to hazardous areas. The variance can, however, be explained by the diversity in these settlements. Our morphological categorization of high-mid density settlements comprises different types of settlements whereby typically, the highest density settlements are closer to the urban core. In contrast, the mid-density settlements are towards the periphery.

Low-density settlements are the least susceptible to hazards since they are located in the periphery. The assessment, however, reveals that some samples of low-density settlements are located in hazardous areas, with the highest variances observed for susceptibility to runoff flooding, epidemic hazards and air pollution. Further, the high variance in the location of low-density settlements reflects the changes that Nairobi has undergone, from the historical colonial period where these settlements were typical to the northwestern region. This is demonstrated especially by the variance in air pollution susceptibility. Whereas the data is positively skewed, indicating that many low-density settlements are in the western region, where exposure to air pollution is higher, there are instances of low-density settlements in the less hazardous eastern region. Similarly, deprived settlements type II in the less hazardous areas, especially the city's western region, reflect these changes.

These changes show the implications of urban governance processes. Like many colonial cities (Home, 2014), Nairobi has a prejudiced history against non-domineering social groups. From the case study review, we find that historical influences are present in the city. These are reflected in both the location of settlements and the type of buildings. In terms of location, we find from the multi-hazard index assessment that the city's central region is most susceptible to hazards. In reference to the colonial planning, these areas were designated for non-Europeans (fig. 6). Additionally, poor urban governance systems demonstrated by the historical development of the city, characterised by city boundary extensions that are unaccompanied by plans, policies that are implemented without adequate systems in place, have propagated inequality. For instance, the privatization of housing, a primary need and basic public good, resulted in inadequate provision of low-cost housing and the failure to adhere to planning and building standards (Gatabaki-Kamau & Karirah-Gitau, 2004). The challenge extends to the lack of services and infrastructures in, especially, deprived settlements. As a result, deprived settlements have proliferated formally planned areas. Further, illegal urban infill practices as observed from satellite imagery (also mentioned by some experts) have resulted in denser planned residential neighbourhoods.

These effects are further demonstrated by assessing indicators used by the analysis to discriminate settlements. We found that deprived settlements are characterized by low vegetation cover, high proximity to rivers, high building density, low road density and high proximity to airports (section 4.3.1), whereas the opposite characterises low-density settlements. These differences highlight the hazardous nature of deprived settlements. They also indicate the stack socio-economic polarization in Nairobi since roads and vegetation are morphological traits that distinguish affluent neighbourhoods from deprived neighbourhoods (Kohli et al., 2012; Kuffer, Barros, & Sliuzas, 2014). On the other hand, high-mid density settlements are primarily characterized by high proximity to major roads, high nighttime LST, CO, high road and building density. These were interesting results of our study that though non-indicative of deprivation, these hazard indicators highlight the more anthropogenic effects of urban development that contribute to hazardous cities.

Further, by assessing the type of buildings and materials used for construction in two deprived settlements (type I and II), we found that buildings made of durable material are much more high priced (monthly rent) and that the location of a dwelling within a settlement influences the type of building materials. Further, more hazard-prone areas are also areas with non-durable housing (both in terms of building material and techniques) and thus cheaper, a technique used by slumlords to minimize losses in the advent of a disaster. Also, from expert discussions, it was inferred that the household's income affects the kind of housing. These findings were reflected by the household survey, whereby affordability of rent was the second most report reason for living in a deprived settlement, following the need to access opportunities. Both are reasons closely tied to the socio-economic well-being of these populations. We, therefore, find that socio-spatial marginalization is rampant at both the intra-city and intra-settlement levels, where the poorest are most exposed to hazards.

We further contrast the multi-hazard index and household survey outcomes. We found that despite some hazards having high scores from the index, their reporting at settlement and household levels was low. The difference in scope of analysis can explain these differences. While the index is a broad assessment, the household responses are more localized. Further, the selection of multi-hazard proxies proved sufficient at capturing the hazards. However, as shown by the household responses, we acknowledge that better proxies can be used for some hazards. For instance, specific to rail and road transport accidents, the absence of pedestrian crossings is one of the leading reasons that was reported. Therefore, these data would better capture these hazards in place of Euclidean measures from road and rail infrastructure. Also, epidemics measured from the presence of dumpsites would be better captured using disaggregated data. Especially given that garbage accumulation was the highest reported hazard at both settlement and household levels. Generally, we found settlement level hazards similar to those reported at the household level, except for hazard exposures influenced by the characteristics of the dwelling, such as extreme temperatures (section 4.3.2).

Given these pre-existing conditions, the projected increase of disasters and climate change-related risks in urban areas will only accentuate the risks inhabitants of deprived settlements face (Revi, Satterthwaite, et al., 2014). Therefore, proactive measures such as adaptation and building resilience must be undertaken and targeted at those living in deprivation: an approach proposed by the United Nations, (2017) New Urban Agenda for the achievement of Sustainable Development Goal (SDG)11 aiming at inclusivity, safety, resilience and sustainability in cities.

5.3. Geospatial Data and Methods

Geospatial data and methods have proven revolutionary in assessing urban poverty by filing a gap that has existed for decades – the mapping of slums. The importance of geospatial data is captured by the discourse on durable housing (section 4.3.1) that demonstrates how the spatial arrangement of settlements provides a way to understand more complex challenges at more local levels, e.g. household level. Additionally, they have proven useful through this study on multi-hazard and deprivation—our attempts at using multi-hazard datasets to map deprivation results in high model accuracies. However, the data generalize the results. As a result, we find that conventional texture features, although obtaining a slightly lower accuracy and rely on many more predictor variables, yield better results.

Furthermore, despite the aforementioned advantages, geospatial methods are generally top-down since they limited interaction with the people being mapped. Furthermore, these techniques have been found complex and needing simplification for comprehension and transferability. The use of satellite imagery has further been found to potentially infringe on the locational privacy of the urban poor, thus putting the already vulnerable in worse situations such as the potential threat of eviction. These issues are alarming and need addressing, especially since higher resolution imagery is being produced and more powerful automated artificial intelligence algorithms are developing. In addition, the challenges of deprived settlements go beyond acknowledging their existence since they are faced by layered deprivations (as highlighted in the definition provided) and therefore require a combination of physical and social scientific approaches.

Dijkstra, Bakker, Dam, & Jensen (2020) highlighted that one way of including social scientific knowledge in science and technology's growing and evolving body is through participatory approaches. Such approaches have produced a wealth of knowledge and the posterior adoption of the tools by communities to address other challenges (Cornwall & Jewkes, 1995). This example shows the advantage of participatory approaches at redistributing power and demonstrates the importance of empowering communities through their inclusion in finding solutions to their problems. However, from the literature review, the lack of integration of participatory approaches has persisted (Cornwall & Jewkes, 1995), creating a nuanced challenge. Where, on the one hand, the physical scientific methods provide transparency by offering a common motivating factor to influence decision making (Borie, Pelling, Ziervogel, & Hyams, 2019), i.e. mapping deprivation; on the other hand, the top-down nature of EO data and methods challenge the notion of creating and strengthening the relationship between science and society.

6. CONCLUSION AND RECOMMENDATIONS

While typical assessments rely on occurrence or magnitude data for hazard analysis, the unavailability of these data resulted in the construction of a multi-hazard susceptibility index. Multi-hazard assessments allow for the estimation of the overall degree of hazardousness in a city (Melanie Simone Kappes, 2011), as demonstrated by the hotspot analysis carried out in this study. Our study shows that settlements located in the inner city are prone to hazards. As a result, both deprived settlements and high-mid density settlements are affected. Therefore, the study agrees with the literature's perception that deprived settlements are located in hazardous areas. However, another category of deprived settlements is located in the city's periphery and are thus less prone to hazards. Hence, better and more consistent approaches for delineating deprived settlements is required for future studies to ensure that the different types of settlements are adequately captured.

Further, since a susceptibility index acts as a useful starting point for localized hazard assessments, we conducted household interviews with two aims in mind. We first localized the hazard assessment, whereby we uncovered an interesting finding that there is intra-settlement socio-spatial marginalization. However, provided the small proportion of households, the survey is considered an indicative sample. Secondly, we found that the proxies selected for indicating respective hazards performed well. Nonetheless, we acknowledge that better-suited indicators can be used for some hazards. Some indicators were acknowledged while constructing the index, but the lack of open data hindered their operation. Nonetheless, all the data used in our analysis comprises open geospatial data, which proved satisfactory for constructing the multi-hazard index. In addition to employing a simple multi-hazard assessment method, our study provides a simple to implement and easily replicable technique for hazard assessment. Further, by considering multi-hazards simultaneously, dependencies and interrelations between hazards were also made (Gallina et al., 2016), leading to the further understanding of hazards (section3.4.3).

Lastly, at a practical level of assessing urban poverty, the importance of the terminology used to refer to deprived settlements is highlighted. The continued efforts to find more dignifying terms is commended and seen to redefine the services and approaches employed in works aimed at improving the lives of the urban poor. Specifically, as demonstrated by the failure of previous 'slum' intervention programmes and as captured by Mayne (2017), this can be done by employing people-centred methods. For instance, we conducted expert interviews to aid in identifying hazards within our study area. We found that interviews by experts in the related fields covered by this study provided insightful information, e.g., identifying hazards. The interview outcomes matched the results of the different assessments conducted in this study, highlight the importance of infusing local knowledge in scientific studies. Thus, the importance of people-centred methods is reflected in this study and recommended for integration in geospatial approaches that are typically top-down.

7. REFERENCES

Abascal, Á., Rothwell, N., Shonowo, A., Thomson, D., Elias, P., Elsey, H., … Kuffer, M. (2021). "Domains of Deprivation Framework" for Mapping Slums, Informal Settlements, and Other Deprived Areas in LMICs to Improve Urban Planning and Policy: A Scoping Review. *Preprints*, (March), 1–37. https://doi.org/10.20944/preprints202102.0242.v1

Adiga, A., Chu, S., Eubank, S., Kuhlman, C. J., Lewis, B., Marathe, A., … Wilson, M. L. (2018). Disparities in spread and control of influenza in slums of Delhi: Findings from an agent-based modelling study. *BMJ Open*, *8*(1), 1–12. https://doi.org/10.1136/bmjopen-2017-017353

Ajami, A., Kuffer, M., Persello, C., & Pfeffer, K. (2019). Identifying a Slums' Degree of Deprivation from VHR Images Using Convolutional Neural Networks. *Remote Sensing*, *11*(11), 1282. https://doi.org/10.3390/rs11111282

Aksu, I., Bazilevskaya, E., & Karpyn, Z. T. (2015). Swelling of clay minerals in unconsolidated porous media and its impact on permeability. *GeoResJ*, *7*, 1–13. https://doi.org/10.1016/J.GRJ.2015.02.003

Baker, J. L. (2008). *Urban Poverty: A Global View*. Washington,DC.

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Borie, M., Pelling, M., Ziervogel, G., & Hyams, K. (2019). Mapping narratives of urban resilience in the global south. *Global Environmental Change*, *54*(December 2018), 203–213. https://doi.org/10.1016/j.gloenvcha.2019.01.001

Brownlee, J. (2014, March 21). Classification Accuracy is Not Enough: More Performance Measures You Can Use. Retrieved August 12, 2021, from Machine Learning Process website: https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/

Calder, A., Medland, A., Dent, A., & Allen, P. (2009). Understanding patterns of deprivation. *Regional Trends*, *41*(1), 93–114. https://doi.org/10.1057/rt.2009.7

Cano, A. B. (2019). Urban Poverty. In *The Wiley Blackwell Encyclopedia of Urban and Regional Studies* (pp. 1–7). https://doi.org/10.1002/9781118568446.eurs0388

Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social Science and Medicine*, *41*(12), 1667–1676. https://doi.org/10.1016/0277-9536(95)00127-S

Dierkx, D. (2019). *The Process Of Urban Renewal In Nairobi , Kenya : A Case Study On Citizen Participation Within The Renewal Of Nairobi City County Estates* (Wageningen University). Retrieved from https://edepot.wur.nl/476345

Dijkstra, A. M., Bakker, L. de, Dam, F. van, & Jensen, E. A. (2020). Setting the scene for Science Communication. In F. van Dam, L. de Bakker, A. M. Dijkstra, & E. A. Jensen (Eds.), *Science Communication, an Introduction* (pp. 1–16). Singapore: World Scientific Publishing Company.

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., Margaret, A., Jonathan, A., … Yetman, G. (2005). Natural Disaster Hotspots: A Global Risk Analysis. In *World bank*. https://doi.org/10.1002/j.1834-4461.1948.tb00495.x

EM-DAT. (2009). General Classification . Retrieved June 16, 2020, from EM-DAT: The International Disaster Database website: https://www.emdat.be/classification

Engstrom, R., Harrison, R., Mann, M., & Fletcher, A. (2019). Evaluating the relationship between contextual features derived from very high spatial resolution imagery and urban attributes: A case study in Sri Lanka. *2019 Joint Urban Remote Sensing Event, JURSE 2019*, (August), 1–4. https://doi.org/10.1109/JURSE.2019.8809041

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE Publications.

Gallina, V., Torresan, S., Critto, A., Sperotto, A., Glade, T., & Marcomini, A. (2016). A review of multi-risk methodologies for natural hazards: Consequences and challenges for a climate change impact assessment. *Journal of Environmental Management*, *168*, 123–132. https://doi.org/10.1016/j.jenvman.2015.11.011

Gatabaki-Kamau, R., & Karirah-Gitau, S. (2004). Actors and Interests: The Development of an Informal Settlement in Nairobi, Kenya. In M. Vaa & K. Hansen Tranberg (Eds.), *Reconsidering Informality: Perspectives from Urban Africa* (pp. 158–175). Uppsala, Sweden: Nordiska Afrikainstitutet.

Gattoni, G., & Patel, P. (1974). *Residential Land Utilization Case Study: Nairobi*. Massachusetts Institute of Technology.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2015). VSURF: An R package for variable selection using random forests. *R Journal*, *7*(2), 19–33. https://doi.org/10.32614/rj-2015-018

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E. (2018). Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience & Remote Sensing*, *55*(2), 221–242. https://doi.org/10.1080/15481603.2017.1408892

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Retrieved from http://www.deeplearningbook.org/contents/intro.html

Government of Kenya. (2009). *National Policy for Disaster Management in Kenya*. Nairobi.

Government of Kenya (GoK). (2010). *National Climate Change Response Strategy*. (April). Retrieved from http://cdkn.org/wp-content/uploads/2012/04/National-Climate-Change-Response-Strategy_April-2010.pdf

Greiving, S. (2006). Integrated Risk Assessment of Multi-Harards: A New Methodology. In *Geological Survey of Finland, Special Paper* (Vol. 42).

Hall, K. K., Evanshen, B. G., Maier, K. J., & Scheuerman, P. R. (2014). Application of Multivariate Statistical Methodology to

Model Factors Influencing Fate and Transport of Fecal Pollution in Surface Waters. *Journal of Environmental Quality*, *43*(1), 358–370. https://doi.org/10.2134/jeq2013.05.0190

Haralick, R. M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Helber, P., Gram-Hansen, B., Varatharajan, I., Azam, F., Coca Castro, A., Kopackova-Strnadova, V., & Bilinski, P. (2018). *Mapping Informal Settlements in Developing Countries with Multi-resolution, Multi-spectral Data*.

Herbert, D. T. (1975). Urban Deprivation: Definition, Measurement and Spatial Qualities. *The Geographical Journal*, *141*(3), 362. https://doi.org/10.2307/1796471

Hoegh-Guldberg, O., Jacob, D., & Taylor, M. (2018). Impacts of 1.5°C of Global Warming on Natural and Human Systems. In S. Bindi, I. Brown, A. Camilloni, R. Diedhiou, K. L. Djalante, F. Ebi, … G. Zhou (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,* (pp. 175–311). Retrieved from https://www.ipcc.ch/sr15

Home, R. (2014). Shaping Cities of the Global South. In S. Parnell & O. Sophie (Eds.), *The Routledge Handbook on Cities of the Global South* (pp. 75–85). https://doi.org/10.4324/9780203387832.ch9

Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, *182*, 147–156. https://doi.org/10.1016/j.geomorph.2012.11.005

Kamengere, R. N. (2014). *University of Nairobi Faculty of Arts Department of Sociology Assessing Fire Hazards Reduction Capabilities in Nairobi'S Kibera Informal Settlements a Project Report Submitted To Faculty of Arts in Partial Fulfillment of Requirements for the Award of Mast*. (October). Retrieved from http://erepository.uonbi.ac.ke/bitstream/handle/11295/76188/Kamengere_Assessing Fire Hazards Reduction Capabilities In Nairobi's Kibera Informal Settlements.pdf?sequence=4&isAllowed=y

Kappes, Melanie S., Keiler, M., von Elverfeldt, K., & Glade, T. (2012). Challenges of analyzing multi-hazard risk: a review. *Natural Hazards*, *64*(2), 1925–1958. https://doi.org/10.1007/s11069-012-0294-2

Kappes, Melanie Simone. (2011). Multi-Hazard Risk Analyses : a Concept and its Implementation. Universitat wien.

Kenya National Bureau of Statistics (KNBS). (2019). Population by County and Sub-county. In *2019 Kenya Population and Housing Census*. Retrieved from https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county

Kim Rutledge, Tara Ramroop, Diane Boudreau, Melissa McDaniel, Santani Teng, Erin Sprout, … Jeff Hunt. (2011, January 21). Runoff. Retrieved May 18, 2021, from National Geographic Society - Resource Library website: https://www.nationalgeographic.org/encyclopedia/runoff/

Kiplagat, S. (2021, July 15). Court orders NMS to close Dandora dumpsite within six months - Business Daily. Retrieved August 14, 2021, from Business Daily website: https://www.businessdailyafrica.com/bd/news/counties/court-orders-nms-dandora-dumpsite-within-six-months-3474214

KNBS. (2018). *Basic Report on Well-being in Kenya -Based on the Kenya Integrated Household Budget Survey (KIHBS)*. Nairobi.

Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, *36*(2), 154–163. https://doi.org/10.1016/j.compenvurbsys.2011.11.001

Krausmann, E., Cozzani, V., Salzano, E., & Renni, E. (2011). Industrial accidents triggered by natural hazards: an emerging risk issue. *Natural Hazards and Earth System Sciences*, *11*(3), 921–929. https://doi.org/10.5194/nhess-11-921-2011

Kuffer, M., Barros, J., & Sliuzas, R. V. (2014). The development of a morphological unplanned settlement index using very-high-resolution (VHR) imagery. *Computers, Environment and Urban Systems*, *48*, 138–152. https://doi.org/10.1016/j.compenvurbsys.2014.07.012

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space-15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6). https://doi.org/10.3390/rs8060455

Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1830–1840. https://doi.org/10.1109/JSTARS.2016.2538563

Kuffer, M., Thomson, D. R., Boo, G., Mahabir, R., Grippa, T., Vanhuysse, S., … Kabaria, C. (2020). The role of earth observation in an integrated deprived area mapping "system" for low-to-middle income countries. *Remote Sensing*, *12*(6). https://doi.org/10.3390/rs12060982

Kuffer, M., Wang, J., Nagenborg, M., Pfeffer, K., Kohli, D., Sliuzas, R., & Persello, C. (2018). The Scope of Earth-Observation to Improve the Consistency of the SDG Slum Indicator. *ISPRS International Journal of Geo-Information*, *7*(11), 428. https://doi.org/10.3390/ijgi7110428

Langat, A., & DiCampo, P. (2019, February 15). In Kenya's Largest Slum, the World Bank Battles Cartels for Control of Electricity | Pulitzer Center. Retrieved June 2, 2021, from https://pulitzercenter.org/stories/kenyas-largest-slum-world-bank-battles-cartels-control-electricity

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Liu, R., Kuffer, M., & Persello, C. (2019). The temporal dynamics of slums employing a CNN-based change detection approach. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232844

Lobo, A. (1997). Image segmentation and discriminant analysis for the identification of land cover units in ecology. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(5), 1136–1145. https://doi.org/10.1109/36.628781

Lou, P., Fu, B., He, H., Li, Y., Tang, T., Lin, X., … Gao, E. (2020). An Optimized Object-Based Random Forest Algorithm for

Marsh Vegetation Mapping Using High-Spatial-Resolution GF-1 and ZY-3 Data. *Remote Sensing*, *12*(8), 1270. https://doi.org/10.3390/rs12081270

Malamud, B. D., Mwangi, E., Gill, J., Hussain, E., Taylor, F., & Sakic, R. (2021). *Multiple-hazards and their interactions in urban low-to-middle income countries : a case study from Nairobi*. https://doi.org/https://doi.org/10.5194/egusphere-egu21-16053

Martínez, J., Pfeffer, K., & Baud, I. (2016). Factors shaping cartographic representations of inequalities. Maps as products and processes. *Habitat International*, *51*, 90–102. https://doi.org/10.1016/j.habitatint.2015.10.010

Mayne, A. (2017). *Slums: The History of a Global Injustice*. Retrieved from https://books.google.nl/books?id=b15xDwAAQBAJ

Mboga, N., Persello, C., Bergado, J., & Stein, A. (2017). Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. *Remote Sensing*, *9*(11), 1106. https://doi.org/10.3390/rs9111106

Meinzen-Dick, R. (2009). Property Rights for Poverty Reduction? In *Economic and Social Affairs* (No. 91). Retrieved from http://www.ifpri.org/sites/default/files/publications/beijingbrief_meinzendick2.pdf

Michalovič, R. (2014). Fire Safety Assessment on Seven Flooring Materials. *Research Journal of Recent Sciences*, *3*(10), 59. Retrieved from www.isca.me

Mitullah, W. (2003). Urban Slums Reports: The case of Nairobi, Kenya Table. In *Understanding Slums: Case Studies for the Global Report on Human Settlements 2003* (pp. 1–22). Nairobi, Kenya: UN-Habitat.

Muguro, J. K., Sasaki, M., Matsushita, K., & Njeri, W. (2020). Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020. *Cogent Engineering*, *7*(1), 1797981. https://doi.org/10.1080/23311916.2020.1797981

Müller, I., Taubenböck, H., Kuffer, M., & Wurm, M. (2020). Misperceptions of Predominant Slum Locations? Spatial Analysis of Slum Locations in Terms of Topography Based on Earth Observation Data. *Remote Sensing*, *12*(15), 2474. https://doi.org/10.3390/rs12152474

Mulligan, J., Harper, J., Kipkemboi, P., Nobi, B., & Collins, A. (2017). Community-responsive adaptation to flooding in Kibera, Kenya. *Engineering Sustainability 170*, *170*(ES5), 268–280.

Mwau, B., & Sverdlik, A. (2020). High rises and low-quality shelter: rental housing dynamics in Mathare Valley, Nairobi. *Environment and Urbanization*, *32*(2), 481–502. https://doi.org/10.1177/0956247820942166

Nagpal, A. (2017). Decision Tree Ensembles- Bagging and Boosting – Towards Data Science. Retrieved July 26, 2021, from Towards data science website: https://towardsdatascience.com/decision-tree-ensembles-bagging-and-boosting-266a8ba60fd9

Ngau, P. M., & Boit, S. J. (2020). Community fire response in Nairobi's informal settlements. *Environment and Urbanization*, *32*(2), 615–630. https://doi.org/10.1177/0956247820924939

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Medeiros, G., Silveira, A., … Saleska, S. (2011). Height Above the Nearest Drainage - a hydrologically relevant new terrain model. *Journal of Hydrology*, *404*, 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

Owusu, M. (2020). *Spatio-Temporal Slum Mapping At Citywide Scale Using High-Resolution Images*. University of Twente.

Pamoja Trust. (2009). *An Inventory of the Slums in Nairobi*. Retrieved from http://www.irinnews.org/pdf/nairobi_inventory.pdf

Persello, C., & Stein, A. (2017). Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2325–2329. https://doi.org/10.1109/LGRS.2017.2763738

Phillips, T., Friedman-rudovsky, J., Gomes, L., Lafreniere, J., Mbachu, D., Lafreniere, J., … Irin, M. W. (2007). *Tomorrow ' s crises today: the humanitarian impact of urbanization*. Malta: OCHA, IRIN & UN-Habitat.

Reitz, A., Hemric, E., & Hall, K. K. (2021). Evaluation of a multivariate analysis modeling approach identifying sources and patterns of nonpoint fecal pollution in a mixed use watershed. *Journal of Environmental Management*, *277*(August 2020), 111413. https://doi.org/10.1016/j.jenvman.2020.111413

Revi, A., David E., S., Aragón-Durand, F., Corfee-Morlot, J., Kiunsi, R. B. R., Pelling, M., … Solecki, W. (2014). Urban areas. In C. B., V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, … L. L. White (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 535–612). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Revi, A., Satterthwaite, D. E., Aragón-Durand, F., Corfee-Morlot, J., Kiunsi, R. B. R., Pelling, M., … Solecki, W. (2014). Urban areas. In C.B., V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, … L. L. White (Eds.), *Climate Change 2014 Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 535–612). Cambridge, United Kingdom and New York,NY,USA: Cambridge University Press.

Saabas, A. (2014, October 19). Interpreting random forests . Retrieved June 18, 2021, from Diving into data website: http://blog.datadive.net/interpreting-random-forests/

Saini, R., Verma, S. K., & Gautam, A. (2021). Implementation of Machine Learning Classifiers for Built-up Extraction Using Textural Features on Sentinel-2 Data. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, 1394–1399. https://doi.org/10.1109/ICACCS51430.2021.9441713

Sanusi, Y. A. (2008). Application of human development index to measurement of deprivations among urban households in Minna, Nigeria. *Habitat International*, *32*(3), 384–398. https://doi.org/10.1016/j.habitatint.2007.11.009

Seto, K. C., Sánchez-Rodríguez, R., & Fragkias, M. (2010). The New Geography of Contemporary Urbanization and the Environment. *Annual Review of Environment and Resources*, *35*(1), 167–194. https://doi.org/10.1146/annurev-

environ-100809-125336
Seto, K. C., & Shepherd, J. M. (2009). Global urban land-use trends and climate impacts. *Current Opinion in Environmental Sustainability*, *1*(1), 89–95. https://doi.org/10.1016/j.cosust.2009.07.012
Silva, S. H. G., Menezes, M. D. de, Mello, C. R. de, Góes, H. T. P. de, Owens, P. R., & Curi, N. (2016). Geomorphometric tool associated with soil types and properties spatial variability at watersheds under tropical conditions. *Scientia Agricola*, *73*(4), 363–370. https://doi.org/10.1590/0103-9016-2015-0293
Thomson, D. R., Kuffer, M., Boo, G., Hati, B., Grippa, T., Elsey, H., … Kabaria, C. (2020). Need for an integrated deprived area "slum" mapping system (IDEAMAPS) in low-and middle-income countries (LMICS). *Social Sciences*, *9*(5), 1–17. https://doi.org/10.3390/SOCSCI9050080
Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*(Proceedings. International Geographical Union. Commission on Quantitative Methods), 234. https://doi.org/10.2307/143141
UN-Habitat. (2003). *Guide to Monitoring Target 11 : Improving the lives of 100 million slum dwellers*. (May), 1–18.
UN-Habitat. (2015). *Slum Almanac 2015/2016: Tracking Improvent in the Lives of Slum Dwellers*. Retrieved from https://unhabitat.org/slum-almanac-2015-2016
UN-Habitat. (2018). *Metadata on SDGs Indicator 11.1.1 Indicator category: Tier I*. Nairobi.
UNDP. (2012). Climate Risks, Vulnerability and Governance in Kenya: A review. *Undp*, 83. Retrieved from http://www.preventionweb.net/files/globalplatform/entry_bg_paper~keynaclimaterisksvulnerabilityandgovernanceinkenyaareviewiisdundpjan13.pdf
UNECE. (2021). The Industrial Accidents Convention and Natural Disasters: Natech. Retrieved August 5, 2021, from Areas of Work website: https://unece.org/industrial-accidents-convention-and-natural-disasters-natech
UNEP. (1992). Agenda 21. *United Nations Conference on Environment & Development Rio de Janerio, Brazil, 3 to 14 June 1992*, *6*(June), 351. Rio de Janerio: United Nations.
UNEP. (2018, December 21). Smoking Nairobi landfill jeopardizes schoolchildren's future. Retrieved June 15, 2021, from Cities and lifestyles website: https://www.unep.org/news-and-stories/story/smoking-nairobi-landfill-jeopardizes-schoolchildrens-future
UNEP. (2020). Cities and climate change . Retrieved October 28, 2020, from Resource efficiency website: https://www.unenvironment.org/explore-topics/resource-efficiency/what-we-do/cities/cities-and-climate-change
UNFPA. (2007). *The State of the World Population 2007: Unleashing the potential of urban growth*.
United Nations. (2017). New Urban Agenda. *Habitat III: United Nations Conference on Housing and Sustainable Urban Development*, 1–66. Retrieved from www.habitat3.org
United Nations. (2019a). The sustainable development goals report 2019. *United Nations Publication Issued by the Department of Economic and Social Affairs*, 64.
United Nations. (2019b). *World Urbanization Prospects: The 2018 Revision*. https://doi.org/10.4054/demres.2005.12.9
USGS. (n.d.). Runoff: Surface and Overland Water Runoff. Retrieved July 30, 2021, from The USGS Water Science School: Learn About Surface Water website: https://www.usgs.gov/special-topic/water-science-school/science/runoff-surface-and-overland-water-runoff?qt-science_center_objects=0#qt-science_center_objects
Vanhuysse, S., Georganos, S., Kuffer, M., Grippa, T., Lennert, M., & Wolff, E. (2021). Gridded Urban Deprivation Probability from Open Optical Imagery and Dual-Pol Sar Data. *IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium*.
Vaughn, L. M., & Jacquez, F. (2020). Participatory Research Methods – Choice Points in the Research Process. *Journal of Participatory Research Methods*. https://doi.org/10.35844/001c.13244
Wang, J., Kuffer, M., Sliuzas, R., & Kohli, D. (2019). The exposure of slums to high temperature: Morphology-based local scale thermal patterns. *Science of The Total Environment*, *650*, 1805–1817. https://doi.org/10.1016/j.scitotenv.2018.09.324
Wang, S., Wang, J., Fang, C., & Feng, K. (2019). Inequalities in carbon intensity in China: A multi-scalar and multi-mechanism analysis. *Applied Energy*, *254*, 113720. https://doi.org/10.1016/j.apenergy.2019.113720
WHO. (2021a). Health Impact. Retrieved June 16, 2021, from Air quality and health website: https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts
WHO. (2021b). Type of Pollutants. Retrieved June 16, 2021, from Air quality and health website: https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants
Windfinder. (2021, June). Monthly wind direction and strength distribution. Retrieved August 5, 2021, from Wind & weather statistics Nairobi/Jomo Kenyatta website: https://www.windfinder.com/windstatistics/nairobi-jomo-kenyatta-airport
World Health Organization (WHO). (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. Geneva.
World Health Organization (WHO). (2018, May 2). Ambient (outdoor) air pollution. Retrieved May 18, 2021, from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health
Zhou, D., Zhao, S., Zhang, L., Sun, G., & Liu, Y. (2015). The footprint of urban heat island effect in China. *Scientific Reports*, *5*(1), 11160. https://doi.org/10.1038/srep11160

8. ANNEX

8.1. Research Design Matrix and Process

Using a case study approach, our research integrates participatory principles and combines qualitative and quantitative methods to answer the research questions. These are summarized in the research design matrix below.

S.O.nr.	Research Question	Data Requirement	Participatory Principle	Method
1	Which hazards are deprived areas predisposed to? Which open geospatial data can be used as hazard indicators?	Key informant interviews (local experts and field experts) Informal meetings with research groups	Consultation	Literature review Database search
2	Are deprived areas more likely to be located in hazardous areas in relative comparison to formal settlements? What share of deprived areas are located in hazard-prone areas? Can a multi-hazard dataset be used to predict deprivation? How do multi-hazard datasets compare to textural features in the prediction of deprivation?	Geo-spatial and EO data Label data (training and testing samples)	-	GIS Spatial analysis Spatial statistical analysis Descriptive statistics Multivariate analysis of variance (MANOVA) Exploratory analysis (Discriminant analysis) Machine learning (RFC)
3	How are hazards spread within a deprived settlement?	Household survey	Consultation and Inclusion	Descriptive statistics

8.2. EM-DAT Nairobi's Recorded Disasters (2009-2019)

Source:	EM-DAT, CRED / UCLouvain, Brussels, Belgium www.emdat.be (D. Guha-Sapir)
Version:	2021-04-21
File creation:	Wed, 21 Apr 2021 11:05:23 CEST

Disaster Group	Disaster Subgroup	Disaster Type	Disaster Subtype
Natural	Biological	Epidemic	Bacterial disease
Natural	Biological	Epidemic	Bacterial disease
Natural	Biological	Epidemic	Bacterial disease
Natural	Hydrological	Flood	Riverine flood
Natural	Hydrological	Flood	Riverine flood
Natural	Hydrological	Flood	Riverine flood
Natural	Hydrological	Flood	Riverine flood
Natural	Hydrological	Flood	
Natural	Hydrological	Flood	
Technological	Technological	Industrial accident	Explosion
Technological	Technological	Industrial accident	Collapse
Technological	Technological	Miscellaneous accident	Fire
Technological	Technological	Miscellaneous accident	Collapse
Technological	Technological	Miscellaneous accident	Fire
Technological	Technological	Miscellaneous accident	Other
Technological	Technological	Miscellaneous accident	Collapse
Technological	Technological	Miscellaneous accident	Fire
Technological	Technological	Transport accident	Road
Technological	Technological	Transport accident	Road
Technological	Technological	Transport accident	Road

8.3. Key Informant Interview Questions

Interview Questions

Interviewer Introduction

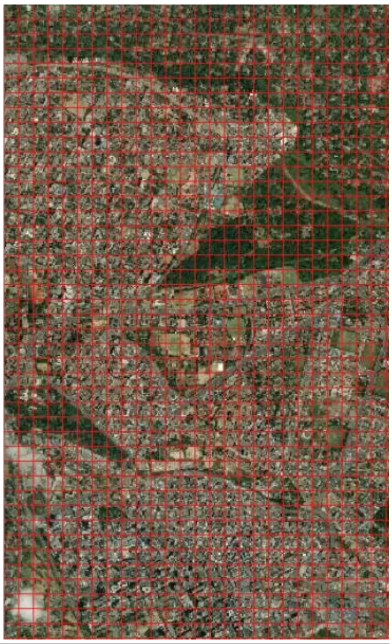
I am a Masters student at the University of Twente, Faculty of Geo-information science and earth observation, undertaking a course in Urban Planning and Management. As part of my course work, I'm conducting research on the mapping of informal settlements using Machine Learning. Particularly my research aims to analyse the relationship of hazardous locations and deprivation (informal settlements/urban poverty).

Respondent

Name and Title/Designation:

Part 1: General Questions on "slums"

1. What is your role in slum planning, management or interventions?
2. In your work, how do you define a slum?
3. What information (geographic/spatial & non-spatial) about slums do you use in your work?
4. What actors are involved in the collection and use of slum data/information?
5. What is the frequency of collecting and updating these data?
6. For which areas is this information collected?
7. How do you use this information?
8. Are there any missing data about slums you think would be helpful to your work?
9. Which level of spatial detail do you use/require for your work?

Administrative units (wards)	Street Blocks	Grids
		

Part 2: Hazards

10. Which hazards affect Nairobi?
11. To your understanding, are the effects of these hazards evenly spread?
12. If not, what are some of the conditions or characteristics that make the spread uneven?
13. Does location and permanency of housing have an effect on perceived (or otherwise) safety and security from adverse weather/climatic/natural disasters and related effects?

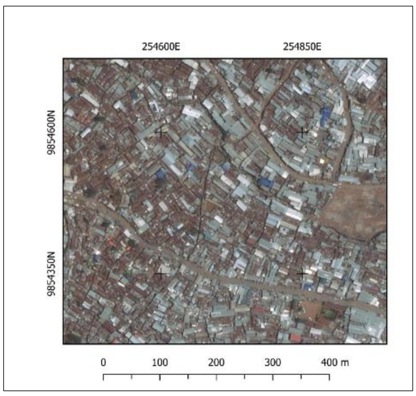
Part 3: Durable Housing

UN-Habitat lists durable housing as a domain in the definition of slums. It captures two main aspects: structural quality of housing and **location** and structural quality of the housing and **permanency** of the structure.

14. Is the quality of housing important in defining settlements?
15. What is your general perception of the location of slums and the type of structures present?
16. What definition of durable housing do you use/ are you familiar with?
17. Do you use this domain in your work on slums? If so, how?
18. What data do you use to capture "durable housing"?
19. What data would you recommend to be used to capture the aspect of durable housing?
20. How do you collect and use these data?
21. On a scale, how would you rank the following types of housing (include others) in terms of durability (tent, pavement dwellings, slums, informal settlements, formally planned settlements, shacks, old Swahili housing)?
22. What do you take into consideration when making the ranking?

Part 4: Ethical concerns

23. What information about slums do you consider sensitive in your work?
24. Who should have access to slum information, and to what extent?
25. What concerns do to have with regards to the use of technology such as AI in the analysis of deprivation?
26. Do you have any "sensitivity" to the use of high resolution (e.g.10m resolution) and very high resolution (e.g. 1.5m resolution) satellite imagery in the analysis of deprivation (Helber et al., 2018)?
    If yes, what are they, and what are probable ways of handling them?



Part 5: Respondent's privacy concerns

How would you like research results and information communicated to you?
Which of the following do you give consent to use in our research? (Designation/Title and Organization

8.4. Household Questionnaire

Priscilla MSc. Research

Please record the date of conducting this survey.

Date of data collection

yyyy-mm-dd

hh:mm

Select your name (enumerator's name)

Select settlement of data collection

- Kibera
- Kariobangi North

We are working with a University called University of Twente. We know that people in Kibera/Kariobangi North face several challenges related to hazards (madhara). We would like to ask you some information about yourself and your household.
Some of the questions may seem personal and we'd like to guarantee that your information will be kept safe and anonymized. It is also not compulsory to answer all questions, if you don't feel comfortable to. Introduction

The survey will take approximately 20-30 minutes. Do you consent to being

- **interviewed?** Yes
- No

Ask for respondent's name.

Age

How old is the respondent?

Sex/Gender

Gender that the respondent identifies with

- Female
- Male
- Other
- Prefer not to answer

If other, are you comfortable telling us which one?

Can you tell us how many people live in this house?

Household size

What is the highest level of education completed by any of the household members?

Anyone in the household

- ○ Primary
- ○ Secondary
- ○ Tertiary
- ○ None of the above

Why did you (or your family) choose to live in this settlement (Kibera/Kariobangi North)?

Ni kwa sababu gani uliamua kuchagua makaazi haya?

- ☐ It's our family home (ushago)
- ☐ Close to opportunities
- ☐ Rent is affordable
- ☐ Close to services
- ☐ Other

If other, please specify

For how long have you lived in this house?

How long they have lived in the structure where the interview is being conducted.

- ○ 0-1years
- ○ 1-2years
- ○ 2-5years
- ○ 5 years and above

Do you pay rent for the house?

- ○ Yes
- ○ No

How much do you pay in rent per month?

- Below 1000 kes
- 1000 - 2000 kes
- 2000 - 3000 kes
- 3000 - 4000 kes
- 4000 - 5000 kes
- 5000 and above

If you don't pay rent, could you please describe for us the living arrangement of this house.

- Landlord (I own the structure)
- Landlord (the structure belongs to my family)
- Tenant (I have an agreement with the owner)
- Other

If other, please describe.

How do you pay for rent?
Kama hulipi kodi ya nyumba kwa pesa, unatumia njia ipi?

As we mentioned in the introduction, we are interested in understanding hazards. The following questions will therefore be related to hazards that affect this community and/or your household.

Maswali yanayofuata ni kuhusu madhara yanayoathiri kijiji/jamii yako

---

Do any of the following hazards affect your community/neighborhood?
Kati ya haya, ni madhara gani ambayo kijiji lako lina pitia?

- [ ] Floods (*mafuriko*)
- [ ] Fire (*moto*)
- [ ] Extreme heat (*joto jingi*)
- [ ] Extreme cold (*baridi mingi*)
- [ ] Disease outbreaks e.g. cholera (*magonjwa ya mlipuko*)
- [ ] Building collapse (*maporomoko ya manyumba*)
- [ ] Road (motor vehicle and boda boda) accidents (*ajali za barabarani*)
- [ ] Rail accidents (*ajali za reli*)
- [ ] Garbage accumulation (*mkusanyiko wa takataka*)
- [ ] Air pollution (uchafuzi wa hewa)
- [ ] Industrial hazards - pollution (air, water, land/soil contamination, noise) (*madhara kutokana na sekta ya viwanda kama vile uchafuzi wa hewa, maji na ardhi*)
- [ ] Industrial hazards - explosions (*milipuko kutokona na sekta ya viwanda*)
- [ ] Others (*mengineo*)

If others, please specify.

If the settlement is affected by flooding, what is the cause of the flooding?
Sababu za mafuriko

- [ ] Closeness to river
- [ ] Insufficient drainage channels
- [ ] Blocked drainage channels
- [ ] Terrain
- [ ] Other

If others, please specify.

If the settlement is affected by fire, what is the cause of the fire?
Sababu za kuchomeka/moto

- [ ] Poor power (electricity) connection
- [ ] Cooking accidents
- [ ] Lighting accidents
- [ ] Burning to evict people - to set up new structures
- [ ] Industrial accidents e.g. explosions
- [ ] Other

If others, please specify.

If the settlement is affected by air pollution, what is the cause of the pollution?

Sababu za uchafuzi wa hewa

- [ ] Industries
- [ ] Transport - roads, rail, airport
- [ ] Burning garbage
- [ ] Traditional cooking means e.g. firewood, charcoal
- [ ] Other

If others, please specify.

If the settlement is affected by disease outbreak, what is the cause?

Sababu za magonjwa ya mlipuko kama vile kipindupindu, COVID-19 etc.

- [ ] Poor sanitation and hygiene at individual level
- [ ] Poor sanitary environment at community level e.g. garbage accumulation
- [ ] Poor water drainage system
- [ ] Contamination of water/food
- [ ] Burst sewerage pipes
- [ ] Unprotected toilets
- [ ] Others

If others, please specify.

If the settlement is affected by road related accidents, what is the cause?

Sababu za ajali za barabara

- [ ] Close distance to transport lines
- [ ] Insufficient transport network (e.g. roads)
- [ ] Insufficient/lack of pedestrian crossing
- [ ] Others

If others, please specify.

If the settlement is affected by rail related accidents, what is the cause?
Sababu za ajali za reli

- [ ] Close distance to transport lines
- [ ] Insufficient transport network (e.g. roads)
- [ ] Insufficient/lack of pedestrian crossing
- [ ] Others

If others, please specify.

If the settlement is affected by collapse of buildings, what is the cause?
Sababu za manyumba kuporomoka

- [ ] Poor construction techniques
- [ ] Poor construction material
- [ ] Unstable ground e.g. swamp
- [ ] Malicious demolitions
- [ ] Others

If others, please specify.

Individual/ Household level
Maswali yanayofuata ni kuhusu madhara yanyo kuathiri nyumbani kwako

---

Do you think that the house you are living in adequately protects you from the mentioned hazards?
Nyumba yako inakupa usalama wa kutosha kutokana na madhara ambayo tumeyaongolea?

- ○ Yes
- ○ No

If your house adequately protects you from hazards, please explain how.
Tuelezee jinsi nyumba yako inakinga tosha kutokana na madhara tunayoyaongelea

Which of the above mentioned hazards has ever affected your household?
Ni madhara gani kati ya haya ambayo yanakuathiri/ yanathiri nyumba yako?

- [ ] Floods (*mafuriko*)
- [ ] Fire (*moto*)
- [ ] Extreme heat (*joto jingi*)
- [ ] Extreme cold (*baridi mingi*)
- [ ] Disease outbreaks e.g. cholera (*magonjwa ya mlipuko e.g. kipindupindu*)
- [ ] Building collapse (*maporomoko ya manyumba*)
- [ ] Road (motor vehicle and boda boda) accidents (*ajali za barabarani*)
- [ ] Rail accidents (*ajali za reli*)
- [ ] Garbage accumulation (*mkusanyiko wa takataka*)
- [ ] Industrial hazards - pollution (air, water, land/soil contamination, noise) (*madhara kutokana na sekta ya viwanda kama vile uchafuzi wa hewa, maji na ardhi*)
- [ ] Industrial accidents - explosions (*milipuko kutokana na sekta ya viwanda*)
- [ ] Not affected at all (*sihathiriwi hata kidogo*)
- [ ] Others (*mengineo*)

If others, please specify.

Why do you think your household was affected by floods?
Mbona nyumba yako iliathiriwa/inaathiriwa na mafuriko?

- [ ] Poor housing structure - roof
- [ ] Poor housing structure - walls
- [ ] Poor housing structure - floor
- [ ] Location of structure
- [ ] Other

If other, please specify

Why do you think your household was affected by fire?
Kwa nini nyumba yako iliathiriwa/inaathiriwa na moto?

- [ ] Type of housing structure - roof
- [ ] Type of housing structure - walls
- [ ] Type of housing structure - floor
- [ ] Location of structure
- [ ] Other

If other, please specify

Why do you think your household was affected by extreme heat?
Kwa nini nyumba yako iliathiriwa/inaathiriwa na joto jingi?

- [ ] Type of housing structure - roof
- [ ] Type of housing structure - walls
- [ ] Type of housing structure - floor
- [ ] Location of structure
- [ ] Other

If other, please specify

Why do you think your household was affected by extreme cold?
Kwa nini nyumba yako iliathiriwa/inathiriwa na baridi nyinigi?

- [ ] Type of housing structure - roof
- [ ] Type of housing structure - walls
- [ ] Type of housing structure - floor
- [ ] Location of structure
- [ ] Other

If other, please specify

Why do you think your household was affected by building collapse?
Kwa nini nyumba yako iliathiriwa/inaathiriwa na maporomoko?

- [ ] Type of housing structure - roof
- [ ] Type of housing structure - walls
- [ ] Type of housing structure - floor
- [ ] Location of structure
- [ ] Other

If other, please specify

To help us understand better the exposure to hazards, we will ask you a couple of more questions about your household. Do you agree to this?
Ili tuweze kupata maarifa ya kutosha, tunataka kukuuliza maswali mbili zaidi kuhusu nyumbani kwako, ni sawa?

- Yes
- No

Which of the following does your household rely on for lighting?
Unatumia aina gani ya taa (ya kuonea)?

- Electricity (stima)
- Kerosene lamp (taa ya mafuta)
- Candles (mishumaa)
- Other

If other, please specify.

Which of the following does your household rely on for cooking?
Unategemea aina gani ya moto kupikia?

- Charcoal (makaa)
- Firewood (kuni)
- Kerosene stove (mafuta ya taa)
- Gas e.g. meko
- Electricity (stima) e.g. coil
- Other

If other, please specify

This marks the end of the household questions. Thank the respondent for their time and ask them if they would be interested to know the outcome of the research.
Tunakushukuru kwa muda wako. Ungeweza kupata taarifa kuhusu utafiti huu?

- Yes
- No

If yes, which mode of communication would they like?
Ungependa kupata taarifa kuhusu utafiti huu kwa njia gani?

- [ ] Posters
- [ ] Booklet
- [ ] Community Workshop
- [ ] Newspaper article
- [ ] Whatsapp link (to a website)
- [ ] Social media e.g. Facebook, Instagram, Twitter
- [ ] Others

If other, please specify.

Which language would you like the communication to be in?

Ungependa taarifa kuhusu utafiti huu uenezwe kwako kutumia lugha gani?

- ○ English
- ○ Kiswahili
- ○ Sheng
- ○ Other

If other, please specify.

In the following sections, kindly record the structure characteristics by observation. This might take some time, therefore inform the respondent that you will be there for a couple of minutes more.

Rekodi aina ya nyumba ambayo umefanya utafiti.

Material used for roofing

- ○ Iron sheet
- ○ Makuti
- ○ Carboard
- ○ Tiles
- ○ Other

If other, please specify

Internal roofing modifications i.e. material for ceiling.

- Polythene ceiling
- Wooden ceiling
- Carboard ceiling
- None

Material used for walls

- Mud
- Mud with concrete plastering
- Iron sheets
- Bricks (matofali)
- Stone
- Wood
- Other

If other, please specify

____

Material used for flooring

- Bare ground
- Bare ground covered with polythene carpet
- Concrete plastering
- Wood
- Other

If other, please specify

____

Kindly record the geo-coordinates of the structure with an accuracy of below 10m

GPS ifikishe chini ya mita 10

latitude (x.y °)

longitude (x.y °)

altitude (m)

accuracy (m)

Point and shoot! Use the camera to take a photo

Take photo of the general neighborhood. Be careful not to take a photo of anyone.

Click here to upload file. (< 5MB)

If consent for being interviewed is not given, thank the respondent and terminate survey.

8.5. Codes used

Process	Platform/IDE	Link
Data collection	Google Earth Engine	https://github.com/MappingMojo/LULC-GLCM-RFC
GLCM feature extraction	R Studio	
Land Cover Classification		
Land Use Classification		