Searching for the built environment

Clustering built environment typologies to find spatial patterns and areas of deprivation using remote sensing techniques.

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

by

Stephan Olde

Student number: 4288548

To be defended in public on February 27th 2023

Graduation committee

Chairperson	:	Dr.ir. S. van Cranenburgh, dept T&L
First Supervisor	:	Dr.ir. T. Verma, dept PA
Second Supervisor	:	Dr. N.Y. Aydin, dept SE&S

Abstract
Cities are experiencing rapid urbanization and with it face the issues it provides. More people are moving into cities and they need to live somewhere. This leads to housing situations with poor living conditions. In order to address this issue, policy makers need information on where people live and in what kind of conditions the people live.

This research uses high resolution satellite images in combination with an unsupervised Convolutional Neural Network Autoencoder to identify features that can be used to cluster different built environment typologies. Previous remote sensing research uses ground truth data which for some areas is not available or needs manually labeled training data. This research attempts to circumvent the issue of information scarcity in order to create a methodology that can be applied on any city as long as satellite images are available. From the resulting clusters, clusters can be selected which represent areas with high levels of deprivation which in turn can help identifying the deprived areas.

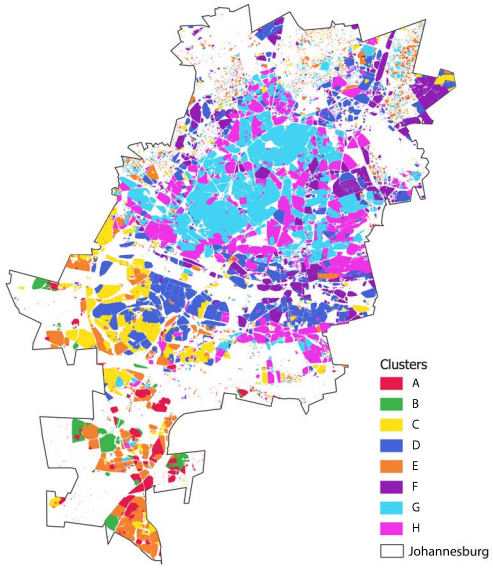

Fig. 1. Overview of the found clusters within the city of Johannesburg.

In order to test this methodology, a case study was performed for the City of Johannesburg which historically has faced issues of racial and spatial segregation. The study showed that the unsupervised clustering methodology was able to identify multiple built environment typologies, including clusters showing deprived areas. The resulting clusters are shown in the figure and show spatial patters of different clusters throughout the city. The results were evaluated using three methods: a visual analysis, spatial analysis and using census data. The visual analysis of the clusters showed that from the eight clusters, five distinct built environment typologies could be identified ranging from informal types of built environment to areas that are highly developed. The spatial analysis and census data confirmed the differences between these typologies but also highlighted that one of the clusters was a mix of both ends of spectrum of the level of development of the built environment

and was not usable. The research showed that it is feasible to use unsupervised machine learning techniques to cluster different built environment types but that human interpretation is required to contextualize the results.

Acknowledgements

This thesis marks the end of nearly 10 years of studying at Delft University of Technology. After starting my journey at this university with a year of Mechanical Engineering and a year of Maritime Engineering, I finally found myself at the faculty of Technology, Policy and Management where I was able to show my talents in a succesfull way and get a bachelors degree. From there it was clear that I wanted to pursue the masters degree of Engineering and Policy Analysis where I felt welcome in the EPA family from the moment I started.

Even though this EPA journey started during lockdowns and remote lectures, I could connect with my fellow students who got me through the courses and helped to distract from the state of the world. When everything went back to normal, we could finally meet each other in real life and friendships could be capitalized on. I want to thank all of my fellow students of supporting each other in these difficult times.

I want to thank my supervisors for providing valueable feedback on my work in this project and helping me get through the journey of what people call a thesis project. In special regards I want to thank Trivik in being understanding in the projects we did together, from guiding my bachelor thesis and now my master thesis as well. You helped me get through my struggles and helped me focus on the parts of my thesis that I neglected as well as reminded to take care of myself outside of my thesis as well.

Furthermore, I would like to thank my friends and family for being there when I needed them and letting me find my own way in this 10 year long journey that has now come to an end.

I hope, you, as a reader enjoy this thesis and be reminded that hard work pays off however long the journey may be.

Table of Contents

List of Figures

List of Tables

1 Introduction

In this section the topic of this research will be introduced together with its scientific and societal relevance. Furthermore, this section will be used to present the structure of the research in this thesis.

1.1 Problem Situation

In 2018, 55% of the world population lived in urban areas and this percentage is expected to grow to 68% in 2050 with most of this increase happening in Asia and Africa (United Nations, 2019). On these continents the population in cities grows between 2.6% and 4.1% annually (Onodugo & Ezeadichie, 2019). With this rapid urbanization, cities struggle to accommodate the growing population, resulting in growing inequalities faced by urban residents (United Nations, 2019). These inequalities often relate to access to basic infrastructure, such as healthcare and education, as well as adequate housing (Abascal et al., 2022). As a result, a growing number of people in rapid urbanizing cities experience poor living conditions and negative health outcomes (United Nations Department of Economic and Social Affairs Population Division, 2019). Estimates by UN Habitat (2016) suggest that approximately one billion people live in slums and informal settlements. The rapid urbanization becomes visible in the forming of new urban peripheries on the outskirts of the existing urban areas (Inostroza, 2017). The formation of informal settlements is difficult to track as they form and develop outside of the reach of the government and are not always included in official data sources (Dos Santos et al., 2017).

In December 2016, the United Nations General Assembly endorsed the New Urban Agenda which shares a common vision on urbanization and a sustainable world. The New Urban Agenda calls for sustainable development of cities along the lines of the Sustainable Development Goals (SDGs) set in the United Nations their 2030 Agenda for Sustainable Development (United Nations, 2016). The relevant SDG for this thesis is SDG 11 which is focused on creating cities that are resilient, safe, sustainable and inclusive for all their citizens (Nations, 2022). In order to track the progress of SDG 11 data is required to monitor the effects of policies in cities around the world. However, there are problems with the current methods of measuring SDG indicators in growing cities such as the data being outdated, inaccurate, or incomplete (Musango, Currie, Smit, & Kovacic, 2020).

The main type of data source used by the UN to track indicators related to the SDGs is a census, which is conducted in many countries around the world (Division, 2022). However, census data does not properly reflect the situation of residents in informal settlements and slums. This is caused by these residents being underrepresented because there is little to no official registration of their residency (Satterthwaite, Sverdlik, & Brown, 2019). Furthermore, some regions might not be identified in the data as informal settlements, which causes these regions to be missed in the results. Moreover, census data has a tendency to focus on the deficits and problems of informal settlements rather than their strengths and potential (Friesen, Taubenböck, Wurm, & Pelz, 2018). Alternatively, some countries have other data available on a regional rather than national level such as small-scale surveys or administrative data like land use records and utility connections (Roy, 2005). A problem with these kinds of data is that this data is not readily available in national databases which makes it difficult to find and access the data. Another problem with local surveys and administrative data is that the indicators will be different for each region which makes it harder to compare cities to each other (Madhavan, Beguy, Clark, & Kabiru, 2018). As a result, traditional data sources do not adequately represent informal settlements and may be misleading in their reporting of these areas.

Remote sensing has emerged as a promising alternative data source for detecting and characterizing informal settlements, particularly in hard-to-reach, dangerous or remote areas (Gevaert et al., 2016; Gram-Hansen et al., 2019; Gränzig et al., 2021). Remote sensing data, such as satellite imagery, can be used to identify and map the spatial distribution and characteristics of informal settlements and can provide a more accurate and up-to-date picture of these areas than traditional data sources (Gram-Hansen et al., 2019). Remote sensing methodologies feature a number of different tools and techniques to provide information on informal settlements such as image classification, feature extraction, object detection and texture analysis (Mahabir et al., 2018). These methods rely on distinguishing visual features that are present in informal settlements compared to other residential areas, making them detectable from satellite images (Gram-Hansen et al., 2019).

However, current remote sensing methodologies have certain limitations in terms of inputs and outputs. One limitation in terms of inputs is that most algorithms are fully supervised, requiring manual labeling of a train and test dataset for training (Plazas, Ramos-Pollán, & Martínez, 2021). This can be a labor-intensive process that must be repeated for each new city, as there can be variations in how informal settlements are visually represented between cities (Hofmann et al., 2015). Manual labeling may also introduce a bias of what constitutes an informal settlement, whereas in practice, there may be more nuance surrounding informal settlements (Kit et al., 2012). For example, some informal settlements may be more organized and developed, while others may be more spontaneous and less developed. In terms of output, many algorithms tend to provide a binary classification of urban areas, categorizing them as either formal or informal, without providing additional context or insights into the spatial structure of the city (Mahabir et al., 2018). Therefore, in this research we will use an adaptation of an unsupervised image classification methodology designed by Singleton et al. (2022a) which classifies houses and their surrounding areas in a city based on satellite imagery without the need for manually labelling data. This adaptation allows for the use of images with a higher resolution which can capture more nuanced differences in the images. Moreover, this algorithm produces a number of classes based on the visual elements which, in our expectation, will include one or more classes of informal settlements as well as a range of formal types of settlements. As the classification algorithm will only provide a number of nameless classes without any context, we will take these classes and contextualize them according to a framework developed by Dovey and Kamalipour (2017) which classifies the built environment on a two axis scale; architectural and infrastructure. This will result in a ranking of the found classes from informal to formal on both axis and helps us to increase our understanding of the found classes. The resulting classes within their spatial context can help with increasing our understanding of the spatial structure of a city and how different built environments are distributed over the city.

1.2 Research Objectives

This research will answer the main research question: *what can be learned from inferring built environment types from satellite images using unsupervised machine learning algorithms?* By answering this question a number of aspects will be explored. Firstly, whether a satellite view of the built environment is sufficiently similar for similar neighbourhood types and a big enough difference between different neighbourhood types to differentiate between different neighbourhood types. Secondly, by answering this question spatial structures within a city can be explored and help with identifying possible spatial segregation between different types of built environments. Thirdly, by exploring the built environment types identified by the classification algorithm, this methodology might allow us to identify areas with high levels of deprivation.

The goal of this research is to explore the capabilities of remote sensing techniques on differentiating different built environments in a meaningful way. To achieve this, a case study of the greater City of Johannesburg in South Africa will be performed. For this city we will gather the data, apply the algorithms and explore different classes. The classes will be ranked based on their architectural and infrastructural appearance as to add context to the nameless classes formed by the machine learning algorithm. This will help to see to what extent the built environments differ from each other. After this, census data will be linked to the found classes that can serve as a proxy for determining whether deprivation is present in any of the classes. This evaluation will help confirm if this research alligns with the assumption that deprivation of areas reflects in the built enviroment as is suggested in literature (Ellena et al., 2020; Molina-García et al., 2017). In the process of answering the main research question, the following set of sub-questions will be answered first:

- *What are the challenges of measuring deprivation according to literature?* This question will explore the literature on deprivation and how this can be evaluated . This will help in understanding the challenges that currently exist in this field.
- *How has remote sensing been used in existing literature to characterize the built environment?* This question will explore what kind of remote sensing methods have been used in literature to characterize the built environment. This can provide insights in different approaches to achieve similar results as well as highlight limitations of past research.
- *Which built environment typologies can be found in the City of Johannesburg using unsupervised remote sensing techniques?* By answering this question the found classes will be explored and given some context in terms of what can be visually explored from the images in our analysis which will help highlight differences between the found classes.
- *What is the spatial distribution of built environments in the City of Johannesburg?* By answering this question, the resulting clusters will be linked to the real world context where they come from. To achieve this we will study the spatial aspect of the resulting clusters. Moreover, the resulting clusters will be linked to existing data about gated communities, informal settlements and census data to further describe the clusters.

1.3 Research Flow and outline

In this chapter, we have explored the topics of remote sensing and the built environment and how we will use unsupervised remote sensing techniques to identify different built environment typologies with the goal of finding spatial structures and with that identifying informal settlements within a rapid urbanizing city. In Chapter 2 we will perform a literature review on what elements form the built environment of residential buildings as well as explore different methodologies that use remote sensing techniques to characterize or classify different built environments. In Chapter 3 we will explore our methodology to classify and contextualize the built environment as well as introduce our case study area which is the City of Johannesburg in South Africa. In Chapter 4 the results of the classification as well as the contextualization of the found classes will be presented. Moreover, in this chapter the resulting classes and their spatial characteristics will be compared to real world data as validation of the unsupervised algorithm. In Chapter 5, the results regarding the case study area will be discussed as well as the used methodology including the found limitations. Chapter 6 presents the conclusion of this research as well as recommendations and paths for future research.

2 Literature review

In this chapter we will explore literature on urban deprivation and on how remote sensing has been used to identify areas subject to deprivation. This exploration is performed in order to find knowledge gaps that remain in literature.

2.1 Deprivation in areas of rapid urbanization

In rapid urbanizing areas one of the key challenges is the issue of deprivation, which is the lack of access to resources and opportunities to achieve a minimal standard of living (Mitlin & Satterthwaite, 2013). Deprivation can have a lot of consequences for people like physical and mental health problems as well as social isolation and reduced life expectancy (Wilkinson et al., 2009). One of the key factors causing deprivation is the lack of affordable housing which has multiple effects like overcrowding and substandard living conditions (D'Alessandro & Appolloni, 2020). Moreover, due to urbanization marginalized communities often get displaced which causes social exclusion and isolation (McGranahan et al., 2016). Another factor contributing to deprivation is the lack of access to basic amenities like healthcare, education and transportation (Palmer et al., 2019). For people that are already vulnerable and marginalized these kinds of factors can stack up further disadvantaging them (Keeley et al., 2019). For example, no access to education can reduce employment options which results in a lower standard of living. Similarly, having no access to healthcare can reduce a person's health and thereby reduce life expectancy. Therefor, it is important to get a grip on deprivation and find ways to improve living conditions for people living in deprived areas.

2.2 Lack of information

The United Nations urges local institutions to try and limit deprivation in rapid urbanizing areas by building low-cost housing and basic infrastructures (Habitat, 2016). As not having interventions that limit the deprivation of areas will lead to massive growth of deprived areas and put more people in a disadvantaged position (Baud et al., 2009). However, in order to implement effective measures that address these challenges, these areas need to be identified and characterized first (dos Santos et al., 2022). This calls for the collection of specific and exhaustive data on the geographic locations, built environments, demographic compositions, and socioeconomic dynamics of these areas (Mitlin & Satterthwaite, 2012). Because a lot of deprivation develops in an informal and unregistered setting little is known in terms of statistical data about the form, size and other characteristics of these areas (Myers, 2021). Because of the information deficit, it is difficult to form adequate policy to limit deprivation in urbanizing areas.

Despite some challenges in identifying and characterizing deprived areas, there are some information sources that support current research and policy making. These sources of information include: census data, community engagement, expert knowledge and remote sensing. Census data is a reliable source of information but is usually only performed once every ten years and does not reflect recent changes as is occuring in areas subject to rapid urbanizing areas (Samper et al., 2020). Another source of information that can be useful in the process of collecting data is community engagement as it allows researchers to gather information from locals and get their perspective of the situation in the area they live in (Goodman & Gatward, 2008). However, these local surveys are costly and time consuming. Moreover, if this data is gathered using a sample of the residents from an area, the results might not reflect the views and experiences of all the members of the community (Baatiema

et al., 2013). A third option to gather data on local deprivation is the use of expert knowledge as it can help understand the complex and multifaced nature of deprivation in their areas of expertise (Bessell, 2015). But, expert may dissagree with each other and their views can be based on empirical evidence which can make it difficult to verify information that is provided by them (Cabrera-Barona & Ghorbanzadeh, 2018). A different way of gathering information is by the use of remote sensing as it allows for the collection of information on physical conditions and infrastructure without the need for fieldwork (Sahriman et al., 2013). However, remote sensing also has limiting factors as it is limited to errors and biases in the used training data (Giri, 2012). Despite the challenges with these methods for finding areas subject to deprevation, the aformentioned methods do provide insights into urban areas and all methods can help increase the understanding of these areas.

2.2.1 Challenges in Measuring Deprivation Another problem which limits researching on deprivation are the measurements methods available. For example, Habitat (2016) has five criteria that are measures of deprivation; lack of access to improved water source, lack of access to improved sanitation facilities, lack of sufficient living area, lack of housing durability and lack of security of tenure. Any house lacking one of these measures is categorized as deprived. As this measure is based on households increases the difficulty of determining research as it requires information on a household basis. As levels of deprivation can differ within a single neighbourhood it can be quite difficult to capture nuanced levels of deprivation using aggregated data sources (Myers, 2011). Therefor, misrepresentation of the actual situation occurs quite easily as it is resource intensive to gather information on a household level of detail.

Another global measure for identifying the level of deprivation is the Human Development Index which measures a country's level of development in three dimensions: health, education and standard of living (Alkire, 2007). It is a measure developed by the United Nations to allow better comparison between countries. However, it can also be calculated on a more local scale as long as information on that scale is available. For example, if one region has a significantly lower Human Development Index score compared to the national average or other regions within the same country, it can be considered more deprived in terms of the Human Developemnt Index (Khalifa & Connelly, 2009). The index can be used to identify areas where improvements in terms of health, education and living standards are needed and to track the progress of these indicators over time (Lai, 2000) . Similar to other measures, the Human Development Index on a local scale requires non-aggregated data as well as being resource intensive to gather the required data. Moreover, it is important to note that the Human Development Index is a broad measure that does not capture all of the aspects that entail deprivation because of the complex nature of deprivation (Alkire & Robles, 2017).

2.3 Remote sensing techniques

The use of remote sensing has grown over the last couple of years as satellite images have become more widely available as well as computing power being more accessible. The advantages of remote sensing over traditional data gathering techniques is that it is less time consuming and more affordable (Klasen, 2000). Remote sensing also has the benefit that it is spatially explicit and can be performed at high levels of geographical accuracy (Arribas-Bel et al., 2017). One of the main approaches to measuring deprivation is using remote sensing to detect slums or informal settlements (Kuffer et al., 2016).

Literature on slum detection and deprivation mapping has been explored and is presented in Table 1. From literature it showed that there are four main methods used for slum detection which are feature extraction, texture based approaches, pixel based approaches and contour detection. In feature extraction, the algorithm is trained to generated features that resemble the input data, this can be predetermined features like statistics from census data but also less interpretable features. Pixel based approaches classify take a sattelite image and classify the land use of each pixel which then will create a land use map of an area. Contour detection methods train to identify the edges of slum areas and accurately predict exact areas that are slums. The last method is texture based which takes a similar approach to pixel based approaches but on a larger scale where the texture of an image is analysed over a larger area and the relation between the different pixels will determine the output of the classification. The majority of the research used high resolution satellite images of less than a meter per pixel resolution for their methodology because of the detail that is captured within the higher resolution. Another thing which stood out from the research is that most research uses supervised machine learning techniques where the data will be manually labeled or a different type of ground truth data is used. The research that used unsupervised methods generated features on the available input images which are then clustered based on similarity of the features. Current research is diverse in the used algorithms which include neural networks, support vector machines and random forests.

Table 1. Remote sensing research related to deprivation mapping and slum detection

Author	Topic	Input data	Method	Machine learning
(Stark et al., 2020)	Slum mapping	High resolution images	Texture	Supervised
(Kit & L"udeke, 2013)	Slum mapping	High resolution images	Feature extraction	Supervised
(Leonita et al., 2018)	Slum mapping	Multispectral satellite images	Feature extraction	Supervised
(Prabhu et al., 2021)	Slum mapping	High resolution images	Feature extraction	Supervised
(Ajami et al., 2019)	Deprivation mapping	High resolution images	Feature extraction	Supervised
(Williams et al., 2020)	Slum detection	High resolution images	Feature extraction	Supervised
(Ghaffarian & Emtehani, 2021)	Slum detection	Satellite images	Feature extraction	Supervised
(Owen & Wong, 2013)	Slum detection	Satellite images and morphology data	Texture	Supervised
(Hofmann et al., 2001)	Informal settlements detection	Satellite images	Feature extraction	Supervised
(Ansari & Buddhiraju, 2019)	Slum detection	High resolution images	Contour model	Supervised
(Ibrahim et al., 2019)	Land cover	Street typology	Feature extraction	Unsupervised
(St Amand, 2014)	Slum detection	Satellite images	Texture	Unsupervised
(Arribas-Bel et al., 2017)	Deprivation mapping	High resolution images	Pixel based	Supervised

One common method for detecting slums is the use of ground truth data, which can be the locations of known slums or a specific set of predetermined features. These features usually include physical characteristics of the built environment, such as roof materials, building size and building density, or census-based statistics, such as population density and poverty rates. The use of a ground truth allows for the comparison and validation of the used algorithm, and helps to ensure the accuracy and reliability of the results. However, there are limitations using a ground truth for training algorithms as it requires the availability and up-to-date data on these locations which is not always available (Mahabir et al., 2018). Moreover, using a predetermined set of features may lead to a limited set of

generated features that does not capture the full complexity and diversity of slums (Leonita et al., 2018). One way to circumvent this issue is to use manually label images as being slums which can be a tedious and time consuming procedure which allows room for error. Another way to circumvent this issue is to use unsupervised machine learning methods which, up to now have been limited in their application.

Another issue found within current research is that it focuses a lot on binary slum detection, an area is either a slum or not. This makes for a harsh cut off in terms of classification and might not capture the diversity that is present in the appearance and structure of slum (Kuffer et al., 2016). In order to solve this problem methods could be used that classify on broader scale like the research by Arribas-Bel et al. that predicts an index score of deprivation rather than the more common binary classification. However, methods like these still require ground truth data to function.

2.4 Knowledge gaps

In this chapter we have explored the topics of deprivation and remote sensing in order to increase our understanding of the current state of literature on these topics. We have seen that deprivation is a complex topic that traditionally is explored by analyzing socio-economic indices from census data or local surveys which has the problem that it is a costly and time consuming endevour to collect this data. Because of these problems, socio-economic data is scarce in a lot of countries. Also, a lot of this data is aggregated to administrative units which makes it harder to identify specific areas that face problems of deprivation. In order to solve the problem of data scarcity, remote sensing techniques have been developed which allow for a more detailed, cheaper and more frequent generation of data on deprived areas. This has mainly been performed using binary slum detection techniques. Research with more nuanced classification of urban areas is limited and mostly focuses on high level classification of areas like an area being residential or industrial. Therefor, further research could focus on developing a methodology that has a low data requirement and provides a broader classification of the urban built environment which is suitable for finding areas with high levels of deprivation.

Current research has the tendency to focus on developing algorithm that try to achieve perfect prediction scores on whether an area is deprived or not. However, reality is more messy and nuanced which requires a more nuanced type of classification. In order to achieve this together with a low data requirement restraint, we will use an unsupervised machine learning algorithm together with an unsupervised clustering algorithm. The inputs for these algorithms are high resolution satellite images and remotely sensed building location data. The outputs of these algorithms are multiple clusters which each represents a different type of built environments. The resulting clusters are collections of satellite images without context. As this methodology is to gain insights in areas with limited information availability we will evaluate the clusters using three different methods that all have weaknesses but combined can help to understand the clusters. These methods are manual visual inspection and classification of a sample set of each cluster, location analysis of the clusters and comparison with available data on the extremes of built environment development, and a socio-economic analysis for the found clusters. The details for each of these steps and methods can be found in Chapter 3 and the results are presented in Chapter 4.

3 Research Design

In this chapter an overview of the design of this research will be presented. First an overview of the methodology will be presented followed by an exploration of the method used for the analysis and clustering of the built environment using unsupervised machine learning methods. Next, we will explore the process of contextualizing the classes found by the clustering analysis. After this, the process for evaluation the resulting will be explained. Finally, we will explore the case for our case study area which is the City of Johannesburg. An visual overview of the methodology is found in Figure 2.

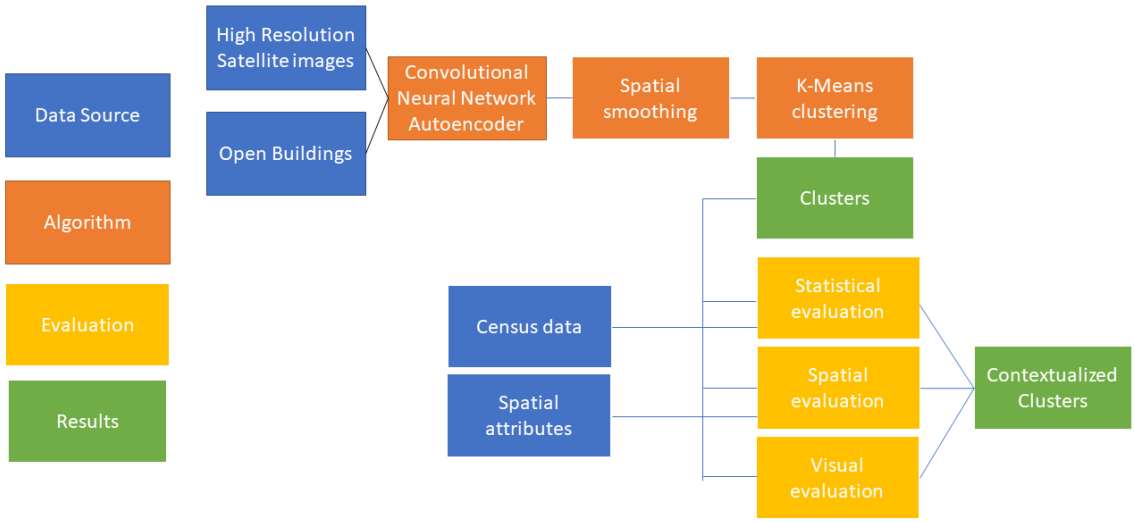

Fig. 2. Overview methodology

3.1 Clustering of the built environment

The process of classifying the built environment consists of three parts; A convolutional autoencoder,a smoothing algorithm and a k-means clustering algorithm. This design is an adaptation Singleton et al.'s (2022a) research on estimating local neighbourhood context using convolution neural networks and satellite images. This method is used because of the low data requirements by only requiring satellite images and the location of buildings, which are both available globally and can be acquired from public data sources. The data sources will be described in further detail in Chapter 3.1.1. Moreover, this method is unsupervised which means that the algorithms are not trained using labelled data but rather uses all the input images to train on and derive the most important features representing their latent representation. A latent representation is a simple representation of the input while retaining information on its most important features. This has the

benefit that this process can be adjusted relatively easily to be applied on any urban area in the world and take into account it's local built environment as the algorithm trains itself on the input data for that specific region.

As mentioned before, the clustering process consists of three main parts which link together in a linear fashion. The first set of input data is a collection of satellite images that together cover an area of interest. The second input dataset consists of the location of all buildings in the same area of interest. These two data sources are then combined to create an individual image for each of the buildings in the dataset, including an area around each of the buildings as to include the built environment in the images. All of the images will be used to train the convolutional autoencoder which can be split into two parts itself, the encoder and the decoder. The encoder part generates a latent representation of the images which the decoder uses to restore the original image. The result of the decoder is not used for further analysis but can be used to check to what extent data is retained in the latent representation. After training the autoencoder, the latent representation is generated for each of the input images. The resulting latent representations are then smoothed where a building is given the average feature values of the buildings within a set radius. This improves similarity within a neighbourhood and helps to remove some of the outliers generated by the autoencoder. These smoothed values are then clustered using a k-means clustering algorithms which is an unsupervised clustering algorithm. The k-means algorithm is run for different k-values from which the optimal k-value is selected which is the final number of classes for the data. After these steps, each of the input data values has an assigned class which is then ready for interpretation and analysis.

3.1.1 Data sources There are two data sources for the clustering of the built environment; high resolution satellite images and a dataset containing building footprints. For both datatypes multiple data sources have been considered which resulted in using high resolution satellite images from the GIS company ESRI and an open source dataset from Microsoft containing remotely sensed building footprints.

There are numerous providers of satellites imagery but most have the problem that they require paid licences to use making the process costly. The world's leading high resolution satellite imagery provider is Maxar Technologies where most companies providing satellite images to the public have license agreements with. Because of our limited research budget we had to choose an image provider that allows for the use of their imagery for our research. We ended up using ESRI's World Imagery layer as they allow the use of their images for scientific research (ESRI, 2022) and the data is formatted in a way which is designed for the use in GIS software. Other public satellite image providers like Google Earth or Bing Maps do not allow for easy downloading of the images or add watermarks to their high resolution imagery. The free image providers like ESA's Sentinel projects or USGS EarthExplorer do not provide for high resolution images which also makes them unfit for this research. Therefor we used the he World Imagery layer is an image layer with global coverage with the highest resolution being 1.9 cm per pixel up to 1.2 meter per pixel resolution for the rest of the world. For most metropolitan areas the image resolution available is 0.3 meter per pixel which works with our methodology which uses 1.0 meter per pixel resolution.

As data source for the building footprint multiple options were considered; Google's Open Buildings project (Google, 2022) and Microsoft's Global ML Building Footprint project (Microsoft, 2022) which both are projects that provide building footprint maps with a large global coverage. As

Google's Open Buildings project is centered around Africa and Asia, we have decided to use Microsoft's dataset as it has a larger global coverage. However, in future cases both projects should be considered as viable sources as the projects keep growing in their coverage. These datasets are generated using deep learning algorithms and therefor are not 100% accurate which is no problem as for our research we are not interested in the individual buildings but their location. This means that errors in house shapes or rotations will not affect our results. Also, we are interested in patterns within cities rather than evaluating individual houses which means that if the data is missing a few houses per neighbourhood it will not really affect the results. Moreover, these datasets have a confidence score for each of the buildings which allows to limit the dataset to only include buildings with high confidence scores.

3.1.2 Pre processing In order to be able to use a machine learning algorithm we need to have input data that is in a useable format. By putting together the locations of the building footprints and the satellite images, a single image, or tensor, can be made for each building in the building dataset. Effectively, this means that the sattelite images covering the area of interest are transformed into a large number of smaller images. The process of transforming satellite image and building footprints to extracted images is shown in Figure 3. The building outline shapes are first converted to single points before being converted to a square box which is the cutout pattern for the resulting images/ As can be seen, the data contains areas where houses are close to one another. This causes a lot of overlap in the resulting images, which increases the training data for similar areas, which will improve feature generation later in the process.

The preprocessing is performed using GPU-accelerated Python packages to increase performance compared to CPU based packages because of the number and size of the images that are processed. The resolution of the images as well as the number of images that can be processed is limited by both CPU RAM and GPU RAM as datasets can get very large. As an example, the satellite images for the City of Johannesburg which covers and area of 1,645 km2 is 20 gb with a resolution of 1 meter per pixel and 220 gb with a resolution of 30 cm per pixel. This needs to be loaded completely into GPU RAM in order to be processed together with leaving enough RAM for the extracted individual images. As a result of these limitations, the data was divided into multiple parts whih after transformation can be joined again as to have a complete file for the later steps in the process. The extracted images for 3 million buildings totaled 41 gb in compressed form. Thus, for the preprocessing one needs to make sure to have enough RAM available in both CPU and GPU and this step might require some trial and error in the segmentation of the input data to make sure that the compute units can process it.

3.1.3 Convolutional Neural Network Autoencoder A convolutional neural network (CNN) Autoencoder is a type of artificial neural network that is used for dimensionality reduction and feature learning (Hinton & Salakhutdinov, 2006). The goal of autoencoders is to re-create the original data that was given to them with the availability of fewer dimensions or features. CNN autoencoders are particularly useful for image data as they can learn hierarchical representations of the input data by taking advantage of the spatial structure inherent in images (Vincent et al., 2010). This means that it can detect patterns in satellite images like road parts being connected to other road parts or that if there is a tree it is likely that there are more trees.

CNN autoencoders function by first encoding the input data into a lower-dimensional latent space through the use of convolutional and pooling layers. The network's encoder component has been

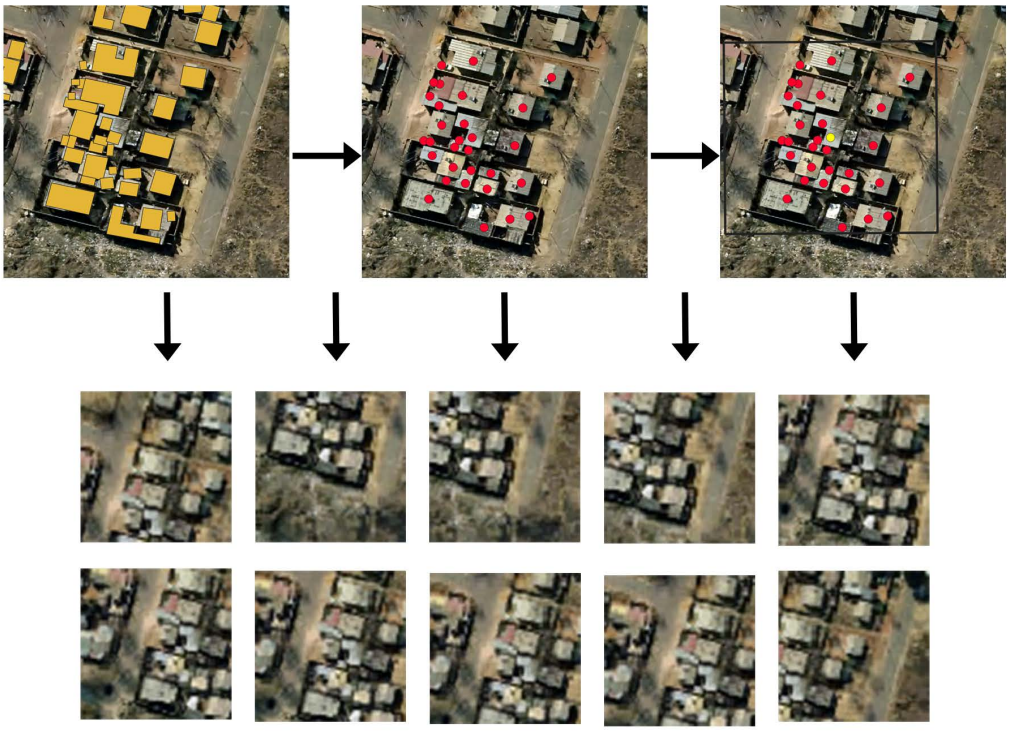

Fig. 3. Example of generated image tensors. Buildings from the Open Buildings dataset are converted to centroids after which an image of 64 meters by 64 meters is extracted from a satellite image for each of the centroids.

trained to maintain the most crucial information while mapping input data to this latent space. The network's decoder component is subsequently trained to recreate the input data from the latent space. The goal is to create a compressed representation of the data that retains the most crucial features while removing the less important ones. For this research it means that this can be leveraged to perform a clustering algorithm on the most important features of images rather than on the original images itself. These original images are 64 by 64 pixels by 3 color layers equaling 12.288 values. With the use of the autoencoder the number of values is reduced to 256 values which equates to 2% of the original values. Another reason why the use of an autoencoder works in this research is because it represent the patterns in the images rather than just color values making images that are similar but not exactly the same still comparable in terms of their latent representation.

Convolutional neural network Autoencoders have several types of layers like convolutional layers, pooling layers and fully connected layers. Convolutional layers are designed to learn local patterns in the input data, pooling layers are designed to reduce the spatial resolution of the feature maps produced by the convolutional layers, and fully connected layers are designed to learn global patterns in the data (Krizhevsky, Sutskever, & Hinton, 2017; LeCun, Bengio, & Hinton, 2015). Convolutional

layers are particularly useful for image data because they can learn features that are translation invariant, meaning that they are not sensitive to the exact position of the features in the input. Pooling layers help to reduce the number of parameters in the network which makes it more robust to small changes in the input data. Fully connected layers are used to learn higher-level features from the output of the convolutional and pooling layers. These three layer types are supported by a number of less complex layers like an input layer which initializes and load the input data, normalizing layers which scales the values of to previous layer to values between 0 and 1 (Goodfellow, Bengio, & Courville, 2016) and upsampling layers which doubles the size of the data by repeating the initial values (Ketkar & Santana, 2017). These layers are not connected like the other types of layers and are there to facilitate the connected layers to work properly. The combination all these types of layers make for a range of applications and including different types of patterns, local and global, small scale and larger scale to be represented in the resulting features in the latent representation.

3.1.4 Parameters An important part of designing a neural network are the hyperparameters which in contrast to the nodes in a neural network are user defined and not learned by training (Bengio, 2012). Hyperparameters exist in two types; layer parameters and model parameters. Both types of hyperparameters affect the complexity of the model and the ability of the model to learn to perform a specific task. As the names suggest, layer parameters affect an individual layer in the network and model parameters affect the running and optimization of the network as a whole. Layer parameters in our model for convolutional layers include the number of filters, kernel size, stride and padding type. While there are numerous different model parameters available and hyperparameter tuning can be a study by itself, this research considered the default model parameters optimizer, metrics and loss which are passed to the model when compiling. A more complete overview of model hyperparameters can be found in Appendix A.

Most layer types have specific hyperparameters that can be configured and tuned for optimal performance. However, as will be shown in the section on model implementation, the model used in this research a combination of convolutional layers, batch normalization layers and upsampling layers. Batch normalization and upsampling layers are layer types that are regularly used as with their default parameters (Szegedy et al., 2016). However, they can be configured to work along a certain axis or use different methods of handling interpolation. In the model used in this research the default parameters were used for the normalization and upsampling layers as these have limited influence on model performance. Convolutional layers are more complex in terms of usage as these layers are locally connected nodes and changes the dimensions of the data in the network (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012). In our model, the hyperparameters filters, kernel size, strides and padding were set for convolutional layers. The kernel size controls the size of the filters used to learn these features, while the number of filters specifies how many different features the layer may learn (Albawi, Mohammed, & Al-Zawi, 2017). Before passing through the convolutional layer, the input data is padded according to a set padding type. Another type of parameter is the activation parameter is a mathematical function applied to the output of a layer and helps to introduce non-linearity allowing it to learn complex relationships in the data (Goodfellow et al., 2016). All-in-all, there are a lot of considerations to make regarding parameters when designing layers in a neural network. However, when designing layers, it relatively easy to configure and make it work, but the hard part is to optimize it (Géron, 2022).

Hyperparameters for the model will determine for what purpose the neural network will be optimized, how fast it will learn the task at hand and what metrics it is evaluating to improve the performance of the model. The optimizer specifies the algorithm which will be used by the model to update its weights and biases (Kingma & Ba, 2014). Experts on neural nets are still divided over which optimizer is the best overall. For example, some optimizers might outperform others in terms of convergence speed and generalization performance but perform worse in terms of overfitting and underfitting (Kingma & Ba, 2014). Nevertheless, some optimizers like Adam, SGD and Adamax all have similar performance for a lot of tasks and it is difficult to predict up front which one will perform better for the task at hand (Gulli & Pal, 2017). The metrics parameter specifies how the performance of the model is tracked over time and can measure the ability of the model to correctly predict values. A different but related parameter is the loss function which measures the difference between the ground truth and the prediction of the neural network (Manaswi, 2018). Both the metrics and loss parameters influence how well the neural net will train and perform over time. It is important to carefully choose the appropriate loss function and metrics for a specific task, like classification or autoencoding, as they directly infuence the model's performance.

3.1.5 Implementation The objective of the CNN Autoencoder in this research is to create latent representations of the input images that have a high retention of information from the original information but in a more compact form. As the first part of this research is an adaptation of Singleton et al. (2022a), the code of that research which was published on a GitHub repository was used as a starting point for our own code. As Singleton et al.'s (2022a) research revolved around classifying residential types for a complete country including rural and urban areas, the design of the CNN Autoencoder needed some adjustments as to be able to distinguish different typologies within an urban context of a single city. This is because we assume that there are smaller differences within the built environments of a single city compared to that of a country after some initial tests using the same CNN Autoencoder that was used by Singleton et al.. As a result, the resolution of the input images was increased together with the size of the images as to keep a similar area covered within each image. These adjustments were limited by the availability of satellite images and the available computing power needed to process large quantities of data. After some experimentation, which is addressed in Appendix A, the resulting input data has a resolution of 1 meter per pixel with a real world size of 64 meter by 64 meter. Because of the limited availability of high resolution satellite images with coverage of multiple color bandwidths, the neural net was designed to work within the red, green and blue (RGB) color space. Thus, the resulting neural network is inspired by the work of (Singleton et al., 2022a) with adaptation to increase the amount of information given to the algorithm.

The resulting CNN Autoencoder design is shown in Figure 4 and shows the encoder and decoder part of the neural net and all of the layers within the network. A full design of the network including hyperparameters can be found in Appendix A. The neural network is designed according to a few common practices in the design of CNN Autoencoders which are the use of a symmetrical architecture, the use of regularization techniques, the use of upsampling techniques and the use of transposed convolutional layers (Hinton & Salakhutdinov, 2006). The symmetrical design ensures that there is a balance between the encoder and decoder which makes sure that both can work properly together. The normalization layers act as way to regulate the values within the network and help with overfitting problems. The upsampling layers layers in the decoder help through

interpolation, to increase the resolution of the layers which in steps helps to get back to the resolution of the original image.

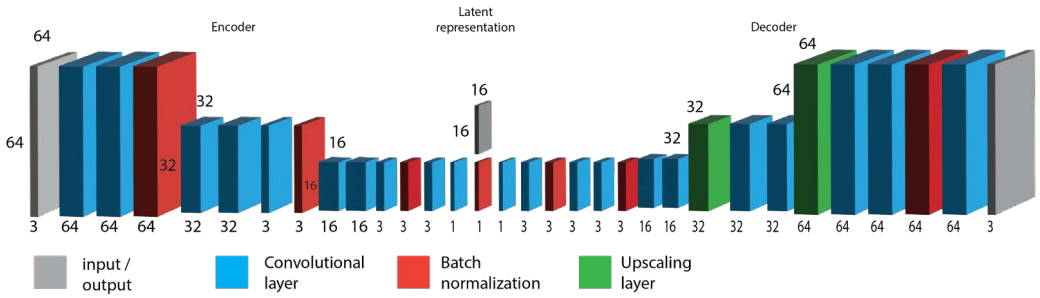

Fig. 4. Design of the Convolutional Neural Network Autoencoder used for generating latent representations of the built environment. The numbers represent the dimensions of all of the layers. The design is nearly symmetrical in its design so the decoder is similar to the encoder part but in reverse. The design shows how a latent representation of the input can be extracted from the center of the algorithm with a resolution of 16 by 16 values.

One of the import design decisions of this neural network is the stepped design of the convolutional layers where in steps the size of the data is reduced. Because this gradually reduces the spatial resolution of the input data as it passes through the network, the computational complexity of the model is reduces and allows the model to focus on higher level features (Krizhevsky et al., 2017). Another aspect of the stepped design is to gradually change the number of filters as the input data passes through the network, which stimulates the model to learn increasingly complex features (LeCun et al., 2015). The stepped design also allows the model to learn features at different scales which is beneficial to capturing different levels of patterns (Szegedy et al., 2016). There is one design feature that is implemented which is more common practice than strictly necessary, namely, working with powers of 2. The reason why this is used in this neural network is because it makes scaling the network easier as well as that it allows for the use of equal padding and strides which simplify the network (Krizhevsky, Hinton, et al., 2009). The design of the layers and their parameters has been done in order to achieve a stepped design to gradually decrease and increase the resolution of the data and capture patterns at different scales.

One thing that might stand out when seeing this design is the lack of pooling and fully connected layers which have been mentioned earlier as being important layer types which together with the convolutional layers make a Convolutional Neural Network. Although pooling layers are not present in the neural network, pooling does occur. The pooling is performed by some of the the convolutional layers where the stride parameter is larger than one and the kernel size is smaller than the input data. The difference between using pooling layers versus convolution layers with pooling parameters is that convolutional layers have learnable parameters while pure pooling layers do not have that ability (Krizhevsky et al., 2017).

The design of the neural network also does not include fully connected layers which is because fully connected layers do not have spatial invariance (Huang, Liu, Van Der Maaten, & Weinberger, 2017). Spatial invariance is the ability of a model to recognize features regardless of its position in

the input data. Therefor, for image data, it is important to preserve spatial information in order to recognize objects or features regardless of their position (Chen, Zhu, Papandreou, Schroff, & Adam, 2018). Convolutional layers, because of their sliding window approach do allow the network to learn local patterns regardless of their location in the input data. Because our input data is satellite images that are not all rotated the same way and can contain buildings on any place in the image, it is important to have convolutional layers with the spatial invariance rather than the fully connected layers.

For this research, the model hyperparameters were not explored further than in simple model exploration. This is because the used parameters are used frequently in these types of CNN Autoencoders. The used optimizer is Adam (Kingma & Ba, 2014), the used metric is accuracy and the loss function is mean squared error. With the Adam optimizer, the model's weights and biases are updated using adaptive learning rates which has been shown to work well on a wide range of tasks in the field of deep learning (Kingma & Ba, 2014). The metric accuracy is used because it is a metric that evaluates the correct predictions made by the model divided by the total number of predictions. In the context of a CNN autoencoder for images, accuracy measures how well the model can reconstruct the input image (Brownlee, 2019). The loss function mean squared error is a function that measures the difference between the input and reconstructed data by calculating the mean squared error. Research shows that Mean Squared Error (MSE) loss function in the context of image autoencoders generally perform better than other loss functions (Krizhevsky et al., 2017; Glorot, Bordes, & Bengio, 2011). This does mean that in some specific cases another loss function may result in better performance of the autencoder but for this research the MSE loss function should function well enough. The three model hyperparameters used in this neural network all are widely used within the field of image CNN Autoencoders and hyperparameter tuning could probably increase performance of the model one or two percent but also would require a lot of trial and error.

The CNN Autoencoder was designed with the limitation of a cloud computing instance with a Tesla P100 GPU with CUDA 11.2, 16GB VRAM, 32GB RAM. The most restrictive part of the process is the preprocessing where the VRAM and regular RAM are utilized to their max and where the computers crash when the limits are exceeded. A solution to these RAM limits is working with segmented data but it does complicate the process. This is where the trade-off was made to use a resolution of 1 meter per pixel and limit the area to 64 meter by 64 meter. Other aspects like the complexity of the neural network were no issue as convergence of the network happened between around 12 hours of training which basically meant that running it overnight would be sufficient when working on it the next day. Increasing the amount of training data might increase the time per epoch but it will not create training times that are outside the scope of reason.

3.1.6 K-means clustering The latent representation that was created using the CNN autoencoder is not open to human interpetation because it is multi-dimensional, contains values between 0 and 1 and it is near impossible for a human to find out what a generated feature represents. However, we know that the latent representation is a reflection of the input image it represents and similar images will have similar features in their latent form. Because we know there are patterns in similar built environments and that a number of those patterns are captured in the latent representation of the dataset, the dataset can be clustered. A common method for the unsupervised clustering of high dimensional data is K-Means clustering (Guo, Liu, Zhu, & Yin, 2017). The K-

Means has proven to be successful in a multitude of unsupervised deep learning tasks including ones with autoencoders (Aljalbout et al., 2018).

The K-Means algorithm is an unsupervised clustering algorithm which clusters data in a dataset in a predefined number of clusters. In order to gain some understanding of how the K-Means algorithm operates, a 2 dimensional example will be visualized and explained, whereas the actual data processed in this research contains 256 dimensions. The algorithm starts by randomly selecting K centroids, which it assumes are the centers of the clusters as shown in Figure 5. Based on the Euclidean distance between each data point and the centroid, the algorithm then assigns each data point to the closest centroid. After assigning all the data points to a cluster, the centroid of each cluster is set to be the mean of all the data points assigned to their cluster. This process is repeated until the centroids stop moving and the clusters are converged. This algorithm works the same for larger amounts of clusters as well as more datapoints and dimensions which makes the complexity scale linearly which makes the algorithm quite efficient.

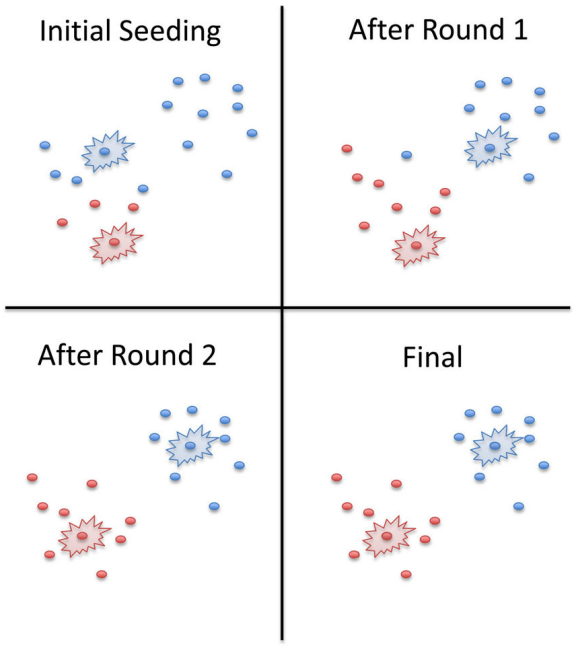

Fig. 5. Example of a Kmeans clustering algorithm showing 3 iterations of finding the optimal centers. The actual algorithm will keep refitting the centers until the change in distance scores gets under a certain threshold or a set number of itterations have been performed. Source: Page et al. (2014)

Because of the convergence process, the final clusters are very sensitive to the random initial position of the centroids. To mitigate this issue, the K-means algorithm needs to be run with different starting conditions multiple times. After all the runs, the final cluster centroids with the lowest sum of squared errors (SSE), can also be referred to as the within cluster sum of errors (WCSS), are chosen and clusters are assigned to the data points. The SSE calculated according to the following

formula: $SSE = (datapoint - centroid)^2$. So the SSE is the distance between the centroids and the data points assigned to the clusters squared as to mitigate negative values. The K-Means run with the lowest SSE will be chosen as the optimal result and used to assign each data point a cluster.

However, as this research uses unsupervised methods to identify patterns and context in the built environment, the number of resulting clusters is unknown to us beforhand. In order to find an optimal number of clusters, k-Means needs to be run for a range of K values. Afterwards statistical analyis can be used to help and find the best value for K. The most common methods to identify the best value of K are the elbow approach or the silhouette score method, which are subjective methods to obtain an optimal K value and thus might not be suitable for all usecases (Pelleg, Moore, et al., 2000). In cases with large, complex datasets it is good practice to pick the method for selecting a K value based on what kinds of clusters are wanted. In our case this means that we select based on what we want our clusters to reflect. As we want to have clusters that are most similar within themselves we have opted to select the cluster with the lowest average WCSS meaning that the points within each cluster are the closest to each other (Charrad et al., 2014). Compared to other methods, this might mean that there are k values with a lower total WCSS score for all clusters combined but that the cluster that we select are most compact, a risk with this method is that in the final result there is a chance of outlier clusters. These potential outlier classes can also be valuable in this research as it will help understand the diversity of built environment types. Thus, the number of clusters or K value we will select for the rest of our research will be determined by analyzing with which K value the clusters are most compact.

3.1.7 Smoothing There is a lot of variation in the built environment surrounding houses, whether this is variation in the location of a house within a neighbourhood or if a car is parked on the street. Every variation influences the latent representation generated for these houses in quite a big way, hence clustering might not be as consistent in assigning the same cluster to neighbouring houses. This can be caused by the CNN autoencoder not having enough layers, not having enough training data or not being optimized enough. As this problem also occurred with the research of Singleton et al. (2022a) (2022a), they introduced a way of improving homogeneity within neighbourhoods by implementing a spatial average on the latent representation.

This method of smoothing the resulting latent representations is based on the assumption that neighboring buildings will have a similar built environment. For all the buildings in the dataset their latent representation was averaged with the latent representations of the buildings within a 124m circular radius. The choice of this 128m radius is that it is four times the distance of the area that is captured by the input images in the CNN Autoencoder, which is a 32m radius, but this distance, it could, in reasonable range, also be a bit higher or lower. This helps to smooth out outliers within neighbourhoods and increase homogeneity of the results (Anselin, 1995). The result of this smoothing procedure is shown in Figure 6 where clustering results before and after smoothing are shown. As there is a significant increase in cluster coherency, the clustering algorithm is performed after all latent representations have been smoothed spatially.

3.2 Typologies

As the resulting clusters from the CNN Autoencoder and the K-means clustering algorithm do not provide a lot of insights themselves, apart from a spatial distribution of these clusters throughout the researched area. For this research, the decision was made to add this context manually for all

(a) Unsmoothed cluster result sample (b) Smoothed cluster result sample

Fig. 6. Effect of the smoothing operation on the uniformity in neighbourhood classifications. Each color represents one of eight clusters. There is no relation between the colors of clusters of both images as they are the results of different k-means clustering calculations with the same settings but the unsmoothed input data versus smoothed input data.

of the found clusters. To achieve this, a representative sample of images for each of the clusters will be ranked based on the architecture (buildings) and the urban design (streets and block layout) in the images. This is done in line with the research of Dovey and Kamalipour (2017) they show how images of neighborhoods can be ranked from informal to formal in each of the two categories. In their research they applied these classifications to multiple cities in the Global South, including Johannesburg, as to show that this classification method can be applied on any city. This will help with interpreting the clusters and increase comparability between clusters and, in further research, this could allow tracking the development of informal settlements in multi-temporal research. Moreover, it can provide a lens for analyzing and understanding the different morphologies that can exist within a given urban area.

The classification process involves categorizing urban areas based on the formality of architecture (buildings) and urban design (street and block layout). Formality refers to the degree in which the architecture and urban design of an area conform to regulations (Marshall, 2009). These regulations may include zoning regulations, building codes, or other regulations governing the design and construction of buildings and the layout of streets and neighborhoods. Formal urban areas are often planned and developed top-down, with a distinct hierarchy of control and a set of rules that govern the physical environment in terms of layout and appearance (Rios, 2014). Informality is characterized by self-organization and micro-spatial adaptation to the environment (Gouverneur, 2014). Informal settlements often emerge in an incremental and bottom-up manner where buildings and streets are modified over time to include the needs and preferences of the residents (Robinson, 2013). Another sign of informality is the use of temporary and makeshift material as well as the lack of carefully laid out streets (Roy, 2011). Thus, by analyzing the level of (in)formality of a cluster, we can gain insights into the development of these areas.

As built environments are not just informal or formal and it can be a mix of both in the architecture and urban design, the clusters will be placed into a typology grid developed by (Dovey & Kamalipour, 2017) and is shown in Figure 7. This grid represents the nine possible combinations of

formal and informal architecture and urban design, ranging from informal settlements (i/i) in the lower left to formal settlements (F/F) in the upper right. Intermediate categories include mixed formal/informal settlements (F/i or i/F), informal mixed formal settlements (i/m or m/i), and formal mixed formal settlements (F/m or m/F). By assigning each of the clusters to one of the types it can provide insights into the various morphologies present in the analyzed city. Moreover, it will help to identify the trends and patterns within the city and where formality and informality intersect.

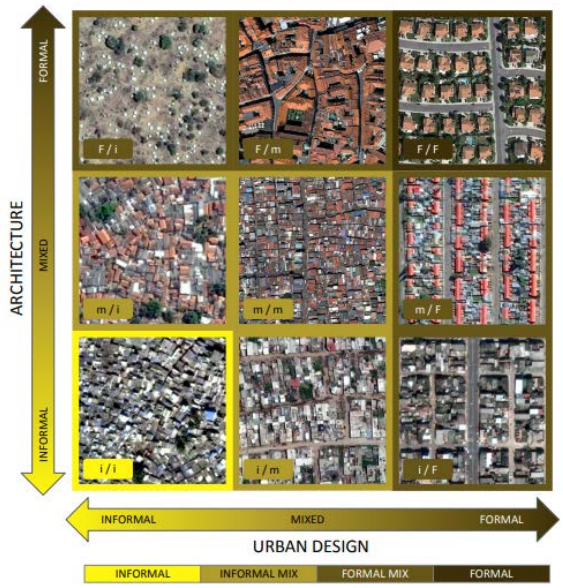

Fig. 7. Different built environments typologies on two axis: architectural and urban design. Source: Dovey et al. (2017)

In order to be able to analyze the clusters and assign a typology to a cluster, 50 random sample images for each of the clusters will be extracted using GIS software. In order to be able to assess the built environment rather than just the building or the 64 meters that was used as an input for the CNN Autoencoder, a 200 meter by 200 meter image will be extracted for each of the random samples to include a proper amount context. This set of images will then be classified by assigning one of the typologies described above. For reference, the selected images and their typology are shown in Appendix B and delineated per cluster. The typology classifications per cluster will then be aggregated in order to reach a final typology for each of the clusters.

3.2.1 Census evaluation In order to evaluate the results of the clustering algorithm and the assigned typologies, statistics for each of the clusters will be calculated by using available census data. There will be an analysis of five variables which each will serve as a proxy variable for one of the measures of deprivation as defined by the UN Habitat programme (Habitat, 2016)

The first step in this evaluation process is to gather socio-economic data for the area of interest. To improve the results, it is desirable to use data with the highest available resolution thus each entry covering a small area. As this research analyzes areas with low information availability this step can be performed with socio-economic information of any resolution but will work better with data on more detailed data.

The second step is to match the spatial location of each of the houses with the area they reside in. This is done by assigning each location the statistics of the place they belong to. Using this method, socio-economic profiles can be made for each of the clusters by aggregating the data of all of the buildings that have socio-economic variables assigned. After this step, the aggregate data is analyzed to identify trends and patterns within each cluster.

Through this process of linking census data to the clusters, real-world context is added to the built environment typologies that are found in the earlier steps of this research. This will increase the understanding of the found clusters and the differences between the different built environment typologies.

3.3 Case Study

In this section the case will be presented on which we will apply the remote sensing techniques presented above.

3.3.1 Case study description The case chosen for this research is the City of Johannesburg Metropolitan Municipality, South Africa, or Johannesburg for short. Johannesburg is the biggest metropolitan area of South Africa with a population of 5.5 million residents (Joburg Municipality, 2022). Johannesburg is located in the northern part of the country and spans 80 kilometers in length and 50 kilometers in width totaling an area of 1644 square kilometers. The city is divided into 7 administrative regions and these regions are subdivided into a total of 135 wards and 804 subplaces (Joburg Municipality, 2022). The wards each have their own ward councilor and ward committees that are elected to represent them in the municipality and help improve their wards (Joburg Municipality, 2022). The subplaces are smaller areas within wards and are administrative areas used for statistical analysis.

The choice for the City of Johannesburg as case is because of a number of reasons. Firstly, Johannesburg has been growing 3% annually over the last 50 years meaning the city is facing urbanization challenges. Secondly, Johannesburg is known for the South African apartheid regime from 1948 to 1990 which means that there is a history of segregation rooted within the city that still exists today. Therefore, Johannesburg has contrasting morphological typologies from luxurious houses in gated communities to shacks in informal settlements which have relatively big visual differences which should be beneficial for our methodology. Lastly, the city and its province have public GIS portals that allow for easy access to data in a geographical form. However, these data sources are limited to the information that is available like census data where the most recent version is from 2011. The combination of rapid urbanization and historically rooted segregation within the city make for Johannesburg to be a suitable city to study within this research.

During the apartheid Johannesburg had distinct residential areas for people with white and black skin colors. Development of the residential areas were designed in order to enforce the boundaries

between different types of neighbourhoods. This created an urban tissue where both natural elements like rivers and man made elements like roads were used to create separation (Bähr & Jürgens, 2006). After the apartheid ended, these natural barriers remained and their effects still affect the separation between neighbourhoods to this day. However, not only physical barriers remained as social barriers remained as well. Both types of barriers have resulted in where the city is today as we will discuss in the following paragraphs.

The urban structure of Johannesburg is a complex one with both planned and unplanned urban development occurring at the same time (Totaforti, 2020). Moreover, the city's urban development was affected heavily by the apartheid policies to segregate different population groups as well as the post-apartheids policy to try and solve previous segregation rules. The city is challenged by a fragmented social and urban structure and will need a lot of time to bring the fragments together (Bremner, 2000). To get some more insights in the city we will discuss the important economic areas of the city as well as the residential extremes in the form of informal settlements and gated communities.

One of the city's most important places was the central business district which was located centrally in the city and due to urban development became the economic centre of the city with large commercial and residential areas. During the apartheid, only white people were allowed to live in the central business district but near the end of the apartheid this rule was overthrown. This resulted in a lot of residents and businesses fleeing the area because of an increase crime in the area. In the 1990's, after apartheid ended, the once prosperous area developed to be a no-go zone for most people. A number of large companies and governmental bodies are still located in the area with it's enormous office buildings but these have their own parking, restaurants and shops protecting their workers from the risks of the streets. After the demise of the central business district, economic hubs got more fragmented which resulted in several economic nodes throughout the city with a focus of economic nodes in the northern part of the city as can be seen in Figure 8.

Johannesburg is a city of extremes where there is a lot of poverty as well as a lot of wealth. This is visible in the existence of a lot of informal settlements as well as gated communities all over the city. There are currently 318 recognized informal settlements in the Johannesburg region with the estimated number of residents exceeding 500.000 (Luvhengo, 2022). With the informal nature of these settlements, the real number of informal settlements is likely to be higher. In contrast, there are around 8000 residential gated properties with each containing one or more houses (Joburg Municipality, 2022). The locations of the registered informal settlements and the gated properties can be seen in Figure 8 where we can see that the gated properties are situated in the north of the city while the majority of the informal settlements are found in the south and near the borders of the city. Moreover, the gated properties are located around the economic nodes and only a few informal settlements are close to economic nodes.

Census data for the City of Johannesburg in is available in different levels of spatial aggregation. the smallest spatial denomination available. The most complete and extensive data found for the City of Johannesburg comes from the 2011 census as the 2022 census results were not yet available during this research. The data was retrieved from the city their online GIS platform GeoLIS (Joburg Municipality, 2022) which provided data on subplace level. These subplaces are administrative areas where one or more blocks are joined together and divide the city in 804 smaller areas with an average of around 5500 residents. However, there are a significant amount of larger subplaces with 127 subplaces having more than 10.000 residents including 41 subplaces with 20.000 or more. The

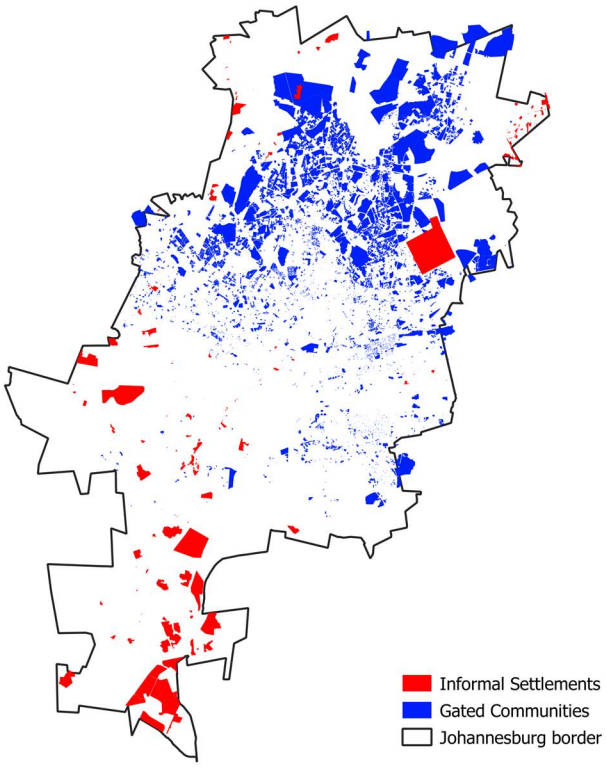

Fig. 8. Overview of gated Communities, registered Informal Settlements in Johannesburg. The overview shows that most gated communities reside in the northern part of the city and the most informal settlements can be found in the south of the city or at the edges of the city in the north. Source: City of Johannesburg (2011)

data on a variety of socio-economic indicators such as income, education, employment and housing characteristics (Statistics South Africa, 2011).

3.3.2 Housing policies The government of Johannesburg recognizes the housing disparity in the region and has actively tried to implement policies that improve living conditions for people with low incomes. This has resulted in low-income housing projects as a result of the national housing policy of 2007, funded by the local and national governments offering affordable accommodation in the city (Charlotn, 2014). However, these projects often reside on the border of the city away from economic activity centers, limiting the opportunities for the residents of these areas (Charlton & Meth, 2017). As the housing policies were not yielding the desired results over time a new, local and inclusive housing policy was adopted by the city council of Johannesburg in 2019 (Webster, 2019). This policy stimulates mixed-income housing development where 30% of the developed houses in projects must be social housing with a capped rent and a set list of requirements for living conditions (Charlton & Meth, 2017). The effects of this new policy still have to show but show the city council's

intention of providing affordable housing for low-income households. There are also critiques on the new policies not being inclusive enough as the proposed rent cap is still too expensive for about half of the households in the city.

Overall, the City of Johannesburg presents a diverse and complex case for this research on the application of remote sensing techniques. Its challenges of rapid urbanization, a history of segregation, and environmental issues provide an opportunity to study the city in depth and potentially identify spatial structures of the built environment that form this city.

4 Results

In this section, the results of the research described in Chapter 3 on the City of Johannesburg are discussed. The results of an initial attempt of the CNN Autoencoder is presented in Chapter 4.1 together with the results of the final CNN Autoencoder. After this, the clustering results as well as the effects of the smoothing are shown in Chapter 4.2. Following this, the clusters will be presented in Chapter 4.3 with their given typologies. In Chapter 4.4, the clusters are analyzed spatially and as well as being compared to known informal settlements that are registered by the city. Finally, the found clusters are compared census data in Chapter 4.5 in order to identify socio-economic patterns in the found clusters.

4.1 CNN Autoencoder

A neural network is simple to set up and get working in today's world, but it is more difficult to get working well. Initially, the methodology of Singleton et al. (2022) was copied, which in terms of setup is similar to the final CNN autoencoder described in Chapter 3 but scaled to a different size of the input images. However, after running the algorithm for on low resolution (10 meter per pixel) images the reconstructed images were disappointing. A sample of the input images and the restored images by the CNN Autoencoder are shown in Figure 9. As can be seen, the input images only contain little detail and the restored images are even less insightful. A map of the initial clustering of these results can be found in Appendix A and this clustering was not showing a lot of potential because of it's seemingly random distribution. Because Singleton's (2022) research covered an entire country rather than just one city, the conclusion was reached that the differences in a single city are likely to be more nuanced than in a country with both urban and rural areas. This introduced the question on how to improve the algorithm and get better results. This resulted in the choice for a higher resolution of the input images as more detail could help in better capturing nuanced differences between different areas. In the end two high resolution iterations of the original algorithm have been explored and the three algorithms and their performance metrics are shown in Table 2 and the network designs are shown in Appendix A. The changes in the algorithm show that the improved algorithm has a higher data retention than the original algorithm which shows that the latent representation is a more detailed representation of reality.

A necessary change that was required between the original low resolution algorithm was that the area of interest needed to be changed from 160 by 160 meters to 64 by 64 meters in the high resolution CNN Autoencoders due to limits of the available RAM in the used compute units. However, adjusting the algorithm to take larger images but keeping the latent representation to be 64 values did improve the results but did have a higher loss value for both the training and test data. This does mean that more data is retained as 0.9039 accuracy on 12.288 values on the initial high resolution algorithm is a better retention of information than an accuracy of 0.09005 for 1024 values. As the input data went increased by a factor 10, an increase in the size of the latent representation was the next step in improving results. This resulted in a better performing neural network that retained 256 values instead of 64 values and there was a 1.7% increase in accuracy compared to the initial high resolution algorithm. Moreover, the loss scored the best of the three algorithms which means that the reconstructed values are closest to the ground truth of the three algorithms. All in all, the use of higher resolution images means that more data is being stored in the latent representation which will improve further results. The retention of details can be seen in the sample images shown in Figure 10.

Table 2. Comparison of three CNN Autoencoders

	Low resolution	High resolution 1	High resolution 2
Input values	16 x 16 x 4 (1024)	64 x 64 x 3 (12,288)	64 x 64 x 3 (12,288)
Resolution	10 meter per pixel	1 meter per pixel	1 meter per pixel
Covered area	160 * 160 meter	64 x 64 meter	64 x 64 meter
Latent representation	8 x 8 (64)	8 x 8 (64)	16x16 (256)
Compression	6,3%	0,5%	2,1%
Layers	21	36	28
Trainable parameters	28.600	422.865	341.897
Non-trainable parameters	82	284	276
Converging epoch	283	401	481
loss (training)	0,0112	0,0132	0,0056
accuracy (training)	0,8907	0,9014	0,9171
loss (test)	0,0086	0,0134	0,0057
accuracy (test)	0,9005	0,9039	0,9181

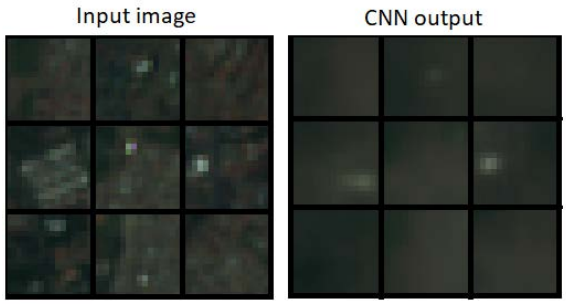

Fig. 9. Input and results for the low resolution CNN autoencoder showing that only little information is input as well as retained for the images.

4.2 Clustering and smoothing

After training the CNN Autoencoder and generating the latent representation for the complete dataset, the data is ready for smoothing and clustering. The data was processed using a spatial smoothing algorithm which averages a data points' latent representation with all the datapoints within a range of 128 meter which is a way of introducing spatial lag to the data and smoothen out outliers within neighbourhoods. The K-means algorithm was used in order to create clusters in an unsupervised manner.

To see the effects of smoothing on the resulting clusters, both local effects and global cluster effects can be observed. An example of local effects of smoothing was shown in Figure 6 in the Chapter 3. For the global effects on the clusters of the smoothing algorithm, the WCSS can be analysed. The result of clustering both the unsmoothed data and the smoothed data with a k-value from 1 to 14 is presented in Figure 11. For each k-value the K-means clustering was performed 5000 times in order to be able to compare the resulting clusters fairly. Overall, the smoothed data has better WCSS scores showing that smoothing made the resulting clusters more compact and therefor more similar to each other. The number of clusters with the lowest average WCSS score is the smoothed

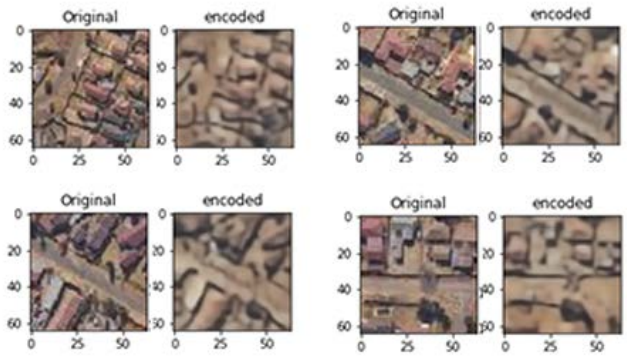

Fig. 10. Input and restored images for the final CNN autoencoder showing the retention of shapes and structures in the images.

data with 8 clusters. The surrouring number of clusters score simillar as the cluster size of 8, the cluster sizes of 6,7 and 9 all being within 2% of the score of 8. The choice for a cluster size was made as it had the lowest score and is on the higher end of the options allowing for, possibly, more nuanced clusters than working with one or two clusters less.

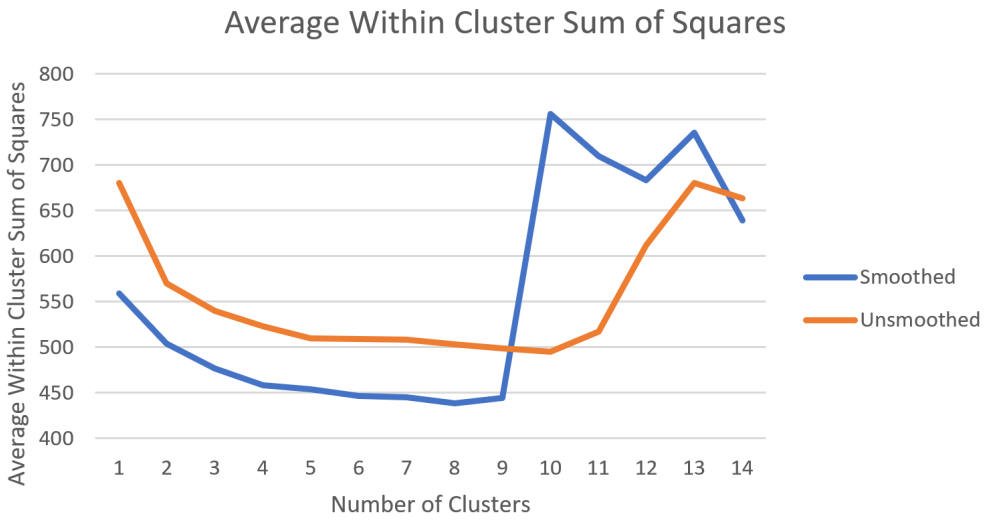

Fig. 11. Comparison of the Average Within Cluster Sum of Squares for a cluster size of 1 to 14.

In Figure 11, the first half of both the smoothed and unsmoothed average WCSS values look like a so-called "elbow shape" which is a typical shape that the WCSS values of K-means clustering

take on. However, in this plot the average WCSS value rises in the end which is atypical behaviour for WCSS plots. This can be explained by the combination of two aspects of our data: the high dimensionality and the use of average WCSS instead of just the WCSS. Because it is an average WCSS value over the clusters rather than an average of all individual values small outlier clusters can affect the results significantly. The high peaks of aver values occurs because of outlier clusters appearing in the K-means clusters as can be seen in Figure 12 which shows that there is a single cluster G with only 54 out of nearly 3 million data points in it which raises the average a lot. This is also one of the side effects of the K-means algorithm which optimizes the clusters on distance to cluster centers rather than finding clusters of equal size and optimal center distances for those. This behaviour of different cluster sizes is also what is good for our use case as it the different built environment types are not gonna be having an equal amount of houses in them.

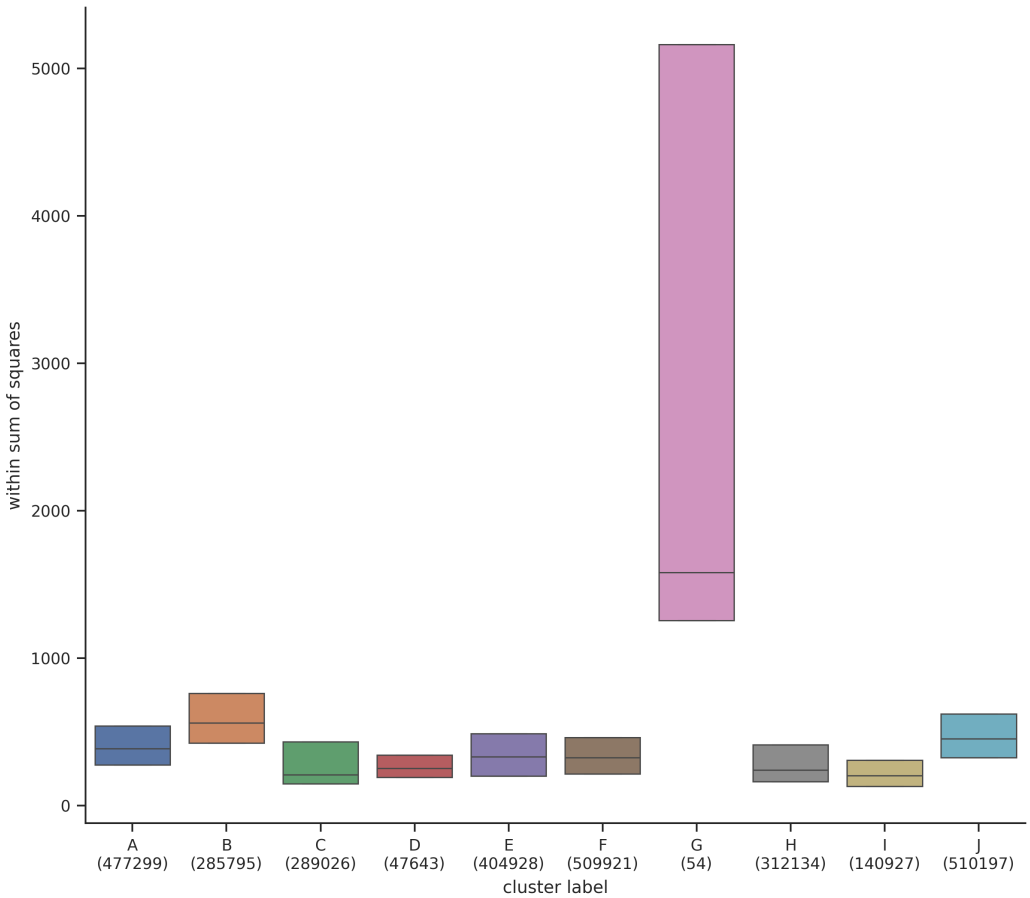

Fig. 12. Outlier behaviour of within sum of squares for a cluster size of 10 within the smoothed data K-means exploration. Showing how cluster G with 54 out of roughly 2.7 million datapoints can skew the average cluster values

The WCSS score for each of the clusters of our final selection of 8 clusters is presented in Figure 13 and shows the spread of values within all of the found clusters. What can be seen is that the size of the 25th to 75th percentile for most clusters is quite similar as half of the clusters span around 280 points and only one has a bigger range at 326 points. This means that a most of the clusters are relatively similar in their compactness of similar values. Further analysis of the actual clusters will show whether the similarity of the built environment found within each cluster is similar or not.

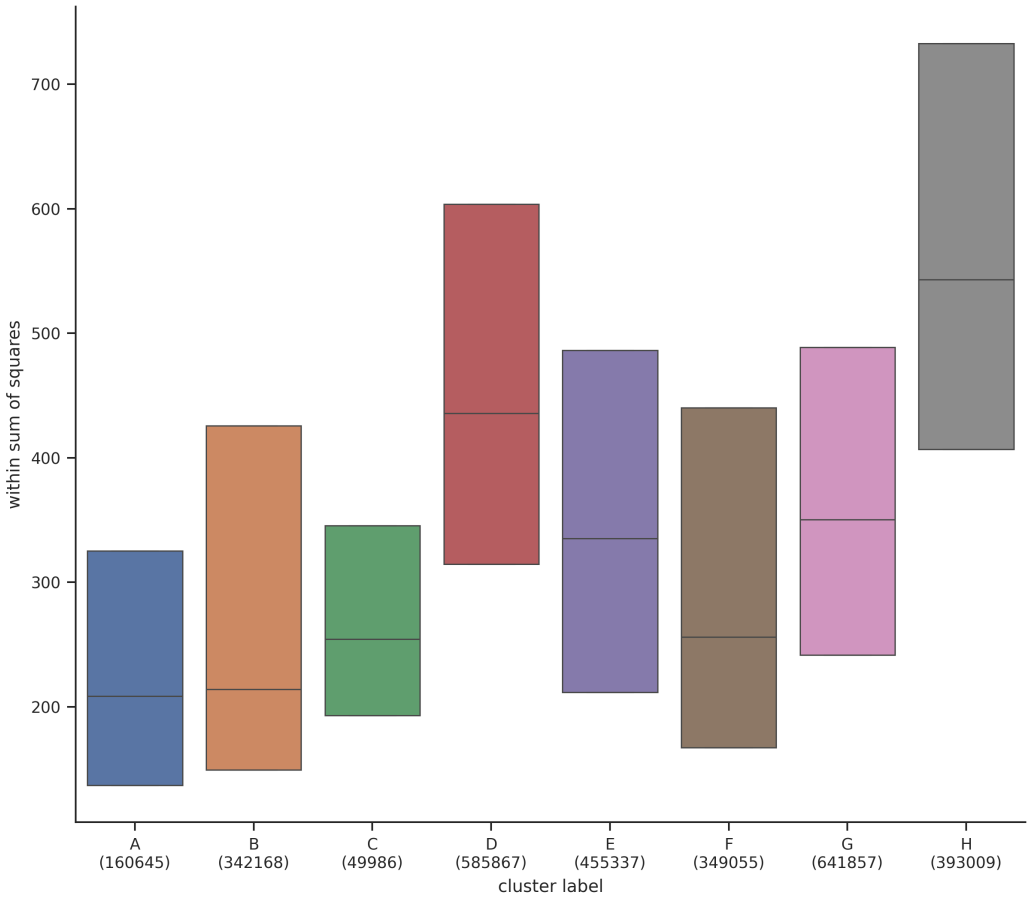

Fig. 13. Comparison of the Within Cluster Sum of Squares for the selected number of 8 clusters. The figure shows the spread of WCCS values for each of the resulting clusters.

4.3 Cluster typologies

In order to give a typology to all of the cluster, the found clusters have to be analyzed. This is done by assessing a sample set of 50 images for each of the clusters on the level of architecture and urban design from informal to formal. This has been done by manually assigning these values using a combined view of the satellite image as well as street view images for each of the sample images. The results of these classification are shown in Table 3 and the classified images per cluster can be found in Appendix B. These should not be considered as a complete truth for each of the clusters as it is a small sample compared to the complete dataset but does allow us to develop our understanding of the clusters and the types of built environments in them. An even smaller overview of the clusters is shown in Figure 14 which shows four images for each cluster.

Table 3. Classifications of cluster sample images based on the architecture and urban design.

	Architecture	Formal	Formal	Formal	Mixed	Mixed	Mixed	Informal	Informal	Informal
	Urban Design	Formal	Mixed	Informal	Formal	Mixed	Informal	Formal	Mixed	Informal
Cluster										
A					29			21		
B								44		6
C					50					
D		20			26	1				3
E			6		21	3		20		
F		6			20	7		6	6	5
G		15	33						2	
H		35	6		6					3

4.3.1 Cluster A: Structured shacks This cluster shows a combination of mixed and informal architecture with a formal. There are clear roads and building plots present in the images which shows that some urban planning was present in these areas. The architecture is split between mixed and informal where the informal areas are mainly shacks while the mixed areas show areas where a part of the buildings show are built using better building materials with permanent roofs but others still have corrugated metal.

4.3.2 Cluster B: Unstructured shacks This cluster shows mainly consists of areas with informal architecture and and a mix of informal and formal urban design. The houses in these areas are built spaciously apart with each clearly on their own assigned plot of land and all having access to roads. The built quality of the houses is low and can be considerated shacks with cheap building materials. Ten percent of the sample images got the label of informal urban design because these buildings are built more randomly throughout the neighbourhoods they are built in but show the same building characterics as the rest of the images in the cluster.

4.3.3 Cluster C: Formal housing with backyard shacks This cluster shows areas with a mixed architecture and a formal urban design. The areas mainly consists of well build, good quality houses however, most of these houses have a small shack attached to them or in their backyard, giving it a mixed architecture overall.

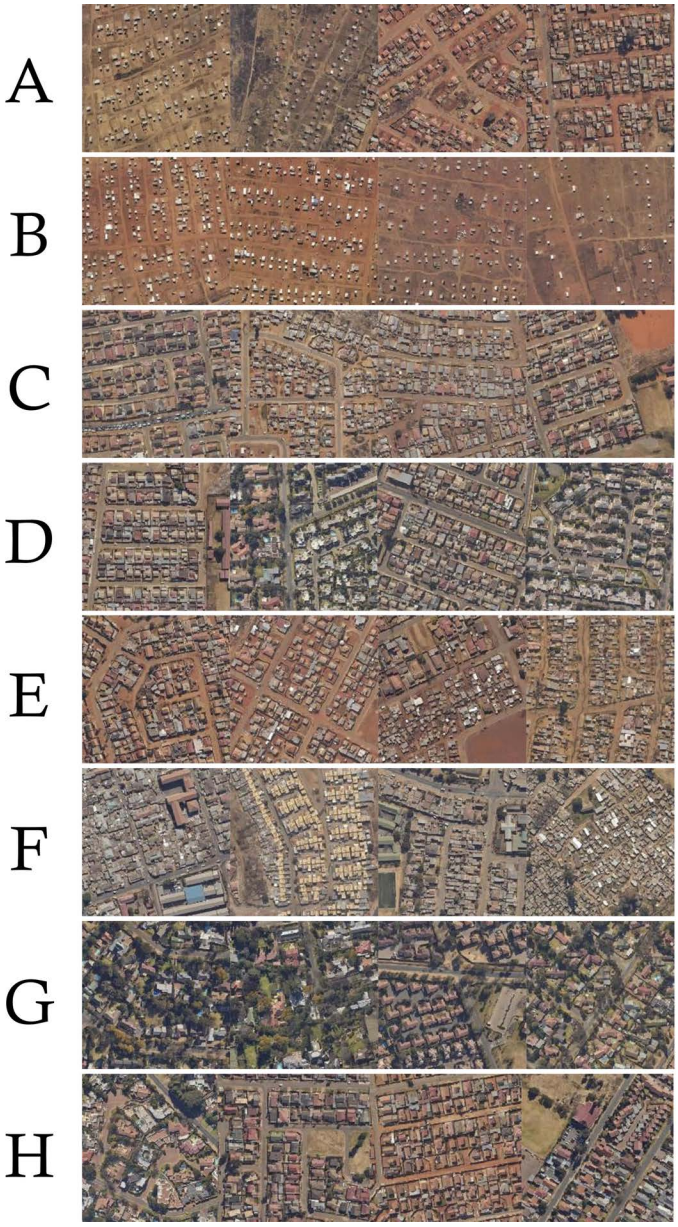

Fig. 14. Sample images per cluster showing differences in built environment

4.3.4 Cluster D: Formal housing with backyard shacks This cluster has a formal urban design and a combination of formal and mixed architecture. The areas with a formal architecture are quite spacious and have relatively large houses and have trees in the streets. The samples with a mixed architecture also have spacious houses but smaller than the previously described houses,

but with the addition of shacks in the backyards. Also the mixed architecture areas show little to no trees in them. So, this cluster is a mix of formal architecture with formal urban design and mixed architecture and formal urban design meaning that these areas are quite well developed areas overall.

4.3.5 Cluster E: Mixed built environments This cluster has mainly got a formal urban design in combination with a combination of a informal and mixed architecture. What stand out in this cluster is the housing density within the building plots as a lot of the buildings within a plot are built touching each other without space between them. As if the living space has been grown increasingly by adding small buildings within a building plot over time. There are also some formal architecture, mixed urban design samples which show a variety of built environments that are quite spatial and do not really show similarities with each other and might be considered outliers within this cluster.

4.3.6 Cluster F: Mix of everything This clusters biggest contributing class is mixed architecture and formal urban design but there 60% of samples can be found in 5 other types which are split more or less equally. The formal architecture, formal urban design, looks like outlier images as it covers some commercial or industrial areas rather than residential areas. Other than that, it is a combination of informal architecture and mixed architecture. The informal architecture does appear to be of the worst possible quality and the mixed architecture appears to be on the lower end of development of formal architecture. The areas appear, in terms of architecture to be on the lower end of the mixed architecture scale or in a stage where the next part would be to develop better housing to become a mixed architecture area.

4.3.7 Cluster G: High end houses This cluster has a formal architecture and shows a combination of formal and mixed urban design. There are two outliers in informal/mixed which can be explained by the proximity to formal/formal areas and the smoothing process. This cluster shows a lot of trees in the sample images and the houses are very spacious. The urban design for a lot of the samples is considered mixed as the buildings in these areas have been built in their plots to match the orientations of the owners of the buildings rather than all being organized as oriented in the same way as the rest of the neighbourhoods. Most of the samples in this cluster can be found in gated communities which probably allows for this less structured urban design within the community.

4.3.8 Cluster H: High end houses This cluster has the majority of the samples assigned to the formal architecture and formal urban design type. There are a number of samples in the mixed architecture and mixed urban design type but visually these neighbourhoods are quite similar. The different in their architecture is some low quality shacks in backyards in the mixed architecture type. There are also a few informal architecture and informal urban design sample images which do not seem to have a relation with the majority of the samples within this cluster. The overarching appearance of the areas in this cluster is that the cluster represents structured neighbourhoods with some trees and buildings that are spacious but not exorbitantly big.

What can be seen in the descriptions of the different clusters is that in terms of the chosen characteristics of architecture and urban design a lot of the clusters show similarity with each other. For

example, 5 out of 8 have a high number of samples in them with mixed architecture and formal urban design. This means that on their face value they could be considered similar clusters. However, the clusters show different combinations of characteristics throughout making them different clusters. Moreover, the discovery when analyzing the clusters samples was that the built environment needs more nuance than just being classified informal, mixed or formal as within these three categories another range of categories could be identified. For example, the difference of mixed architecture neighbourhoods with shacks and brick houses alternate on plots of lands or where brick houses have shack like side buildings connected to them or brick houses with shacks in the backyard all categorize as being mixed architecture but in practice can represent different types of neighbourhoods. And these types of nuanced differences can be found in all of the characterizations. Another interesting discovery was the level of urban planning found in the sample image where even most areas with informal architecture are well structured with clear building plots and roads in them. What also was discovered during the analysis is that there are outliers within the data and that some of these are quite different from the rest of the data like industrial areas and that some of these outliers are likely to be caused by the smoothing process and the proximity of multiple different built environment types. Thus, the clusters are by no means perfect seperations of each others but do show a lot of promise in terms of their visual appearance and in order to further explore the clusters the spatial aspect of the clusters will be explored in the next section.

4.4 Spatial analysis

Besides the visual aspects of the clusters, there is also a spatial component which can be explored. In Figure 15 there is an aggregated display of the clusters on the map of Johannesburg as on the scale of the complete city plotting just the houses results in a faded figure as is shown in Appendix B. On a first glance there is a clear spatial difference between a number of the clusters. A first glance of Figure 15 shows most of the houses in clusters F, G and H are in that north of the city and clusters A, B, C and E are found through the southern part of the city. Cluster D is found predominantly in the northern part of the city but also has a high concentration just below the geographical center of the city.

If we compare Figure 15 with Figure 16, which is the same figura as in Chapter 3, we can see a similar divide of the city with gated communities in the north and informal settlements in the south. This gives us the notion that on a high level, the algorithm was able to differentiate less developed areas from beter developed areas. By overlapping both the registered informal settlements and the registered gated communities with the found clusters, the comparison can be enumerated. In Table 4 the results of this comparison are shown and provide further insights into the found clusters. Cluster A and E both have high percentages of houses in informal settlements with nearly half of the houses of A and a quarter of the houses in E being found areas that are registered informal settlements. The clusters with the highest percentage found in gated communities are clusters E and H with 23.5% and 31% of the houses being attributed to gated communities. What is interesting is that the numbers for A and H could be expected as these clusters could be considered the extremes of an informal and formal built environment which should reflect with the numbers of informal settlements and gated communities. However, cluster E stands out with almost similar percentages of the houses being found in informal settlements as well as in gated communities. In terms of spatial aspects this spread comes from a big part of the cluster being in the south where most of the informal settlements can be found. Moreover, the contribution of the gated communities comes from the scattered properties that can be found in the northern part of the city and specifically in

the less densely built areas. This means that this cluster might not have a clear typology and might actually represent two types of built environments.

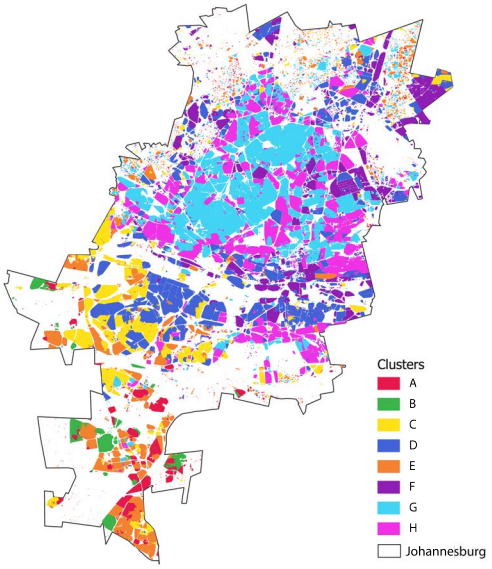

Fig. 15. Simplified overview of the clusters over the City of Johannesburg. The simplification is that houses within the same class within 50 meter of each other are joined to a single entity in an itterative process and has been performed for visualization aspect and does not affect the underlying data.

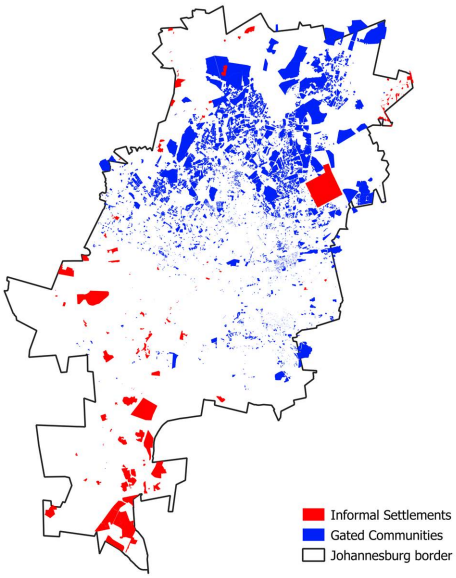

Fig. 16. Overview of gated Communities, registered Informal Settlements in Johannesburg. The overview shows that most gated communities reside in the northern part of the city and the most informal settlements can be found in the south of the city or at the edges of the city in the north. Source: City of Johannesburg (2011)g

One thing to note in Table 4 is that a lot of clusters have 6 to 13 percent of the houses in informal settlements or gated communities or both while not being clusters that necessarily represent these types. These numbers can be explained by multiple aspects and can be any combination of these aspects. Firstly, the algorithm is not perfect and the built environment has a lot of variation in it which can be cause for outlier classifications within areas. Secondly, there can be anomalies in neighbourhoods like schools and hospitals that can alter the classification of that specific area while the most of its surroundings are classified differently. Thirdly, the smoothing algorithm did have some effects on influencing the outer rings of adjacent neighbourhoods which can have influenced these numbers. Another aspect which is of influence is the comparison of data sources as the satellite imagery used is from July 2022 and the information on the informal settlements and gated communities are from 2012. In these 10 years the city grew a lot and areas changed a lot. As a result, these numbers presented are indicative but do show that the extremes have been captured in the classifications.

As this research has the aim to find informality, the above results can be studied on a combined map with the known informal settlements as well as the clusters that have the most overlap with these points. As most informality is present in the south of the city, Figure 17 shows the combination of clusters A and E in combination with the known informal settlements. As can be seen, there is a lot of overlap with these informal settlements and clusters but there is also large parts of both of these clusters that are not within these known informal. This means that with the knowledge of the built environments of both clusters and the informality overlap they show, that the informal settlements may have grown in the 10 year time period between the registration date of the informal settlements and the date the satellite images were taken. As the used classification was not completely accurate, we can't conclude definitively that all the areas that are highlighted in Figure 17 by clusters A and E are informal settlements and the rest of the areas are not but we can say that they are likely to be areas with a higher chance of deprivation than other areas.

4.5 Census data

In order to verify the information that we have found previously, the clusters can be compared to each other using the 2011 census data. The data in this census covered a lot of topics from income to education to housing and facilities. As in this research the focus is on informality and deprivation, the covered census statistics will be related to the five measures of deprivation of UN Habitat habitat2016urbanization: improved water source, lack of access to improved sanitation facilities, lack of sufficient living area, lack of housing durability and lack of security of tenure. Not all these five point translate directly to the information that is available in the census data but for each of the five points a proxy has been chosen and will be evaluated for all the clusters in this section.

4.5.1 Water access The first measure of deprivation is access to an improved water source. The census data includes a number of variables on different types access to clean water sources including

Table 4. Distribution of classified houses in clusters in registered informal settlements and registered gated communities

Cluster	Total number of houses	In Informal settlements	In gated communities
A	161747	75630 (46.76%)	3213 (1.99%)
B	50160	3624 (7.22%)	17 (0.03%)
C	459411	39044 (8.5%)	35939 (7.82%)
D	639747	40207 (6.28%)	82634 (12.92%)
E	350728	88046 (25.10%)	82634 (23.56%)
F	342242	23372 (6.83%)	14641 (4.28%)
G	391889	5806 (1.48%)	26746 (6.82%)
H	586340	8945 (1.53%)	182034 (31.05%)

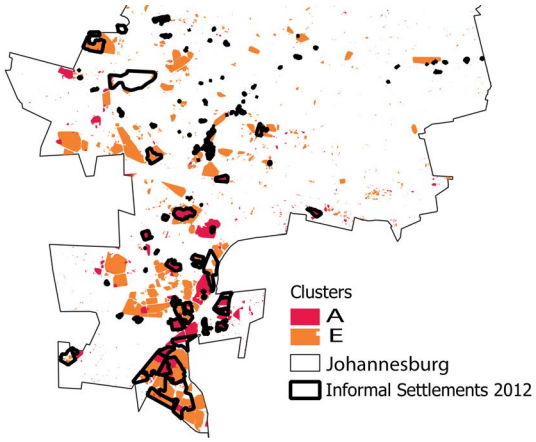

Fig. 17. Informal settlement development in the south of the city. The figure shows that there is a lot of overlap between the informal settlements acknowledged by the city in 2012 and the clusters A and E from our results but also suggest that the informal settlements have changed and grown since 2012

a number involving the location of the water source. As a proxy for improved water sources, the variable, general access to tap water was chosen. As can be seen in Figure 18, houses in cluster A have the least access to piped tap water followed by cluster E. This means that the areas in cluster A and E are the least developed in terms of access to water.

4.5.2 Sanitation facilities The second measure for deprivation is the access to improved sanitation facilities. As there is a range of different types of sanitation options from holes in the ground to septic tanks to sewage system connected toilets. As a proxy for access to improved sanitation options, the variable on houses with flush toilets connected to the sewage system was chosen as it is the highest level of sanitation. Figure 19 shows that clusters A has the least access to the sewage system followed by cluster E and F that have a wide spread of values within their data. Meaning that there are significant amount of houses in those clusters that lack access to the sewage system. This means that these three clusters have areas are least connected to the sewage system and might be forced to less hygienic means of sanitation.

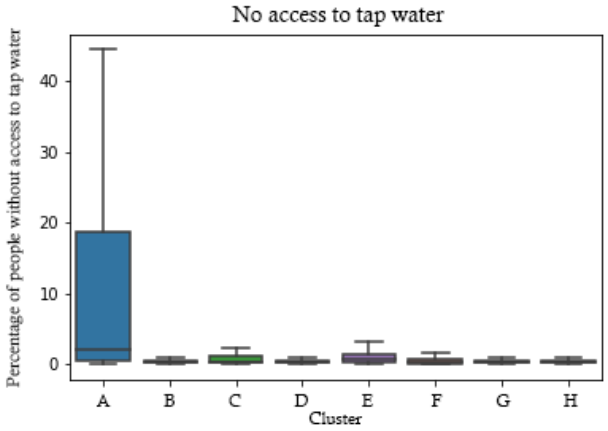

Fig. 18. Access to piped water

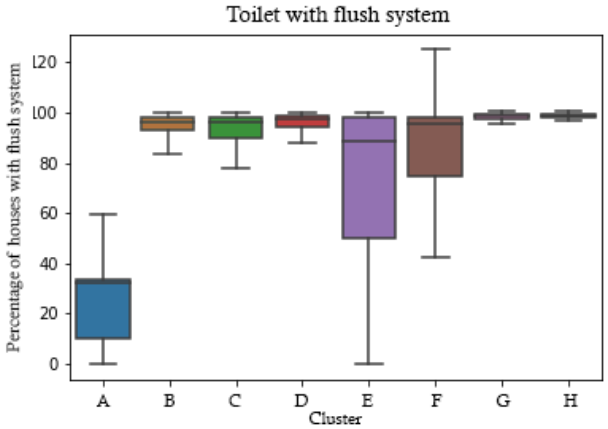

Fig. 19. Access to toilet with sewage access

4.5.3 Living areas For the third measure of deprivation, the living area can be measured. As a proxy for this, the area of the houses are used. Instead of using reported data from the census records, the input data file on the houses for this research was used as it contains the shape and area for all of the houses. This means that it is probably more accurate for the size of the buildings than self-reported and aggregated census data. However, this data does not account for the number of floors in houses as well as that it might detect connected shacks as separate buildings. Figure 20 shows the distribution of the area per building in square meters for all of the clusters. The smallest average house size can be found in clusters A and B with houses that, on average, are smaller than 20 square meters while in the other clusters the average house sizes are similar around 30 square meters. This means that according to house sizes, clusters A and B score the worst.

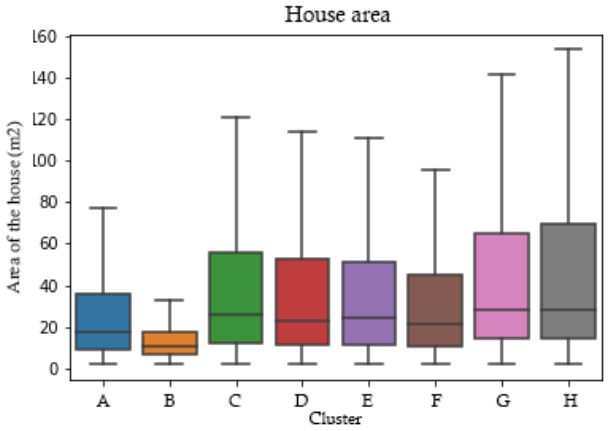

Fig. 20. Living area per cluster in square meters

4.5.4 Housing durability The fourth measure of deprivation is the lack of housing durability. As a proxy for housing durability a variable on shacks was selected that represents people living in shacks that are not in the backyard of another building. As shacks are know for their cheap building materials it can be used as a proxy for information on housing durability. Figure 21 shows that, on average the most shacks can be found in cluster A. Clusters E and F both have low percentages of shacks on average in their areas but do have a large quantile range meaning that there are neighbourhoods in the clusters with a significantly higher number of shacks in them.

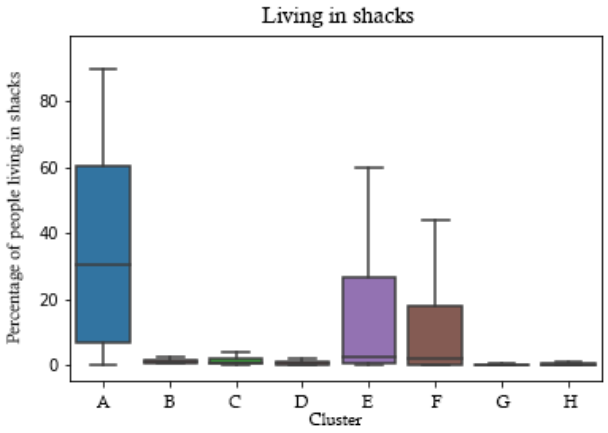

Fig. 21. Shack as residence that is not located in the backyard of another house

4.5.5 Security of tenure The last measure of deprivation is the lack of security of tenure, which of the five measures might be the most difficult to find an appropriate proxy for as the census does not contain a lot of information on security of tenure. But, the census does include data on whether people own the house they live in or rent the house. Therefor as a proxy for tenure security, the variable house ownership was chosen as having ownership of the house you live in provides a certain level of security of tenure.Figure 22 shows that clusters G and H have the highest level of property ownership with near 40% of the people owning their houses. The clusters that have the smallest percentage of home owners are clusters A and F. Therefor, the houses in clusters A and F probably have more people renting or have a different living arrangement which means that a higher percentage of people in these areas will probably have a lower security of tenure.

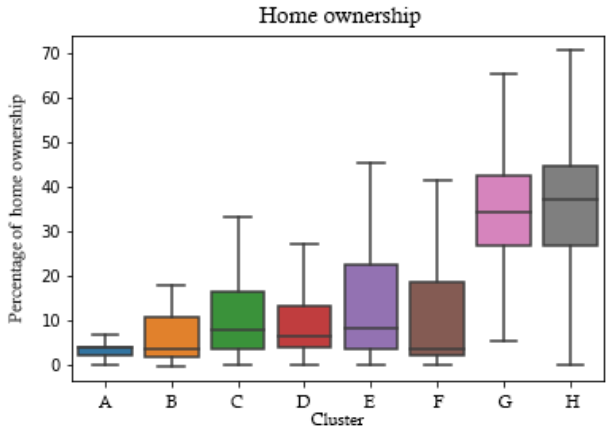

Fig. 22. Complete ownership of property

So, after comparing the clusters for the five proxy variables for deprivation, there are four main takeaways. First, cluster A has the most deprived areas in it with quite some distance to the other clusters. Secondly, clusters E and F have some areas in them that have show more signs of deprivation but on average do not deviate a lot from the averages of the other clusters. Thirdly, clusters B, C and D show little signs of deprivation and are quite similar to each other in terms of the analyzed proxy variables. As a fourth take away, the data suggests that clusters G and H are the least deprived areas as these clusters are mainly located in areas with low scores for the deprivation variables. All in all, the census data shows that there is a difference between most of the clusters found by the clustering algorithm.

4.6 Summary of results

The results of the CNN Autoencoder, the smoothing process and the K-means clustering algorithm resulted in 8 different clusters of different classifications of the built environment. In this chapter, these classifications have been evaluated in three different ways: using the visual appearance of sample images, by comparing spatial data with known extremes within the City of Johannesburg

and by comparing census data for the found clusters. These analysis have shown that a number of the found clusters are distinctly different and that others more similar to each other. The cluster that shows the most deprivation and informality is cluster A which shows in all three evaluation methods that scores the lowest in terms of level of urban development and has the most overlap with known informal areas. This is followed by cluster E which has some unexpected characteristics. A part of the cluster shows high levels of deprivation characteristics but simultaneously this cluster also has a big part of its buildings located in gated communities that have the least amount of deprivation in them. The inclusion of both of these extremes within this clusters makes it a difficult cluster to use in drawing conclusions on whether an area can be considered deprived or not. Clusters C, D and F are all similar in terms of the characteristics found of these clusters. There are nuanced differences in terms of the built environment but all clusters show similar levels of urban development. Of these four clusters, cluster F stands out the most in terms of census data as this cluster does seem to have a bit more areas in it with higher deprivation scores but on average scores similar to the other three clusters. According to the analysis, clusters G and H can be considered the most developed as these clusters have a formal architecture and have low deprivation scores based on the census data. Also, these houses in these clusters are mainly located in the areas where the least amount of deprivation is expected within the city. Cluster B is a cluster with contradicting evaluations, the visual inspection of the cluster showed a lot of informality which implies that the areas should show signs of deprivation in the other two evaluations. However, the other two evaluations do not highlight this cluster as having a low level of development. Inspecting the location of this cluster shows that it is located near other deprived areas in the south part of the city. This could mean that the data used to compare to clusters is outdated or does not have the correct information on these areas. Considering the differences between the different clusters, the results shows a spatial segregation of different built environments within the City of Johannesburg with a further developed northern part of the city. All in all, the results show that the used methodology was able provide clusters that include different types of built environments that can help to gain insights into the spatial structure of the researched city and highlight areas subject to deprivation.

5 Discussion

In this chapter, the key findings of this research will be discussed. These results will be discussed in relation to the research questions and how this research fits in with the existing literature on the topic. The limitations of the research will also be discussed as well as possible policy uses for the results of this research.

5.1 Main findings

The main goal of this research was to use identify spatial structures within a city using remotely sensed data. To achieve this goal, a number of steps had to be taken which include 1) identifying suitable data sources, 2) implementing multiple machine learning algorithms, and 3) analyzing the resulting clusters. By applying these steps 8 built environment clusters were identified for the City of Johannesburg. The resulting clusters showed different types of the built environment with some clusters showing similar characteristics and others being distinctively different. The results showed a clear distinction of different built environments between the north and the south of the city which aligns with the historic racial segregation of the city and the vast difference in economic possibilities and resources available in these two parts of the city (Ballard & Hamann, 2021). As the built environment, to a certain extent, reflects the living conditions of the residents of these areas, the analysis of the clusters helped to identify a cluster where the areas within it show signs of high levels of deprivation. Most of the clusters showed a limited scope of built environment styles within the individual clusters but there was also a cluster that was more divided in the found built environments within that cluster. Also, the analysis showed that the clusters that were in between the extremes were similar to each other in terms of the variables on which they were compared and while they could contain differences in reality these were not explicit from the evaluation processes that were performed. The found clusters showed that even though there were areas with low quality housing, these areas did show levels of structural organization where buildings were built on designated plots of land in a structured manner. The results also showed signs that the official information sources on the location of informal settlements were outdated and that is likely that growth of informal settlements has taken place compared to the available data from 2012. Moreover, these results can help identify areas where the extremes in terms of the built environment are located next to each other and can be a first step in the process to make the gap smaller.

5.1.1 CNN Autoencoder As the CNN Autoencoder is a black box model, it is not really possible to know what specific features of the built environment are considered most important within the model. However, when looking at the resulting images produced by the decoder part of the features that were retained compared to the original images were the buildings, roads, trees of the input images as well as the color of some of these features. But, there might also be some more subtle features that are not that apparent to the human interpretation of the images. The clustering of this information using the latent representation of the input images showed that the CNN Autoencoder is very sensitive to differences in the latent representation and produced neighborhoods with a lot of different classifications scattered throughout. By applying a spatial smoothing algorithm neighborhoods became more homogenous in their classification at the cost of houses at the borders of neighborhoods sometimes being classified as the built environment of the adjacent neighborhoods. So, the smoothing algorithm did improve results but also allowed for some noise in the resulting clusters. In the end, the complete process allowed for the identification of

different built environments within the City of Johannesburg and analyze their spatial distribution in the city. What can be learned from this is that, for the City of Johannesburg, the appearance and relationship between buildings, roads and trees allows for a successful clustering of the built environment and can help identifying the different types of built environments in the city.

5.2 Evaluation

One of the main criteria when evaluating machine learning classification algorithms is the accuracy of the classification it provides. However, in order to evaluate the accuracy in a traditional sense, a ground truth is required. As this research classified the built environments of buildings in an area that did have a ground truth available, this method of evaluation was not possible. In order to evaluate the resulting clusters of the algorithm, three different methods were applied to evaluate the resulting clusters. First, a visual inspection of a sample of 50 images of each cluster was performed where the images were classified based on the architecture and the urban design of that area. Each image was given a classification of informal, mixed or formal on both of these criteria which provided an overview for each cluster and provided information on what different types of built environments could be found in each cluster and how these were distributed. As only 400 out of nearly 3 million images were manually classified, it did provide insights into each cluster but the number of images analyzed was too small to draw definite conclusions from. Also, the images that were manually inspected covered a larger area than was used as input for the algorithm which can allow for a different interpretation of the image than when only area that was classified by the algorithm was considered. However, as we wanted to gain insights on houses in their environment, analyzing a larger area made for easier and more coherent human interpretation and classification of these images in their context. Another, weakness of this method is that human interpretation is prone to human error where misclassification might occur and which can skew the results with these relatively small samples.

Another way in which the resulting clusters were evaluated was by using known information on informal settlements and gated communities as these can be considered extremes in the level of development in the built environment within the context of Johannesburg. This analysis made for an objective comparison of the resulting clusters and to which extent these clusters had areas within them that belonged to each of the two extremes. The results of this evaluation showed that there were two clusters that scored higher than a 25% overlap for one of the two extremes showing that these clusters could be considered as representing an extreme. However, one of the clusters scored high on both extremes which introduce questions on the homogeneity of the built environment of that specific cluster. Other clusters also showed overlap one or two of extremes which can be caused by either misclassification, the smoothing issue, or that the areas that overlapped with the extremes changed in the 10 years between the information on these areas was published and the date of the satellite images. This last point is one of the main issues with this evaluation method as the information on these areas was outdated and these areas have undergone some changes in these 10 years as the city grew around 3% annually. However, as this was one of the few pieces of information available and a lot of these areas are likely to have retained some of its built environment from 10 years ago, it does help to understand the found clusters as it showed that two of the clusters did have significant overlap with these extremes and it pointed out how one of the clusters did not properly contain a single built environment.

The third way in which the clusters were evaluated was by comparing the clusters with census data from 2011. This analysis showed the same pattern that the clusters that were earlier identified as

being on opposite ends of the level of urban development also showed this in the census data. As this analysis highlighted the extremes, it did not help in finding significant differences between the more similar clusters. The census data that was used was aggregated to subplace level which did divide the city into 804 smaller areas but did not match the level of detail of the individual houses that were classified in this research. This means that in a lot of these 804 areas, multiple neighborhoods were included which influences the evaluation of these areas as each house was assigned the census data of the subplace they were located in. Even though with less aggregated census data the evaluation of the clusters could have been more precise, the evaluation did provide insights into the level of deprivation of the areas within the clusters.

5.3 Similar research

This research was based on the research performed by Singleton et al. which used remote sensing to classify different built environments in the United Kingdom. Singleton's research included both urban and rural areas and used the UK's index of deprivation to give context to the found clusters. This research deviated by applying the remote sensing methodologies to a single urban area which introduced the problem that the areas within a single city are quite similar and differences are more nuanced than between urban and rural areas. Because of the need for more information, high resolution satellite images with a resolution of 1 meter per pixel were used rather than lower resolution images of 10 meter pixel. Also, as one of the goals was to be able to identify informal areas, official data sources on the locations of houses were suitable for this research. As a solution, an open source remotely sensed data source from Microsoft was used which included building shapes and locations for the City of Johannesburg but is available for the whole of South Africa as well as most other African countries. The combination of both the higher resolution images and the building dataset allowed for a successful clustering of the buildings within the dataset.

Remote sensing research of the built environment mainly focuses on the detection of slums or informal settlements in a binary fashion, an area is either slum or not (Kuffer et al., 2016). Moreover, current research uses supervised machine learning methods to achieve this goal. Our research has shown that it is possible to extract features from satellite images using unsupervised methods that are useful to create clusters of the built environment. From the resulting clusters from our methodology, the researcher can identify the clusters that are most in line with the label slum or informal settlement. However, the homogeneity within the clusters found in this research are not perfect and might not perform as well as algorithms that are solely focused on detecting slums (Ajami et al., 2019) . However, this research did not only detect slums, it created clusters of different built environments within the city which creates an overview of the city's spatial built environment structure. This can be useful to identify areas within a city where the difference between the development of the built environment is the greatest. This can be useful information in identifying inequalities between adjacent neighborhoods which can be a first step in developing policy that can reduce the gap.

Research on classifying the built environment in a more general sense is generally focused on land use classification which classifies areas as being residential or industrial. While our research did not detect industrial areas in a separate cluster, our research did provide a more nuanced interpretation of the residential archetype by finding different residential built environment typologies which is useful for understanding the dynamics of the spatial structure of the city. But, our research is limited in the fact that the research is performed without set archetypes for the built environment

which can make comparing results from multiple cities more difficult. However, this does allow for a broader discovery of archetypes for each individual city and does not limit the research to a set number of built environment types.

Lastly, the value that this research adds to the current literature is what features are important for understanding differences in the built environment to differentiate built environment typologies. This research has shown that appearance and shape of buildings, roads and trees are likely to be the most dominant features in identifying different types of built environment. This can be useful knowledge for the most dominant stream of research where built environment clustering is performed which is using machine learning algorithms based on object extraction (Kuffer et al., 2020).

5.4 Policy implementation

The aim of this study was to use remote sensing to gain insights into the spatial structures of a city and identify areas subject to deprivation for an area with low data availability. This research has shown that a range of different built environments can be identified using only remotely sensed data which means that it could be suitable to apply to different cities around the world that have low data availability as well. This means that data on the spatial distribution of different built environment styles can become more accessible for policymakers as an aid for their urban planning processes. This can be used for identifying areas where the segregation between built environments is the largest which might be a first step in identifying isolated communities within a city.

This research has shown that it is able to identify areas subject to deprivation. While the process did require human interpretation to identify which clusters show high levels of deprivation, this is a smaller task than labeling training images in more traditional remote sensing methods. Identifying these areas subject to deprivation is key for policymakers if they want to offer support to these developing communities. Traditional ways of identifying informal settlements it a lengthy process on the ground and this research offers a method to identify these areas. Moreover, this methodology could be applied on a regular basis which can enable the tracking of the development of deprived areas over time and see whether policy interventions have an effect on the region.

The city of Johannesburg has a history of social housing projects from which the locations are known to the municipality. Combining the information gathered with this research and their knowledge of these projects can help evaluate whether the social housing projects have resulted in the development of these areas and whether these areas have had a positive impact on their surroundings rather than being isolated projects. Extending this research in a temporal fashion can also help to track these changes in relative urban development to the rest of the city.

On a higher scale, this research could be extended to institutions like the United Nations to help and track the development of cities facing rapid urbanization. This research has shown that it can infer spatial structures within a city which can aid in the process of tracking urban development in line with SDG 11. This can be achieved by extending the methodology to include a spatial segregation score for a city which will allow for better comparison between cities.

5.5 Limitations

The results of this research provide insights into the spatial structure of a city in terms of the built environment using unsupervised methods and with further analysis can identify the areas most

subject to deprivation. Nevertheless, the research was subject to a number of limitations regarding the methods, the available data and the analysis of the results.

The unsupervised nature of the feature extraction and clustering methodology allows for a process that is not subject to prejudices on what the most important features are in classifying the built environment. However, the unsupervised nature is also one of its biggest limitations as it does not allow much space for adjustments in terms of the generated features. As the CNN Autoencoder recreates the input image there is no way of prioritizing features over others and there is no guarantee that the algorithm will include the features that can be most explanatory for the built environment. However, as Singleton's research and our research showed, enough data can be retained by the algorithm to distinguish different types of built environments from each other. Moreover, the results of feature extraction and clustering change with every tweak of parameters or different neural network layers leading to different clusters which are difficult to compare. As a result, only the training statistics of the neural network can be compared in order to improve the results based on the reconstructed images rather than the resulting clusters. This limits the extent to which resulting clusters from different runs can be compared. Nonetheless, this research does show that clustering is possible using used methodology and while the result might not be optimal, it is good enough to provide insights on the spatial structures of the city and the different types of built environments.

Another limitation in the methodology is the need for smoothing of features which highlighted the sensitivity of the CNN Autoencoder to the input images. The spatial smoothing of the latent representations made the classifications of the houses within neighborhoods more homogenous but did, in some cases, introduce the influence of adjacent neighborhoods on the features within a neighborhood. For general interpretation of the results this was not a big problem as we were not looking for patterns on the level of single houses but more on a neighborhood level. However, this did add to the number of misclassifications of built environments within clusters which added noise to the outcomes of the research.

Furthermore, the K-means clustering algorithm introduced limitations regarding the selection of the number of clusters to select. As there is no set number of cluster known beforehand, this number was selected based on the lowest average WCSS score of the range of different cluster sizes. However, the WCSS scores of a number of subsequent cluster sizes was very close to each other which does not give the researcher a lot of information on what the best number of clusters is. This resulted in choosing the lowest score which resulted that a few clusters were quite similar in their built environment that with human interpretation are difficult to distinguish from each other and might have been better as a single joined cluster. However, in this research, we tried to prevent having clusters that had mixed built environment styles and therefore chose for a higher number of clusters rather than lower. With more time, the final number of clusters could be chosen based on the analysis of the clusters rather than a mathematical statistic of the clustering algorithm.

Limitations with data availability were apparent in the evaluation stages of this research where the data used for evaluation had a 10 year difference with the satellite images that were evaluated. That there is limited data availability for a lot of urban areas around the world is one of the reasons to use remote sensing for the collection of data for these areas. However, the old data did make it harder to draw definite conclusions on the validity of the clusters and highlights that there is a need for more recent and accurate data in order to evaluate the resulting clusters. Even though the data was old, it did show that there were differences between the found clusters and allowed us to

confirm that one of the clusters indeed was representative of areas with high levels of deprivation and informality.

6 Conclusion and future work

In this chapter, the research questions will be revisited and conclusions for each of the sub-questions and the main research question will be provided. Furthermore, recommendations for future research will be provided.

6.1 Sub-questions

6.1.1 Sub-question 1 *What are the challenges of measuring deprivation according to literature?*

Deprivation measurement is a challenging task that has received a lot of attention in the academic literature. Deprivation is a multidimensional concept that involves a wide range of social, economic, and environmental elements, which makes assessing it difficult. It is challenging to create a universal method for quantifying deprivation because these characteristics might differ greatly across various geographic regions, cultural contexts, and demographic groups.

One example of an attempt to measure deprivation on a global scale is the United Nations Human Settlements Programme's five criteria for measuring deprivation. These criteria include lack of access to improved water sources, lack of access to improved sanitation facilities, lack of sufficient living area, lack of housing durability, and lack of security of tenure. These criteria aim to capture the multiple dimensions of deprivation, including basic needs and living standards, as well as issues related to housing and infrastructure. These factors are difficult to assess as they mostly depend on census data which is aggregated to administrative areas that can cover large regions which combine multiple neighbourhoods through which deprived areas might be hidden in the statistical aggregation methods.

Measuring deprivation is frequently complicated by other factors like poverty, inequality, and social exclusion. These issues are in addition to those linked to the complexity of deprivation and the availability of data. Although they are frequently closely associated to deprivation, these elements are not always the same as deprivation. When attempting to precisely assess deprivation, it can be difficult to untangle the intricate interactions between these variables. Deprivation can be difficult to quantify, which highlights the necessity for a nuanced and context-specific approach to understanding and resolving deprivation in various groups and circumstances.

6.1.2 Sub-question 2 *How has remote sensing been used in existing literature to characterize the built environment?*

The use of remote sensing has grown in popularity in recent years as satellite images have become more widely available and computing power has become more accessible. Remote sensing has several advantages over traditional data gathering techniques, including cost effectiveness, speed, and high levels of geographical accuracy. One area in which remote sensing has been particularly useful is in the detection and mapping of slums and informal settlements, which is the main approach to measuring deprivation.

According to the existing literature, there are four main methods used for slum detection using remote sensing: feature extraction, texture-based approaches, pixel-based approaches and contour detection. The most used method is feature extraction, which involves generating features that resemble the input data. Pixel-based approaches classify the land use of each pixel to create a land

cover map of the area. Contour detection methods use machine learning algorithms to identify the borders of slum areas and accurately predict their size and locations. Texture-based approaches analyze the texture of an image over a larger area to determine land use. The majority of research in this area has used high resolution satellite images with a resolution of less than one meter per pixel in order to capture fine-grained details. Most studies have relied on supervised machine learning techniques, where the data is manually labeled or a different type of ground truth data is used. However, there has also some research that uses unsupervised techniques, which generate features on the input images and cluster them based on similarity of the features.

One limitation of current research is that it focuses on binary slum detection, meaning that an area is either classified as a slum or not. This approach misses nuance and may not adequately capture the diversity of slum settlements. Moreover, there is a need for more research on the use of unsupervised machine learning techniques, which has the potential to overcome some of the limitations of supervised methods, such as the need for ground truth data or manually labeled images. The use of remote sensing in the characterization of the built environment has the potential to provide valuable insights into the spatial patterns and dynamics of slum settlements which can help to form policies to help these areas.

6.1.3 Sub-question 3 *Which built environment typologies can be found in the City of Johannesburg using unsupervised remote sensing techniques?*

The results of this study show that the City of Johannesburg is a complex and diverse urban area with a range of built environment typologies. Using the unsupervised CNN Autoencoder to generate latent representations for all of buildings in the city and clustering these latent representations, we were able to find eight different clusters of built environments. These clusters represent a range of built environments with informal, mixed, and formal architectural styles and urban design styles. Some of the clusters represented more informal built environment styles while others showed formal styles. The clusters with informal built environments contained a lot of shacks that were built using cheap building materials but, in general still were built in an organized manner in terms of urban design implying that there people in these areas do built their houses on assigned plots of land rather than just anywhere they find space to build. This pattern of structured neighbourhoods was visible throughout all of the clusters showing the effects of urban planning policies in the city. Some of the found clusters were quite similar in the appearance of their built environment which shows that algorithms were quite sensitive to small differences in the input images. One thing that stood out was the presence of mixed built environments in many of the clusters. Mixed built environments, which contain elements of both formal and informal architecture and urban design. These are a common feature in developing cities and can be seen as a transitional stage between more informal and more formal built environments. The found typologies showed that for Johannesburg this is also true as a large part of the houses in the city showed this type of mixed built environments.

In the eight clusters we found five distinct types of built environments and one cluster that contained a mix of everything. These found typologies were:

- Structured shacks, in this typology shacks were built on designated plots in a structured manner.
- Unstructured shacks, in this typology shacks were built scattered throughout an area.
- Formal housing with backyard shacks, in this typology, built environments were found with developed housing that had shacks in the backyard.

- Mixed built environments, in this typology, neighbourhoods had alternating houses of cheaper materials and further developed housing.
- High end houses, in this typology, large houses were include in spacious neighbourhoods with lots of trees.
- Mix of everything, this outlier cluster included both the lower end as well as the higher end houses which means that this cluster had some issues.

That some of the clusters showed similar built environment styles shows that there is room for improvement in this methodology where further analysis of these clusters can help identifying if there are nuanced differences between these similar built environments or whether these cluster should actually be joined together. Also further inspection could reveal that there are suppressed built environment types hidden in the clusters that did not appear in our evaluation of the clusters. Overall, this study contributes to our understanding of the built environment typologies that are present in the City of Johannesburg.

6.1.4 Sub-question 4 *What is the spatial distribution of built environments in the City of Johannesburg?* The spatial distribution of built environments in the City of Johannesburg was analyzed in this study using the eight clusters found in the study. The results of this analysis showed that the clusters are distributed unevenly throughout the city, with some clusters being concentrated in certain areas and others being more widely dispersed. These patterns of spatial distribution were found to be similar to the patterns of informal settlements and gated communities in the city. The clusters with less developed built environments mostly found in the south, while the highly developed areas were concentrated in the northern part of the city. Historically, the northern part of the city had more economic possibilities than the south which resulted in a higher level of development in the north. The found clusters were compared to 10 year old data on informal settlements and gated communities which are the extremes in terms of development of the built environment within Johannesburg and there were clusters that showed a lot of overlap with these areas. However, assuming that the clusters are homogeneous internally, it also showed new areas where informal settlements could be located. This overlap of the clusters with the old data shows that there is likely to be a correlation with the found clusters and the spatial structures that exist in the city.

6.2 Main research question

What knowledge can be gained from inferring built environment types from satellite images using unsupervised machine learning algorithms? This research demonstrates the feasibility of using unsupervised remote sensing techniques to identify and classify different built environment typologies in the City of Johannesburg. By combining high resolution satellite images with remotely sensed data on building locations, we were able to create clusters of the built environment in an unsupervised way. These clusters were not all perfectly distinct but do provide a good starting point for understanding the characteristics and distribution of different types of built environments in the city.

One important aspect of this research is the need for human interpretation to give context and meaning to the clusters that were identified. While the methodology is able to identify different built environment typologies, it does not provide labels for these typologies and further human

analysis is needed to understand the specific characteristics of all of the clusters. In addition to providing insights into the built environment, this method can be used for identifying informal settlements.After contextualizing the clusters, we were able to identify locations in the city that have a low level of development of the built environment which is sign of an area being an informal settlement.

Finally, the range of different built environment typologies identified in this study can provide insights into the spatial structures and segregation patterns in the city. By understanding the distribution of different types of built environments, it is possible to get a general idea to which extent different built environment types are segregated from each other throughout the city. This information can be useful for policymakers and planners that want to address issues of inequality and segregation in the city.

6.3 Future research

As urbanization is an ongoing process that is believed to intensify in the coming years, the need for information on urbanizing areas also increases. This research has provided a methodology to remotely sense information on the built environment in a city and the spatial distribution of different types of the built environment. This research has shown that this type of unsupervised classification has potential for further exploration. At first, this research can be replicated on different urban areas to see if this methodology also works in different parts of the world. Also, a next step would be to test the capabilities of this methodology for the detection of informal settlements by testing it on an area where more recent and accurate information on the location of informal settlements is available and how it compares to supervised methods.

Another avenue in future research is to improve the neural network used in this research as it was a relatively small neural network compared to main stream image analysis algorithms that have more than 10 times more parameters and achieve better image reconstruction results on general image dataset. This could allow for more details to be captured in the latent representation and possibly a better distinction of clusters.

As the current state of this research is specific to a single city, it lacks a way to compare different cities with each other in an objective way. Therefor, future research could involve a way to calculate a segregation score of the city based on the distribution and isolation of different built environment types. Also, a next step in this research would be to spatially group houses from a similar type together in order to be able to identify hotspots of built environment types within a city rather than just having data on individual houses. All in all, this research has shown feasibility of unsupervised built environment classification in an urbanizing region and it could be useful to investigate how the found information can be extended to allign with the needs of urban planners.

References

Abascal, A., Rothwell, N., Shonowo, A., Thomson, D. R., Elias, P., Elsey, H., ... Kuffer, M. (2022). "domains of deprivation framework" for mapping slums, informal settlements, and other deprived areas in lmics to improve urban planning and policy: A scoping review. *Computers, Environment and Urban Systems*, *93*, 101770. Retrieved from https://www.sciencedirect.com/science/article/pii/S019897152200014X doi: https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2022.101770

Ajami, A., Kuffer, M., Persello, C., & Pfeffer, K. (2019). Identifying a slums' degree of deprivation from vhr images using convolutional neural networks. *Remote Sensing*. doi: https://doi.org/10.3390/rs11111282

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (icet)* (pp. 1–6).

Aljalbout, E., Golkov, V., Siddiqui, Y., Strobel, M., & Cremers, D. (2018). Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648*.

Alkire, S. (2007). The missing dimensions of poverty data: Introduction to the special issue. *Oxford development studies*, *35*(4), 347–359.

Alkire, S., & Robles, G. (2017). Multidimensional poverty index summer 2017: Brief methodological note and results. *OPHI Methodological Notes*, *45*.

Ansari, R. A., & Buddhiraju, K. M. (2019). Textural segmentation of remotely sensed images using multiresolution analysis for slum area identification. *European Journal of Remote Sensing*, *52*(sup2), 74–88.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, *27*(2), 93–115.

Arribas-Bel, D., Patino, J. E., & Duque, J. C. (2017). Remote sensing-based measurement of living environment deprivation: Improving classical approaches with machine learning. *PLoS one*, *12*(5), e0176684.

Baatiema, L., Skovdal, M., Rifkin, S., & Campbell, C. (2013). Assessing participation in a community-based health planning and services programme in ghana. *BMC health services research*, *13*(1), 1–13.

Bähr, J., & Jürgens, U. (2006). Johannesburg: life after apartheid. In *Cities in transition* (pp. 175–208). Springer.

Ballard, R., & Hamann, C. (2021). Income inequality and socio-economic segregation in the city of johannesburg. In *Urban socio-economic segregation and income inequality* (pp. 91–109). Springer, Cham.

Baud, I. S., Pfeffer, K., Sridharan, N., & Nainan, N. (2009). Matching deprivation mapping to urban governance in three indian mega-cities. *Habitat International*, *33*(4), 365–377.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. , 437–478.

Bessell, S. (2015). The individual deprivation measure: Measuring poverty as if gender and inequality matter. *Gender & Development*, *23*(2), 223–240.

Bremner, L. J. (2000). Post-apartheid urban geography: a case study of greater johannesburg's rapid land development programme. *Development Southern Africa*, *17*(1), 87–104.

Brownlee, J. (2019). *Deep learning for computer vision: image classification, object detection, and face recognition in python.* Machine Learning Mastery.

Cabrera-Barona, P., & Ghorbanzadeh, O. (2018). Comparing classic and interval analytical hierarchy process methodologies for measuring area-level deprivation to analyze health inequalities. *International Journal of Environmental Research and Public Health*, *15*(1), 140.

Charlotn, S. (2014). Public housing in johannesburg. In *Changing space, changing city: Johannesburg after apartheid - open access selection* (pp. 176–193). Wits University Press. Retrieved 2023-02-13, from http://www.jstor.org/stable/10.18772/22014107656.13

Charlton, S., & Meth, P. (2017). Lived experiences of state housing in johannesburg and durban. *Transformation: Critical Perspectives on Southern Africa*, *93*(1), 91–115.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). Nbclust: an r package for determining the relevant number of clusters in a data set. *Journal of statistical software*, *61*, 1–36.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the european conference on computer vision (eccv)* (pp. 801–818).

Division, U. N. S. (2022). *UNSD — Demographic and Social Statistics.* Retrieved from https://unstats.un.org/unsd/demographic-social/census/

dos Santos, B. D., de Pinho, C. M. D., Oliveira, G. E. T., Korting, T. S., Escada, M. I. S., & Amaral, S. (2022). Identifying precarious settlements and urban fabric typologies based on geobia and data mining in brazilian amazon cities. *Remote Sensing*, *14*(3), 704.

Dos Santos, S., Adams, E., Neville, G., Wada, Y., De Sherbinin, A., Bernhardt, E. M., & Adamo, S. (2017). Urban growth and water access in sub-saharan africa: Progress, challenges, and emerging research directions. *Science of the Total Environment*, *607*, 497–508.

Dovey, K., & Kamalipour, H. (2017, 9). Informal/formal morphologies. *Mapping Urbanities*, 223-248. Retrieved from https://www.taylorfrancis.com/books/9781315309163/chapters/10.4324/9781315309163-13 doi: https://doi.org/10.4324/9781315309163-13

D'Alessandro, D., & Appolloni, L. (2020). Housing and health: An overview. *Annali di Igiene*, *32*(5), 17–26.

Ellena, M., Breil, M., & Soriani, S. (2020). The heat-health nexus in the urban context: A systematic literature review exploring the socio-economic vulnerabilities and built environment characteristics. *Urban Climate*, *34*, 100676.

ESRI. (2022). *"world imagery" [basemap].* Retrieved from https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9

Friesen, J., Taubenböck, H., Wurm, M., & Pelz, P. F. (2018). The similar size of slums. *Habitat International*, *73*, 79–88.

Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow.* " O'Reilly Media, Inc.".

Gevaert, C., Persello, C., Sliuzas, R., & Vosselman, G. (2016). Classification of informal settlements through the integration of 2d and 3d features extracted from uav data. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, *3*, 317.

Ghaffarian, S., & Emtehani, S. (2021). Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery. *Climate*, *9*(4). Retrieved from https://www.mdpi.com/2225-1154/9/4/58 doi: https://doi.org/10.3390/cli9040058

Giri, C. P. (2012). *Remote sensing of land use and land cover: principles and applications.* CRC press.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodman, A., & Gatward, R. (2008). Who are we missing? area deprivation and survey participation. *European journal of epidemiology*, *23*(6), 379–387.

Google. (2022). *Open buildings v2.* Retrieved from https://sites.research.google/open-buildings/

Gouverneur, D. (2014). *Planning and design for future informal settlements: shaping the self-constructed city*. Routledge.

Gram-Hansen, B. J., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V., & Bilinski, P. (2019). Mapping informal settlements in developing countries using machine learning and low resolution multi-spectral data. , 361–368.

Gränzig, T., Fassnacht, F. E., Kleinschmit, B., & Förster, M. (2021). Mapping the fractional coverage of the invasive shrub ulex europaeus with multi-temporal sentinel-2 imagery utilizing uav orthoimages and a new spatial optimization approach. *International Journal of Applied Earth Observation and Geoinformation*, *96*, 102281.

Gulli, A., & Pal, S. (2017). *Deep learning with keras*. Packt Publishing Ltd.

Guo, X., Liu, X., Zhu, E., & Yin, J. (2017). Deep clustering with convolutional autoencoders. In *International conference on neural information processing* (pp. 373–382).

Habitat, U. (2016). Urbanization and development: emerging futures. *World cities report*, *3*(4), 4–51.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504–507.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hofmann, P., et al. (2001). Detecting informal settlements from ikonos image data using methods of feature oriented image analysis-an example from cape town (south africa). *J¨urgens, C.(Ed.): Remote Sensing of Urban Areas/Fernerkundung in urbanen R¨aumen*, 41–42.

Hofmann, P., Taubenböck, H., & Werthmann, C. (2015). Monitoring and modelling of informal settlements-a review on recent developments and challenges. *2015 joint urban remote sensing event (JURSE)*, 1–4.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4700–4708).

Ibrahim, M. R., Titheridge, H., Cheng, T., & Haworth, J. (2019). predictslums: A new model for identifying and predicting informal settlements and slums in cities from street intersections using machine learning. *Computers, Environment and Urban Systems*, *76*, 31–56.

Inostroza, L. (2017). Informal urban development in latin american urban peripheries. spatial assessment in bogotá, lima and santiago de chile. *Landscape and Urban Planning*, *165*, 267–279. doi: https://doi.org/10.1016/j.landurbplan.2016.03.021

Joburg Municipality. (2022). *GeoLIS.* Retrieved from https://ags.joburg.org.za/cgismobi/

Keeley, B., Little, C., & Zuehlke, E. (2019). The state of the world's children 2019: Children, food and nutrition–growing well in a changing world. *UNICEF*.

Ketkar, N., & Santana, E. (2017). *Deep learning with python* (Vol. 1). Springer.

Khalifa, M. A., & Connelly, S. (2009). Monitoring and guiding development in rural egypt: local sustainable development indicators and local human development indices. *Environment, Development and Sustainability*, *11*(6), 1175–1196.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kit, O., & L¨udeke, M. (2013). Automated detection of slum area change in hyderabad, india using multitemporal satellite imagery. *ISPRS journal of photogrammetry and remote sensing*, *83*, 130–137.

Kit, O., Lüdeke, M., & Reckien, D. (2012). Texture-based identification of urban slums in hyderabad, india using remote sensing data. *Applied Geography*, *32*(2), 660–667.

Klasen, S. (2000). Measuring poverty and deprivation in south africa. *Review of income and wealth*, *46*(1), 33–58.

Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing*, *8*(6), 455.

Kuffer, M., Thomson, D. R., Boo, G., & Mahabir, R. (2020). The role of earth observation in an integrated deprived area mapping "system" for low-to-middle income countries. *Remote sensing*, *12*(6), 982. doi: https://doi.org/C:7

Lai, D. (2000, September). Temporal Analysis of Human Development Indicators: Principal Component Approach. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, *51*(3), 331-366. Retrieved from https://ideas.repec.org/a/spr/soinre/v51y2000i3p331-366.html doi: https://doi.org/10.1023/A:1007065804509

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Leonita, G., Kuffer, M., Sliuzas, R., & Persello, C. (2018). Machine learning-based slum mapping in support of slum upgrading programs: The case of bandung city, indonesia. *Remote sensing*, *10*(10), 1522.

Madhavan, S., Beguy, D., Clark, S., & Kabiru, C. (2018). Measuring extended families over time in informal settlements in nairobi, kenya: Retention and data consistency in a two-round survey. *Demographic research*, *38*, 1339.

Mahabir, R., Croitoru, A., Crooks, A. T., Agouris, P., & Stefanidis, A. (2018). A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Science*, *2*(1), 8.

Manaswi, N. K. (2018). Understanding and working with keras. In *Deep learning with applications using python* (pp. 31–43). Springer.

Marshall, S. (2009). *Cities, design and evolution*. Routledge.

McGranahan, G., Schensul, D., & Singh, G. (2016). Inclusive urbanization: Can the 2030 agenda be delivered without it? *Environment and Urbanization*, *28*(1), 13–34.

Microsoft. (2022). *Global ml building footprints.* Retrieved from https://github.com/microsoft/GlobalMLBuildingFootprints

Mitlin, D., & Satterthwaite, D. (2012). *Urban poverty in the global south: scale and nature*. Routledge.

Mitlin, D., & Satterthwaite, D. (2013). Addressing deprivations in urban areas–. , 266–298.

Molina-García, J., Queralt, A., Adams, M. A., Conway, T. L., & Sallis, J. F. (2017). Neighborhood built environment and socio-economic status in relation to multiple health outcomes in adolescents. *Preventive medicine*, *105*, 88–94.

Musango, J. K., Currie, P., Smit, S., & Kovacic, Z. (2020). Urban metabolism of the informal city: Probing and measuring the 'unmeasurable'to monitor sustainable development goal 11 indicators. *Ecological Indicators*, *119*, 106746.

Myers, G. (2011). *African cities: Alternative visions of urban theory and practice*. Bloomsbury Publishing.

Myers, G. (2021). Urbanisation in the global south. , 27–49.

Nations, U. (2022). *Goal 11* Department of Economic and Social Affairs. *Retrieved from https://sdgs.un.org/goals/goal11*

Onodugo, V. A., & Ezeadichie, N. H. (2019). Future planning of global south cities with inclusive informal economic growth in perspective. In Sustainability in urban planning and design. *IntechOpen.*

Owen, K. K., & Wong, D. W. (2013). An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics. Applied Geography, 38, *107-118. Retrieved from https://www.sciencedirect.com/science/article/pii/S0143622812001592 doi: https://doi.org/https://doi.org/10.1016/j.apgeog.2012.11.016*

Palmer, R. C., Ismond, D., Rodriquez, E. J., & Kaufman, J. S. (2019). Social determinants of health: future directions for health disparities research *(Vol. 109) (No. S1). American Public Health Association.*

Pelleg, D., Moore, A. W., et al. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In Icml *(Vol. 1, pp. 727–734).*

Plazas, M., Ramos-Pollán, R., & Martínez, F. (2021). Ensemble-based approach for semisupervised learning in remote sensing. Journal of Applied Remote Sensing, 15*(3), 034509.*

Prabhu, R., Parvathavarthini, B., & Alagu Raja, R. (2021). Slum extraction from high resolution satellite data using mathematical morphology based approach. International Journal of Remote Sensing, 42*(1), 172–190.*

Rios, M. (2014). Learning from informal practices: Implications for urban design. The informal American city: Beyond taco trucks and day labor, *173–191.*

Robinson, J. (2013). Ordinary cities: between modernity and development. *Routledge.*

Roy, A. (2005). Urban informality: toward an epistemology of planning. Journal of the american planning association, 71*(2), 147–158.*

Roy, A. (2011). Slumdog cities: Rethinking subaltern urbanism. International journal of urban and regional research, 35*(2), 223–238.*

Sahriman, N., Abiden, M. Z. Z., Rasam, A. R. A., Tarmizi, N. M., et al. (2013). Urban poverty area identification using high resolution satellite imagery: A preliminary correlation study. In 2013 ieee international conference on control system, computing and engineering *(pp. 430–434).*

Samper, J., Shelby, J. A., & Behary, D. (2020). The paradox of informal settlements revealed in an atlas of informality: Findings from mapping growth in the most common yet unmapped forms of urbanization. Sustainability, 12*(22), 9510.*

Satterthwaite, D., Sverdlik, A., & Brown, D. (2019). Revealing and responding to multiple health risks in informal settlements in sub-saharan african cities. Journal of urban health, 96*(1), 112–122.*

Singleton, A., Arribas-Bel, D., Murray, J., & Fleischmann, M. (2022a). Estimating generalized measures of local neighbourhood context from multispectral satellite images using a convolutional neural network. Computers, Environment and Urban Systems, 95, *101802. doi: https://doi.org/10.1016/j.compenvurbsys.2022.101802*

Singleton, A., Arribas-Bel, D., Murray, J., & Fleischmann, M. (2022b, 7). Estimating generalized

measures of local neighbourhood context from multispectral satellite images using a convolutional neural network. Computers, Environment and Urban Systems, 95, *101802. doi: https://doi.org/10.1016/J.COMPENVURBSYS.2022.101802*

St Amand, F. (2014). Identification of slums in mumbai, india: Unsupervised classification techniques.

Stark, T., Wurm, M., Zhu, X. X., & Taubenböck, H. (2020). Satellite-based mapping of urban poverty with transfer-learned slum morphologies. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, *5251-5263. doi: https://doi.org/10.1109/JSTARS.2020.3018862*

Statistics South Africa. (2011). 2011 Census Statistics South Africa. Retrieved from https://www.statssa.gov.za/?page$_id = 3839$

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2818–2826).

Totaforti, S. (2020). Urban planning in post-apartheid south african cities: The case of johannesburg. *Open Journal of Political Science*, *10*(3), 507–520.

United Nations. (2016). The new urban agenda - habitat iii. The United Nations conference on housing and sustainable urban development (Habitat III) held in Quito, Ecuador.

United Nations. (2019). World urbanization prospects the 2018 revision.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, *11*(12).

Webster, D. (2019, 2). *New inclusionary housing policy shakes up Joburg.* Retrieved from https://obs.org.za/cms/wp-content/uploads/2019/09/New-inclusionary-housing-policy-shakes-up-Joburg.pdf

Wilkinson, R. G., Pickett, K., et al. (2009). *The spirit level: Why more equal societies almost always do better* (Vol. 6). Allen Lane London.

Williams, T. K.-A., Wei, T., & Zhu, X. (2020). Mapping urban slum settlements using very high-resolution imagery and land boundary data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 166-177. doi: https://doi.org/10.1109/JSTARS.2019.2954407

A Appendix A: Convolution Neural Network

A.1 Research flow

To guide the research, a research flow diagram has been designed and is presented in Figure 1. The diagram gives a short overview of the structure of the research, the different phases and what to expect in each phase.

Fig. 23. Research flow

A.2 Low resolution CNN Autoencoder

```
Model: "sequential"
Optimizer='adam', metrics=accuracy, loss=mean_squared_error
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 16, 16, 16)        1040
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 16, 16, 16)        4112
_________________________________________________________________
batch_normalization (BatchNo (None, 16, 16, 16)        64
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 8, 8, 16)          4112
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 8, 8, 16)          4112
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 8, 8, 4)           1028
_________________________________________________________________
batch_normalization_1 (Batch (None, 8, 8, 4)           16
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 8, 8, 4)           260
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 8, 8, 1)           65
_________________________________________________________________
batch_normalization_2 (Batch (None, 8, 8, 1)           4
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 8, 8, 1)           17
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 8, 8, 4)           68
_________________________________________________________________
batch_normalization_3 (Batch (None, 8, 8, 4)           16
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 8, 8, 4)           260
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 8, 8, 16)          1040
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 8, 8, 16)          4112
_________________________________________________________________
up_sampling2d (UpSampling2D) (None, 16, 16, 16)        0
_________________________________________________________________
conv2d_12 (Conv2D)           (None, 16, 16, 16)        4112
_________________________________________________________________
conv2d_13 (Conv2D)           (None, 16, 16, 16)        4112
_________________________________________________________________
batch_normalization_4 (Batch (None, 16, 16, 16)        64
_________________________________________________________________
conv2d_14 (Conv2D)           (None, 16, 16, 4)         68
=================================================================
Total params: 28,682
Trainable params: 28,600
Non-trainable params: 82
_________________________________________________________________
```

Fig. 24. CNN autoencoder design for images with a 10 meter per pixel resolution with 64 features.

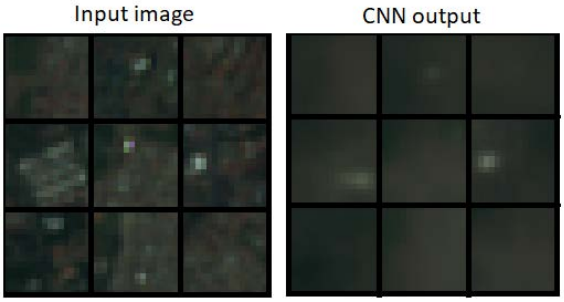

Fig. 25. Input and predicted images

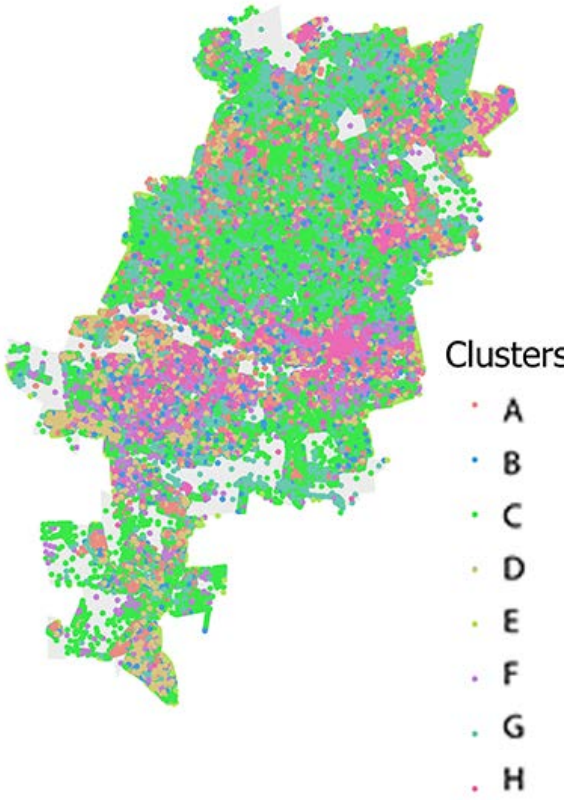

Fig. 26. Initial clustering results using low resolution CNN algorithm

A.3 High resolution CNN Autoencoder with limited features

```
Model: "sequential"
Optimizer='adam', metrics=accuracy, loss=mean_squared_error

_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)              (None, 64, 64, 64)        1792

 conv2d_1 (Conv2D)            (None, 64, 64, 64)        36928

 batch_normalization (BatchN  (None, 64, 64, 64)       256
 ormalization)

 conv2d_2 (Conv2D)            (None, 32, 32, 64)        36928

 conv2d_3 (Conv2D)            (None, 32, 32, 64)        36928

 conv2d_4 (Conv2D)            (None, 32, 32, 3)         1731

 batch_normalization_1 (Batc  (None, 32, 32, 3)         12
 hNormalization)

 conv2d_5 (Conv2D)            (None, 16, 16, 64)        1792

 conv2d_6 (Conv2D)            (None, 16, 16, 64)        36928

 conv2d_7 (Conv2D)            (None, 16, 16, 3)         1731

 batch_normalization_2 (Batc  (None, 16, 16, 3)         12
 hNormalization)

 conv2d_8 (Conv2D)            (None, 8, 8, 64)          1792

 conv2d_9 (Conv2D)            (None, 8, 8, 64)          36928

 conv2d_10 (Conv2D)           (None, 8, 8, 3)           1731

 batch_normalization_3 (Batc  (None, 8, 8, 3)           12
 hNormalization)

 conv2d_11 (Conv2D)           (None, 8, 8, 3)           84

 conv2d_12 (Conv2D)           (None, 8, 8, 1)           28

 batch_normalization_4 (Batc  (None, 8, 8, 1)           4
 hNormalization)

 conv2d_13 (Conv2D)           (None, 8, 8, 3)           30

 conv2d_14 (Conv2D)           (None, 8, 8, 1)           28

 batch_normalization_5 (Batc  (None, 8, 8, 1)           4
 hNormalization)

 conv2d_15 (Conv2D)           (None, 8, 8, 1)           10

 conv2d_16 (Conv2D)           (None, 8, 8, 3)           30
```

Fig. 27. CNN autoencoder design for images with a 1 meter per pixel resolution with 64 features, part 1.

```
batch_normalization_6 (Batc  (None, 8, 8, 3)          12
hNormalization)

conv2d_17 (Conv2D)           (None, 8, 8, 3)          84

conv2d_18 (Conv2D)           (None, 8, 8, 64)         1792

conv2d_19 (Conv2D)           (None, 8, 8, 64)         36928

up_sampling2d (UpSampling2D  (None, 16, 16, 64)       0
)

conv2d_20 (Conv2D)           (None, 16, 16, 3)        1731

conv2d_21 (Conv2D)           (None, 16, 16, 64)       1792

conv2d_22 (Conv2D)           (None, 16, 16, 64)       36928

up_sampling2d_1 (UpSampling  (None, 32, 32, 64)       0
2D)

conv2d_23 (Conv2D)           (None, 32, 32, 64)       36928

conv2d_24 (Conv2D)           (None, 32, 32, 64)       36928

up_sampling2d_2 (UpSampling  (None, 64, 64, 64)       0
2D)

conv2d_25 (Conv2D)           (None, 64, 64, 64)       36928

conv2d_26 (Conv2D)           (None, 64, 64, 64)       36928

batch_normalization_7 (Batc  (None, 64, 64, 64)       256
hNormalization)

conv2d_27 (Conv2D)           (None, 64, 64, 3)        195

=================================================================
Total params: 423,149
Trainable params: 422,865
Non-trainable params: 284
```

Fig. 28. CNN autoencoder design for images with a 1 meter per pixel resolution with 64 features, part 2.

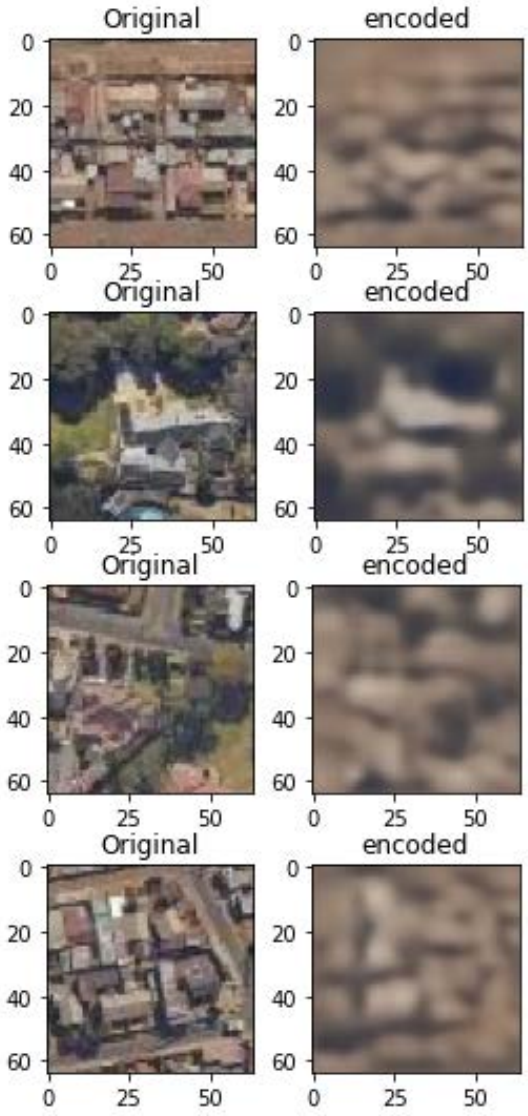

Fig. 29. Input and predicted images

A.4 High resolution CNN Autoencoder with more features

```
Model: "sequential"
Optimizer='adam', metrics=accuracy, loss=mean_squared_error
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)              (None, 64, 64, 64)        1792

 conv2d_1 (Conv2D)            (None, 64, 64, 64)        36928

 batch_normalization (BatchN  (None, 64, 64, 64)       256
 ormalization)

 conv2d_2 (Conv2D)            (None, 32, 32, 64)        36928

 conv2d_3 (Conv2D)            (None, 32, 32, 64)        36928

 conv2d_4 (Conv2D)            (None, 32, 32, 3)         1731

 batch_normalization_1 (Batc  (None, 32, 32, 3)         12
 hNormalization)

 conv2d_5 (Conv2D)            (None, 16, 16, 64)        1792

 conv2d_6 (Conv2D)            (None, 16, 16, 64)        36928

 conv2d_7 (Conv2D)            (None, 16, 16, 3)         1731

 batch_normalization_2 (Batc  (None, 16, 16, 3)         12
 hNormalization)

 conv2d_8 (Conv2D)            (None, 16, 16, 3)         84

 conv2d_9 (Conv2D)            (None, 16, 16, 1)         28

 batch_normalization_3 (Batc  (None, 16, 16, 1)         4
 hNormalization)

 conv2d_10 (Conv2D)           (None, 16, 16, 1)         10

 conv2d_11 (Conv2D)           (None, 16, 16, 3)         30

 batch_normalization_4 (Batc  (None, 16, 16, 3)         12
 hNormalization)

 conv2d_12 (Conv2D)           (None, 16, 16, 3)         84

 conv2d_13 (Conv2D)           (None, 16, 16, 64)        1792

 conv2d_14 (Conv2D)           (None, 16, 16, 64)        36928

 up_sampling2d (UpSampling2D  (None, 32, 32, 64)        0
 )

 conv2d_15 (Conv2D)           (None, 32, 32, 64)        36928

 conv2d_16 (Conv2D)           (None, 32, 32, 64)        36928

 up_sampling2d_1 (UpSampling  (None, 64, 64, 64)        0
```

Fig. 30. CNN autoencoder design for images with a 1 meter per pixel resolution with 256 features, part 1.

```
 2D)

 conv2d_17 (Conv2D)          (None, 64, 64, 64)        36928

 conv2d_18 (Conv2D)          (None, 64, 64, 64)        36928

 batch_normalization_5 (Batc  (None, 64, 64, 64)       256
 hNormalization)

 conv2d_19 (Conv2D)          (None, 64, 64, 3)         195

=================================================================
Total params: 342,173
Trainable params: 341,897
Non-trainable params: 276
```

Fig. 31. CNN autoencoder design for images with a 1 meter per pixel resolution with 256 features, part 2.

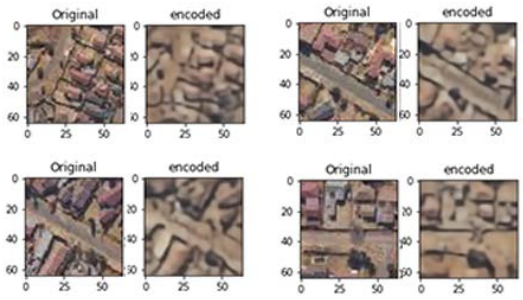

Fig. 32. Input and predicted images

B Appendix B: Cluster classifications

B.1 Cluster overview

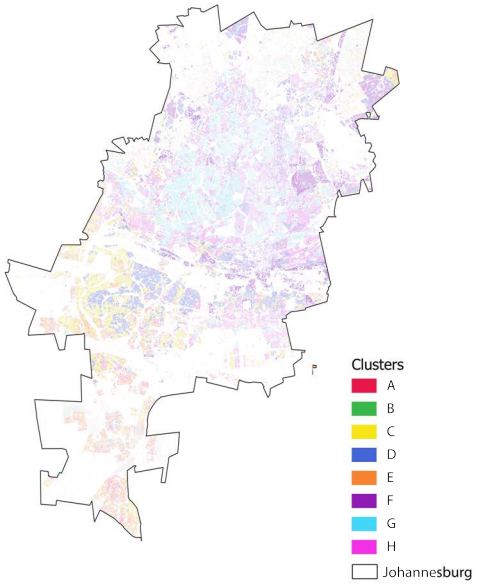

Fig. 33. Overview of the found clusters in original form showing a faded image

B.2 Cluster A

Fig. 34. Cluster A sample images with classification of a mixed architecture and a formal urban design

Fig. 35. Cluster A sample images with classification of a informal architecture and a formal urban design

B.3 Cluster B

Fig. 36. Cluster B sample images with classification of a informal architecture and a formal urban design

Fig. 37. Cluster B sample images with classification of a informal architecture and a informal urban design

B.4 Cluster C

Fig. 38. Cluster C sample images with classification of a mixed architecture and a formal urban design, part 1

Fig. 39. Cluster C sample images with classification of a mixed architecture and a formal urban design, part 2

B.5 Cluster D

Fig. 40. Cluster D sample images with classification of a formal architecture and a formal urban design

Fig. 41. Cluster D sample images with classification of a informal architecture and a informal urban design

Fig. 42. Cluster D sample images with classification of a mixed architecture and a formal urban design

Fig. 43. Cluster D sample images with classification of a mixed architecture and a mixed urban design

B.6 Cluster E

Fig. 44. Cluster E sample images with classification of a formal architecture and a mixed urban design

Fig. 45. Cluster E sample images with classification of a informal architecture and a informal urban design

Fig. 46. Cluster E sample images with classification of a mixed architecture and a formal urban design

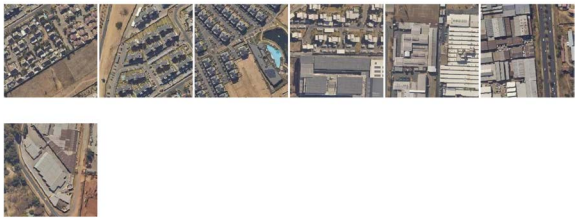

Fig. 47. Cluster E sample images with classification of a formal architecture and a formal urban design

B.7 Cluster F

Fig. 48. Cluster F sample images with classification of a formal architecture and a informal urban design

Fig. 49. Cluster F sample images with classification of a informal architecture and a formal urban design

Fig. 50. Cluster F sample images with classification of a informal architecture and a informal urban design

Fig. 51. Cluster F sample images with classification of a informal architecture and a mixed urban design

Fig. 52. Cluster F sample images with classification of a mixed architecture and a formal urban design

B.8 Cluster G

Fig. 53. Cluster F sample images with classification of a formal architecture and a formal urban design

Fig. 54. Cluster F sample images with classification of a formal architecture and a mixed urban design

Fig. 55. Cluster F sample images with classification of a informal architecture and a mixed urban design

B.9 Cluster H

Fig. 56. Cluster H sample images with classification of a formal architecture and a formal urban design

Fig. 57. Cluster H sample images with classification of a formal architecture and a mixed urban design

Fig. 58. Cluster H sample images with classification of a informal architecture and a informal urban design

Fig. 59. Cluster H sample images with classification of a mixed architecture and a mixed urban design