Leveraging the Potential of Convolutional Neural Network and Satellite Images to Map Informal Settlements in Urban Settings of the City of Kigali, Rwanda

Esaie Dufitimana[1,2 *] and Theophile Niyonzima[3]

1. Quantum Leap Africa research center of African Institute for Mathematical Sciences-AIMS Rwanda, KN3 Kigali, Rwanda
2. School of Information Communication Technology (SoICT), College of Science and Technology of the University of Rwanda, P. O. Box 3900, Kigali, Rwanda
3. Department of Spatial Planning, School of Architecture and Built Environment, College of Science and Technology of the University of Rwanda, P. O. Box 3900, Kigali, Rwanda

*Corresponding Author: edufitimana@quantumleapafrica.org

Abstract

The urban population is rapidly increasing and more than half of the world's population is currently living in cities. About 1 billion of city dwellers are living in slums and informal settlements. Addressing the issue of slums and informal settlements require information on these areas. This study explored the potential of Convolutional Neural Network (CNN) and Very-High Resolution satellite image to map the informal settlements in urban areas of the city of Kigali, Rwanda. The study applied modified U-Net model with MobileNetV2 model as the base model to discriminate areas with informal settlements from other areas. The model was obtained by modifying the original U-Net architecture to incorporate dilated convolutional operations at the beginning of the network. The findings demonstrate that based on the spatial characteristics of informal settlements, the model was able to detect informal settlements in urban areas with a recall of 0.862, a precision of 0.810 and an F1-Score of 0.809. Based on these results, the study can be a basis for finding relevant information about informal settlements that are of concern in the implementation of SDGs, especially goal 11 addressing the issues of safe and inclusive cities and human settlements.

Keywords: Convolutional Neural Network, informal settlements, slums, VHR images

1. Introduction

1.1. Background of the Study

The urban population is rapidly increasing and more than half of the world's population is currently living in cities (United Nations, 2018), which require a well-planned and arranged urban development process to accommodate the population. The increase in urban population has been one of the global challenges over the last decades. Cities, especially in developing countries, have not managed to deal with the challenge caused by urban population growth

https://dx.doi.org/10.4314/rjeste.v5i1.2

(Mahabir et al., 2016). The high increase in urban population has exceeded the capacity of developing countries in urban planning, provision of adequate and affordable housing, and infrastructure provision to most urban dwellers. At present, about 1 billion of turban population is living in deprivation (UN-Habitat, 2016). Deprivation is regarded as a lack of benefits considered to be basic necessities in a society and is linked to the way people live and work (Baud, Sridharan, & Pfeffer, 2008). Most of the urban population that lacks such necessities live in urban deprived areas, known as slums and/or informal settlements (Friesen, et al., 2020; Kuffer et al., 2020). Failure to adopt adequate planning policies or poor implementation of existing ones for inclusive cities force high proportions of urban dwellers to live in those slums and informal settlements.

Slums and informal settlements often exhibit physical and socio-economic features characterized by lack of basic infrastructure, poor housing, overcrowding, unhealthy living conditions, poverty, social exclusion, and tenure insecurity (UN-Habitat, 2003). On the other hand, solutions for slums and informal settlements are difficult to implement due to a lack of reliable information on the status of existing challenges. The common source of geo information has been over the last decades, manual image digitization. However, the information from this source is outdated, time and resources consuming for its creation and presents discrepancy of the reality on the ground due to methods applied (Kuffer et al., 2018). The lack of information about slums and informal settlements is a challenge for monitoring the implementation of goal 11 of the Sustainable Development Goals (SDGs) aiming at making safe and inclusive cities and human settlements (United Nations, 2015). To address the aforementioned challenges, there is a need for other methods that can provide useful and updated information.

In Kigali city, the largest and capital city of Rwanda, the implementation of goal 11 of SDGs is embedded in urban policies redevelopment aimed, among others, at coping with the growth of informal settlements across the city, which occupy about 65% of the built-up residential areas of the entire city (Manirakiza et al., 2019). Of these policies, the master plan aims to improve urban dwellers' living conditions by ensuring adequate land use, provision of modern housing solutions, provision of adequate transportation, and a healthy environment (City of Kigali, 2019; (Uwayezu, and de Vries, 2020). Its implementation involves property redevelopment arrangements that require demolition of existing and inadequate housing facilities in slums and informal settlements characterized by low-quality building materials (Hitayezu et al., 2018). Thus, there is a need for adequate information about slums and informal settlements to support the implementation of urban redevelopment.

Timely and adequate information about slums and informal settlements is recognized as crucial for implementing urban redevelopment policies (UN-Habitat, 2018). Furthermore, it provided a guideline for ensuring justice and equal opportunities for poor and low-income urban dwellers, who predominantly reside in slums and informal settlements, to live and benefit from cities (Chigbu et al., 2019). In this regard, information about slums and informal settlements status unravel the situation concerning spatial and social (in)justice that arises from the implementation of urban development policies (Moroni, 2018). The need for up-to-date information about slums and informal settlements and the availability of Very High Resolution (VHR) earth observation images have triggered the use of earth observation as an alternative way to source information that helps understand the physical aspects of slums and informal settlements (Kuffer et al.,

https://dx.doi.org/10.4314/rjeste.v5i1.2

2016). Earth observation can enable to capture detailed and enough information for spatially identifying and characterizing these areas (Kuffer et al., 2018; Mahabir et al., 2016). Earlier researches demonstrated that spatial information derived from earth observation has the potential to facilitate the understanding of the nature, evolution and growth of slums and informal settlements (Wang et al., 2019; Arribas-Bel et al., 2017; Kohli et al., 2016; Kuffer et al., 2017; Mboga et al., 2017), and their environment conditions (Friesen et al., 2020; Georganos et al., 2019).

Earth observation has shown the advantages of detecting high spatial and temporal characteristics of urban deprived areas, useful for effectively monitoring and tracking urban deprived areas' growth and providing essential information to support urban development policies formulation and implementation (Mahabir et al., 2016). However, despite the effectiveness of earth observation-based information for spatially characterizing and understanding slums and informal settlements, there is a need to explore further methods for extracting adequate information about slums and informal settlements in the global south. In this regard, exploring new methods can bridge data scarcity about slums and informal settlements. Therefore, this research explores the potential of VHR earth observation image to provide information about slums and informal settlements based on their spatial characteristics. The following section will present slums, informal settlement and deprivation, followed by a section on the techniques for slum and informal settlements mapping. Then, the sections on method used and results obtained will follow ending with sections on the discussion and recommendations.

1.2. Slums, Informal Settlements and Urban Deprivation

The term deprivation is conceptualized in many different ways, but in general, it is often used to indicate the lack of a wide range of necessities in a society (Baud et al., 2008). The lack of necessities in life is often used interchangeably with poverty in some social studies and denote slums and informal settlements. However, urban scientists use the term deprivation instead of poverty, and it goes beyond poverty. Deprivation depicts multiple social, environmental (built and natural) and ecological aspects that significantly impact dwellers of urban deprived areas (Kuffer et al., 2020). It is crucial to note that some aspects are specific for deprivation while others are not (Thomson et al., 2019). For instance, environmental and ecological aspects such as flood plain and slope can also exist in non-deprived areas, whereas poor sanitation and inadequate housing are specific to deprived areas (Uwayezu, and de Vries, 2020). In a more general sense, deprivation represents multi-dimensional factors that describe sub-standard socio-economic and physical living conditions (UN-Habitat, 2015; Lilford et al., 2019). Researchers group these factors in domains to elaborate more comprehensive frameworks for measuring urban deprivation. For instance, Thomson et al. (2019) described a framework that groups deprivation factors in five domains: social and environmental risk, lack of facilities, unplanned urbanization, contamination, and tenure insecurity. Very recently, the research by Abascal et al. (2021) conceptualizes deprivation in the so-called Domains of Deprivation Framework. It encompasses concepts resulting from the review of various literatures on deprivation alongside the validation by experts.

The Domains of Deprivation Framework allows a detailed understanding of deprivation, and the data for all its dimensions play a crucial role in its understanding (Abascal et al., 2021). However, its multi-dimensionality requires a broad range of data that is not easily available or obtainable. (Lilford et al., 2019) recognized three essential methods for obtaining data about

https://dx.doi.org/10.4314/rjeste.v5i1.2

deprivation: household surveys for socio-economic data, ground surveys for identifying physical features, and earth observation to identify physical features. Though surveys provide reliable data, they are time-consuming, economically inefficient and do not reflect the fast-changing pace of urban deprived areas. Earth observation is an important source of up-to-date information in deprived areas. It is more efficient for capturing spatial and temporal physical properties of urban deprived areas. Moreover, earth observation alongside advanced machine learning and deep learning methods have shown the potential to capture socio-economic information. In this context, it is claimed that the physical appearance of deprived areas reflects the socio-economic status of their dwellers (Arribas-Bel et al., 2017; Duque et al., 2015).

1.3. Spatial Characteristics of Urban Deprived Areas

Urban deprived areas present physical aspects such as poor structural quality of housing, overcrowding, high building density, solid waste management, environmental conditions (such as proximity to the wetland and steep slopes), and inadequate access to infrastructure such as roads, water supply, electricity (UN-Habitat, 2015). Different researches (Lilford et al., 2019; Kuffer et al., 2018; Wurm & Taubenböck, 2018; Taubenböck & Kraff, 2014) describe urban deprived areas as areas characterized by high building densities, small buildings, irregular arrangement of buildings and street networks as well as their location, which, in some cases is high-risk zones. The physical characteristics of deprived areas differ across locations and can be explained differently depending on the context (Kuffer et al., 2020). However, urban deprived areas present similar morphological characteristics. Kohli et al. (2012) conceptualized these morphological characteristics in three levels, namely object, settlement and environs, in the generic of slum ontology. Object-level represents characteristics such as building roof type, footprint, shape and orientation. The settlement level presents characteristics such as the irregular shape of roads and building density. Lastly, environs level presents characteristics linked to the location, such as proximity to hazardous places, such as flooding areas, wetland areas and steep slopes. Lilford et al. (2019) have extended the concept of slum ontology by including more physical characteristics to describe urban deprived areas. They suggested that deprived areas can be characterized based on their built environment, ecology, and services, but the main characteristics remain the same as presented in slum ontology.

Different researches studying urban deprivation have identified the spatial characteristics of urban deprived areas that can be extracted using earth observation methods. Those characteristics serve as guidance for efficient use of earth observation to produce reliable information on deprived areas. In this regard, this research gathered spatial characteristics of deprived areas based on a series of literature on deprivation and categorized them in three main categories adapted from Lilford et al. (2019). Table 1 illustrates a compilation of spatial characteristics of urban deprived areas.

Table 1: Spatial characteristics of deprived areas from literature

Category	Characteristic	Research	Comment
Built environment	▪ Small buildings; ▪ High building density; ▪ Irregular building layouts;	(Wurm & Taubenböck, 2018), (Lilford et al., 2019),	Image analysis techniques are used to extract these characteristics from VHR remote sensing images.

https://dx.doi.org/10.4314/rjeste.v5i1.2

	▪ Poor roofing materials; ▪ Low road coverage, ▪ Presence of unpaved roads; ▪ Irregular road networks; ▪ Lack of access to electricity; ▪ Lack of streets lightning	(Kohli et al., 2012), (Kuffer et al., 2018), (Kuffer et al., 2020), (Taubenböck & Kraff, 2014).	Mapping night light from VHR remote sensing image indicate the existence of electricity services in areas
Ecology	▪ Gradient and altitude (floodplain, steep slope for landslides, other hazards) ▪ Green spaces ▪ Air quality	(Graesser et al., 2012), (Kohli et al., 2016), (Georganos et al., 2019),	The use of multispectral remote sensing images and the Digital Elevation Model (DEM) to get information on topographic conditions of areas can reveal the possibility of disasters such as landslides or flood.
Services	▪ Presence of open sewers and solid waste disposal	(Arribas-Bel et al., 2017), (Friesen et al., 2020)	Detected from VHR remote sensing images and reveal the unhealthy living conditions.

1.4. Techniques for Mapping Slums and Informal Settlements

Researchers have employed different methods and algorithms for detecting spatial characteristics of slums and informal settlements from VHR remote sensing images and aerial images. The methods such as Object-Based Image Analysis (OBIA) and machine learning have been used to detect spatial characteristics of areas from VHR remote sensing images (Kohli et al., 2016). OBIA segments input image into several homogeneous contiguous groups before classification (Blaschke, 2010). OBIA approaches have shown capabilities of detecting both area and object-based information in slums and informal settlements mapping (Kuffer et al., 2016). Random Forest, a machine learning approach, has shown the potential to achieve relatively high performance for urban mapping (Sun et al., 2017). The Support Vector Machine, which is also a machine learning method, was shown to have improved accuracy for detecting features from earth observation data (Huang & Zhang, 2013). Though the above-mentioned methods have gained more credit for detecting spatial characteristics of areas from VHR remote sensing images, they present challenges for detecting spatial characteristics of complex spatial contextual and texture features (Sameen et al., 2018) and intra-class spectral variability (Chen et al., 2014) mostly in case of mapping slums and informal settlements due to their morphology. Moreover, these methods fail to distinguish different urban structures in VHR images and this is challenging because of the abstract semantic definition of the classes as opposed to the separation of standard

https://dx.doi.org/10.4314/rjeste.v5i1.2

land-cover classes (Persello & Stein, 2017). Therefore, the increase in VHR images' availability and the challenges associated with the commonly used methods have left researchers' room to explore and propose new approaches for detecting spatial characteristics of areas from VHR remote sensing images.

There is a trend in applying deep learning methods for mapping slums and informal settlements since they have shown the advantage of automatically learning and extracting spatial features from images with high accuracy than the commonly used methods (Bergado et al., 2016). Deep learning models have gained popularity for analyzing remote sensing images. They originated from Artificial Neural Networks developed as an advance in perceptron, and they consist of three categories of layers: the input layer, hidden layers, and output layers (Lecun et al., 2015). Deep learning models are built to learn from the known data and predict the other data based on what they learnt. Typically, Convolutional Neural Networks (CNNs), a type of deep learning models, have gained more popularity in processing image through image classification and semantic segmentation, which is the classification that occurs at a pixel level of an image (Sameen et al., 2018a). CNNs are built of one or more convolutional layers made of sliding filters over the input. The advantages of CNNs are that they are found in different architectures, and they allow flexibility for modification whereby the layers can be modified according to the user's need (Indolia, et al., 2018).

In the context of image processing, deep learning through CNNs had demonstrated its potential for remote sensing image analysis. Since their breakthrough in semantic segmentation, known as image classification in remote sensing (Shelhamer et al., 2017), different researchers have applied CNNs with different architectures to analyze VHR earth observation and aerial images in urban studies. Researchers (e.g., Ajamie et al., 2019; Wurm & Taubenböck, 2018; Kuffer et al., 2017; Persello & Stein, 2017) employed deep learning methods to map slums and informal settlements. These researchers illustrated that application of deep learning result in more relevant spatial information compared to other methods. Deep learning methods have illustrated the advantage of classifying the image into some predefined labels at higher accuracy than commonly used methods such as Support Vector Machine (SVM) and Random Forest (Sameen et al., 2018). The advantage of deep learning methods over other commonly used method is their capacity to learn spatial contextual information from image and produce more accurate results (Bergado et al., 2016). Moreover, the research by Mboga et al. (2017) explored the performance CNNs for detecting informal settlement from VHR images, and they compared their performance to that of SVM and found out that CNN was able to lead to better classification accuracy. (Wang et al., 2019) explored CNN's performance through U-Net architecture for identifying pockets of deprivation and found that the U-Net architecture used was able to detect and map the distribution and variation of deprivation. Sameen et al. (2018) benefited from the flexibility of CNNs and designed the architecture that could classify VHR aerial photo with high accuracy.

2. Materials and Methods

2.1. Study Area

Given the conditions of Kigali City in terms of undergoing city redevelopment strategies, the research has identified three research areas: Gatsata, Gitega and Kimisagara. These areas were

https://dx.doi.org/10.4314/rjeste.v5i1.2

purposively selected because they are known to be dominated with informal settlements and present major characteristics of slum areas such as high buildings density, small buildings and limited access to roads and they are among priority areas for urban transformation in accordance to the implementation of the master plan (City of Kigali, 2019). Figure 1 below show the location of these areas.

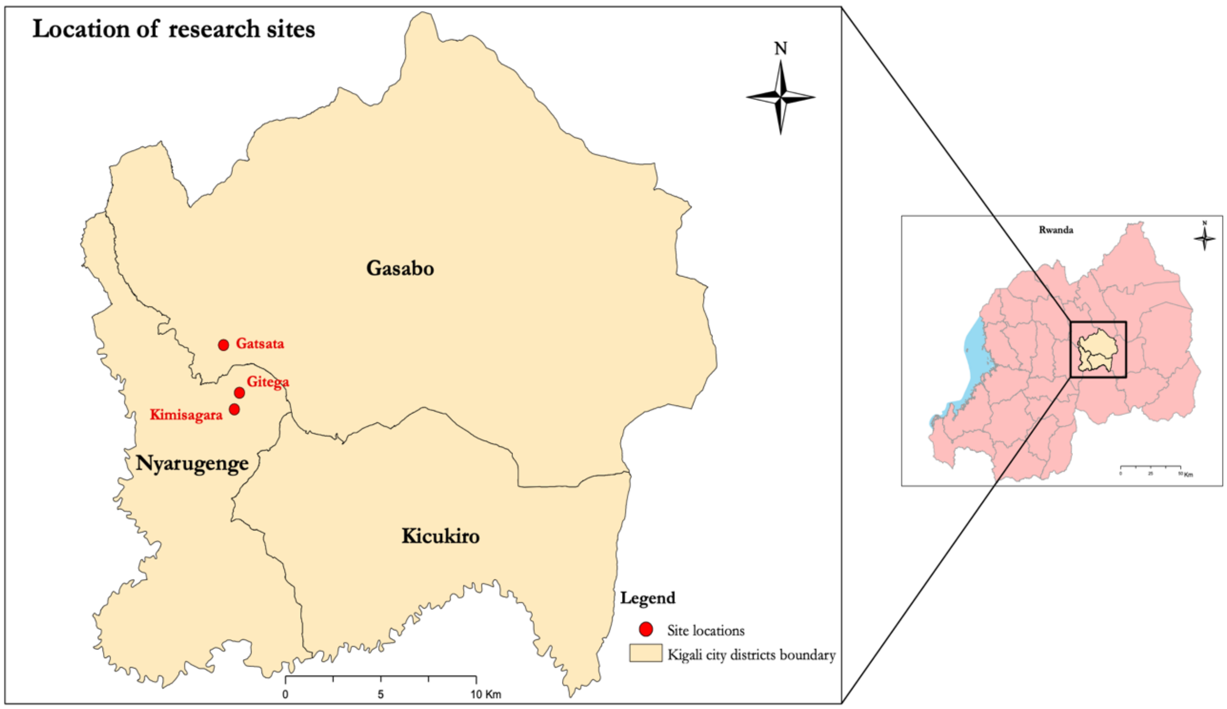


Figure 1: Localization of the study areas: Gatsata, Gitega and Kimisagara in Kigali city

2.2. Data

The research employed a VHR Google Earth (GE) image of Kigali city, which was downloaded with enough zoom level similar to VHR imagery with sub-meter pixel size. The image was downloaded using the SAS Planet tool, a free and open tool for downloading high-resolution satellite images from Google Earth (GE), Bing and Esri Imagery services (http://www.sasgis.org/sasplaneta/). The VHR GE image comprised of 3 visible bands: Red, Green and Blue. Though the VHR GE image quality may be lower than some VHR remote sensing imageries acquired by different commercial platforms, VHR GE images are freely available to the public (with respect to their terms and conditions). VHR GE images' availability gives an advantage for areas and cities with limited resources for purchasing standard VHR satellite images. For instance, VHR GE images have been used to study living environment deprivation in Liverpool, England (Arribas-Bel et al., 2017), to explore the potential of machine learning for automatic slum identification in Latine America (Duque et al., 2017), and to map squatter settlements in South Africa (Gunter, 2009). These studies illustrated that using VHR GE image is a good alternative to commercial VHR images for their respective applications.

Apart from the survey data and VHR GE image, the research used administrative boundaries sourced from the Institute of Statistics of Rwanda and informal settlements boundaries in Kigali

https://dx.doi.org/10.4314/rjeste.v5i1.2

city sourced from the Rwanda Land Management and Use Authority. The research also employed the data sourced from the literature, including the spatial characterization of deprived areas and physical proxies for characterizing perceived tenure insecurity in deprived areas and the model for detecting spatial characteristics of deprivation. Table 2 illustrates the data used for this research.

Table 2: Summary of the data used in the research

Data	**Source**	**Acquisition year**	**Specification**
Household data for 120 respondents and their locations	Field survey	2021	Characteristics of physical environment and perceptions of respondents on tenure insecurity + locations
Satellite image	Google Earth/ SAS Planet	2020	VHR GE image (RGB) with sub-meter pixel resolution (0.65m)
Administrative boundaries	National Institute of Statistics	2012	Shapefile
Informal settlements in Kigali city	Rwanda Land Management and Use Authority	2018	Shapefile

2.3. Methods

The CNN-based method for detecting spatial characteristics in the study area is based on a supervised VHR earth observation image classification. The CNN model learns from a set of ground truth data, also called training data, to predict and classify the image. Training data is a set of observations expected by CNN as examples of all classes to be classified (Chen et al., 2014). This means that the CNN model uses the training data as an example for classes, and hence when new data are passed, it predicts their classes based on what it learnt from the training data. The workflow for image classification for this project consists of 3 steps: data preparation, setting and training the model for image classification and model validation. The workflow is shown in figure 2 below.

https://dx.doi.org/10.4314/rjeste.v5i1.2

Rwanda Journal of Engineering, Science, Technology and Environment, Volume 5, Issue 1, June 2023
eISSN: 2617-233X | print ISSN: 2617-2321

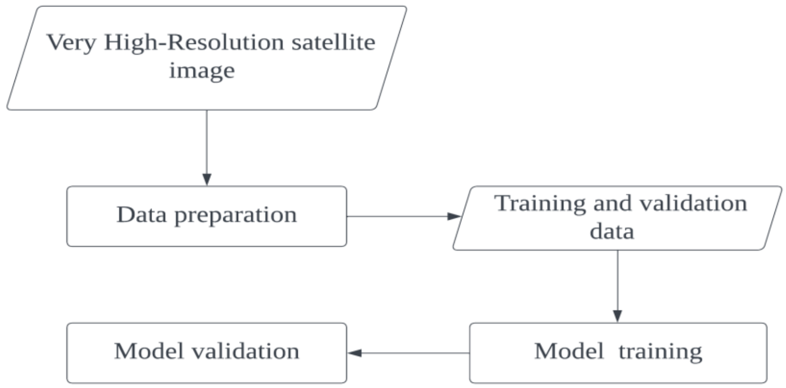


Figure 2: Workflow for remote sensing image classification (Source: authors)

2.3.1 Data Preparation

The study experimented on 14 tiles of 1500×1500 pixels for each of VHR GE images covering the three study sites. Since deep learning models require enough training data, there was a need for sufficient time to prepare labelled data to train our classification model. Training data were prepared delineated slum polygons obtained from the former Rwanda Land Management and Use Authority (RLMUA) currently Known as National Land Authority (NLA). These polygons were checked one by one on top of the available images (VHR GE images). Nevertheless, almost all polygons were not accurate enough to confidently be used for training a CNN model. This might be attributed to the temporal gap between the time the images were captured and the time the polygons were delineated. Moreover, this may have originated from the error in its creation. Thus, they were manually edited to match with the informal settlements in images. The obtained polygons were hence converted into raster format since the input of the classification model was in raster format as labels.

Python libraries used for data preparation include: Geospatial Data Abstraction Library (GDAL) for reading and writing raster and vector data, NumPy library for multi-dimensional arrays and matrices functions to operate these arrays, Matplotlib for plotting, OS library for interacting with data directory, Skimage for image processing and sklearn for splitting training data into train data and test data, and pyrsgis for exporting GeoTIFFS. Figure 3 below illustrate an example of one image tile on the left with its corresponding labels. Eight tiles were used for training the model, two for testing and the last four were reserved for inferencing (prediction on the new/unseen data).

https://dx.doi.org/10.4314/rjeste.v5i1.2

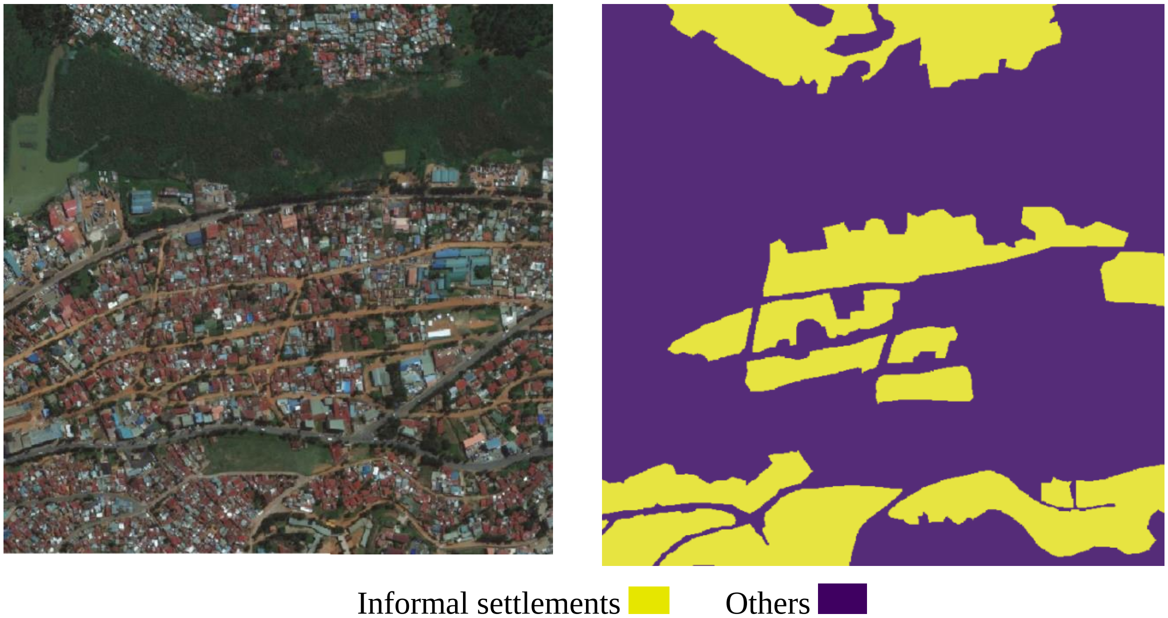


Figure 3: Input image tile (Left) and its corresponding label (Right)

2.3.2 The Model

This study adapted a Pixel-based image classification model that uses a Fully Convolutional Network (FCN) through U-Net architecture. The U-Net architecture, which was originally designed for analyzing medical images, has gained popularity in satellite image pixel classification, commonly known as semantic segmentation (Pham & Nguyen, 2020). In this study, a modified U-Net architecture developed by Wang et al. (2019) was employed for detecting and mapping informal settlements in the study area. The model was developed for detecting small pockets of deprivation. It was obtained by modifying the original U-Net architecture to incorporate dilated convolutional operations at the beginning of the network. These operations generated low-level feature maps at multiple scales, which helped prevent information loss during convolutional and max pooling operations. The modified U-Net architecture showed significant improvements in detecting small pockets of deprivation compared to the traditional FCN model of U-Net architecture. This was achieved by incorporating dilated convolutional operations, which allowed the network to capture more contextual information while still maintaining the resolution of the input image. Thus, it fits for the purpose of this study. The modified model architecture is shown in figure 4.

Libraries used for deep learning model include: Keras, TensorFlow and Hdf5 for saving deep learning model. Keras is an open-source library that provides a Python interface for artificial neural networks. Keras is a powerful and easy-to-use free open-source Python library for developing and evaluating deep learning models. It acts as an interface for the TensorFlow library. This combination of frameworks is a popular choice for deep learning tasks and provides a user-friendly interface to design, train, and evaluate deep learning models.

https://dx.doi.org/10.4314/rjeste.v5i1.2

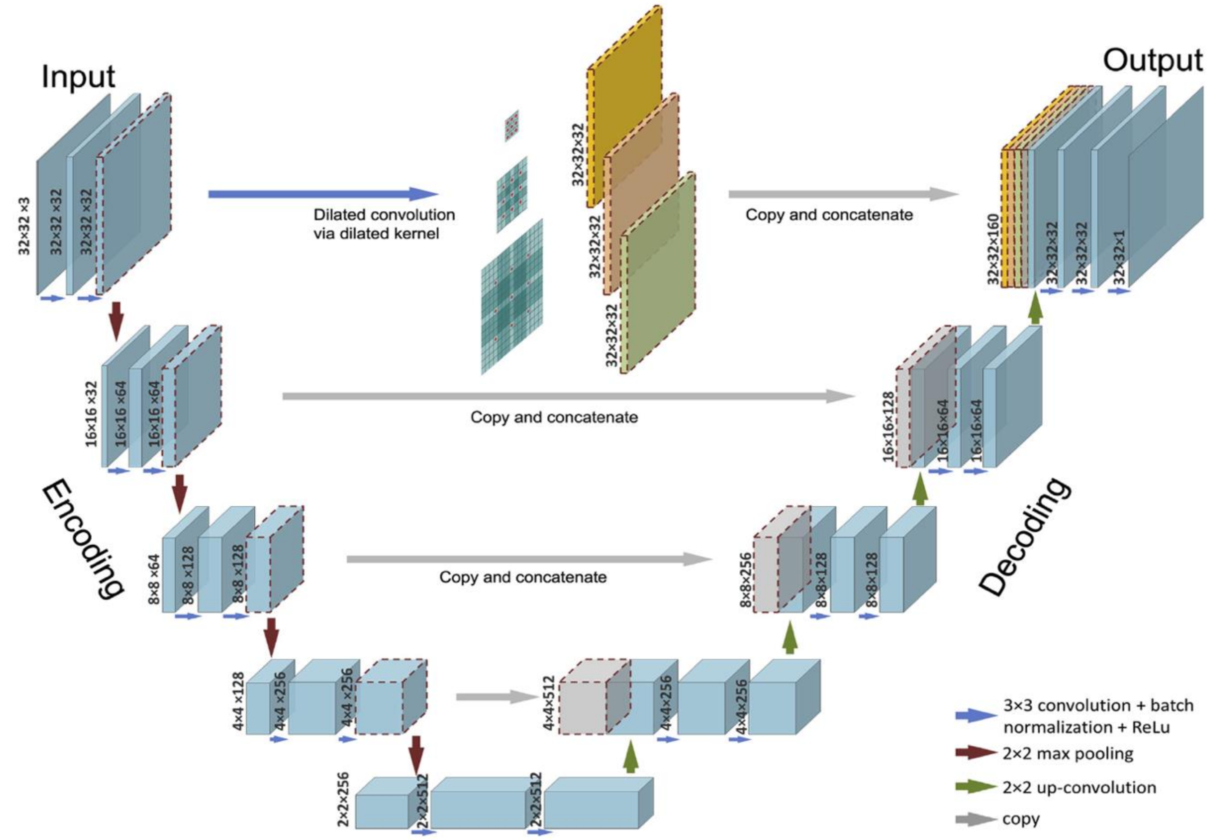


Figure 4: U-Net model architecture (Source: Wang et al., 2019)

The model starts by the section of codes defining the encoder part of the model. The model uses the MobileNetV2 model as the base model. The MobileNetV2 model is commonly used as the base model for U-Net due to its computational efficiency, good feature extraction capabilities, and pre-trained weights. It has a small number of parameters, making it ideal for low-power hardware platforms (Wang et al., 2019). The MobileNetV2 model is followed by the actual U-Net model shown in the figure 4 above. It starts by indicating the size of input images followed by convolutional layer with 16 filters each of size 3x3. The batch normalization layer is then used followed by a relu activation layer and then another convolutional layer with 16 filters of size 3x3, batch normalization layer, relu activation layer and a maxpooling layer of 2x2 size. It is important to know that the output of previous layer is used for the next layer as input. Blocks similar to the above are repeated 4 times with exponential growth of convolutional layers. The fifth block is also similar to the other with convolutional layers with 256 and 512 filters but does not have a maxpooling layer. The above-describes blocks building a part of the model called an encoder and then follows codes sections that made the second part of the model called decoder. In decoder parts, blocks are written by transposing convolutional layers. Means, the blocks in decoder part are duplicates of the encoder in opposite direction and hence use transpose as a transformation to achieve that. Batch normalization help the model to regularize and learn faster.

https://dx.doi.org/10.4314/rjeste.v5i1.2

The blocks are defined in a Python function defining the model. They are followed by blocks of functions for metrics that helps to access the behavior of the model. That block is followed by the block defining the optimization, calling the model function and model compilation to have the complete built U-Net model followed by a line of codes that visualize the model architecture. The line that compiles line of codes features a loss function which is a binary cross-entropy. This is due to the problem in hand which is binary classification whereby the image is classified in two classes (informal settlements and non-informal settlements). If the user prefers to implement multi-classes classification, the loss function should be replaced by the categorical cross-entropy. The functions for accessing model behavior do not play any role in learning but are only useful for assessment of the model performance.

2.3.3 Training and Validating the Model

Training data were split into two sets: train set counting for 80% of all training data and test set counting 20% of all training data. The training set was used for training the model, whereas the test set was used to evaluate the model's performance. The research used more training data than testing data to ensure that enough training data are fed to the model to identify the features in the input images. Image tiles and their corresponding labels were reduced into the size (patches) adequate for the model to train the model. The image patch size used for this research ranges from 64×64 and 128×128 pixels. The model was trained on 200 epochs. The patch size plays a significant role in computational cost in terms of the memory required and the model's accuracy. Reducing the image tiles into small patches help to deal with computational limitations. The implementation was evaluated through model accuracy on the test set and metrics such as precision, recall and F1-score. The recall, precision, and F1-Score are commonly used evaluation metrics for binary and multi-class classification tasks (Nivaggioli & Randrianarivo, 2019). Precision measures the ratio of correctly classified pixels to the total number of all classified pixels in the classified map. The recall measures the ratio of correctly classified pixels in a classified map to the total number of pixels in the reference map. Finally, the F1-score is obtained based on the precision and the recall value and indicates the model's performance by showing the harmonic value balancing both precision and recall. In other words, the recall measures the proportion of actual positives that are correctly identified by the model, while precision measures the proportion of predicted positives that are actually true positives. The F1-Score is the harmonic mean of precision and recall and provides a balance between the two metrics.

Training the model takes a long time due to the capacity of the computer being used and the size of the data. After the training process is completed. The model behavior is evaluated through plotting training curves (accuracy and loss) and the section of codes for evaluating the model performance on training set and test set. After being satisfied by the model performance, the last part is to use the model to predict/inference on the unseen image tiles for prediction. The first section of codes prepares the image tile to be fed into the model for prediction. The whole process was implemented in Python 3. As for the hardware specifications, the process took place on a personal computer with Intel(R) Core (TM) i7-9750H CPU at 2.60 GHz, a RAM of 16 GB, and dual GPU: Nvidia Quadro T1000 and Intel(R) UDH Graphics 630. These specifications were able to handle deep learning tasks on small to medium-sized datasets. Thus, the available computational resources would not allow the smooth performance on a bigger dataset. If the dataset is too large, it is necessary to use a more powerful computing platform such as a server or

https://dx.doi.org/10.4314/rjeste.v5i1.2

cloud-based service with advanced Graphic Processing Units (GPUs) to accelerate the training process, especially for computationally intensive deep learning tasks.

3. Results

The research relied on how different studies on deprivation conceptualize and spatially characterize urban deprived areas using information derived from earth observation. Typically, the research took advantage of the spatial characteristics of urban deprived areas. Training accuracy and training loss graph were used to understand the behavior of the model during training and guided the decisions on how the model's performance would be improved. The training accuracy, on the other hand, measures how often the model correctly predicts the output label compared to the actual label during training. The graphs in figure 5 and 6 below presenting training and testing accuracy show an increasing trend for the training and testing accuracy over the course of the training process. An increasing trend in training accuracy indicates that the model was learning to fit the training data better and was becoming more accurate in its predictions. On the other hand, an increasing trend in testing accuracy was a good sign that the model was generalizing well and was able to make accurate predictions on new, unseen data. This suggests that the model was not overfitting to the training data and was learning to identify the key features of informal settlements that are present in the testing data.

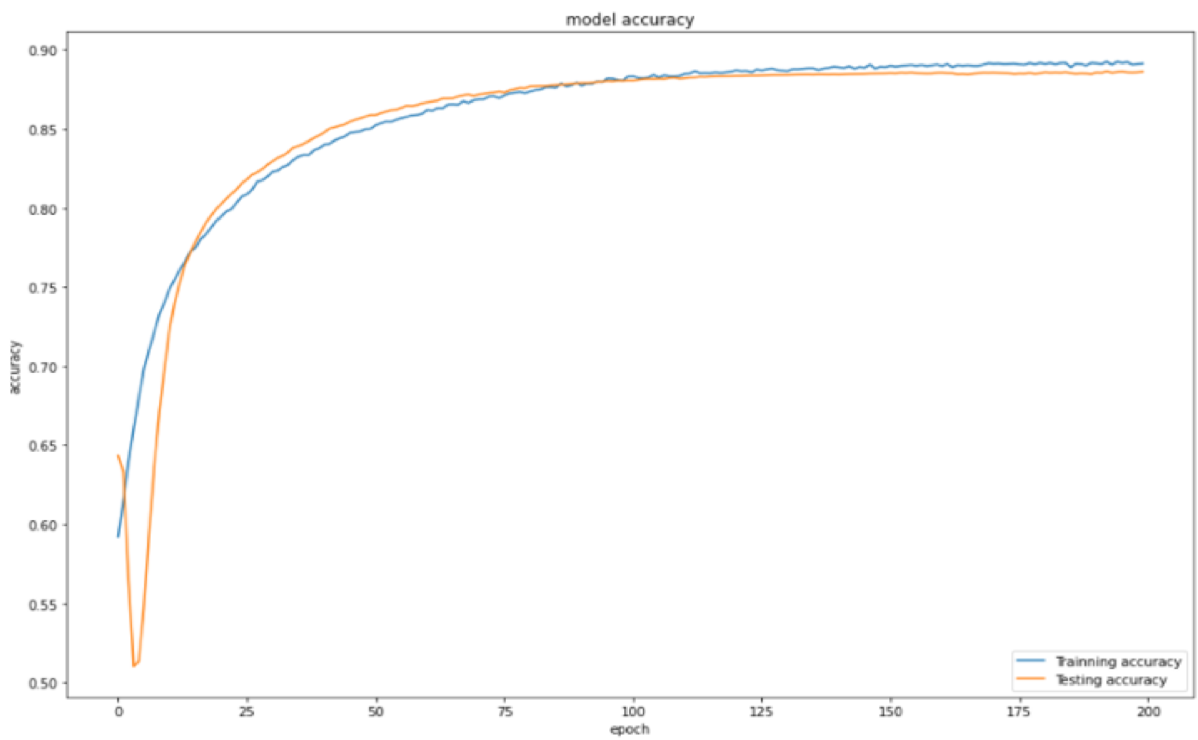


Figure 5: Training and testing accuracy for the dataset with 128 by 128 pixels size

https://dx.doi.org/10.4314/rjeste.v5i1.2

Rwanda Journal of Engineering, Science, Technology and Environment, Volume 5, Issue 1, June 2023
eISSN: 2617-233X | print ISSN: 2617-2321

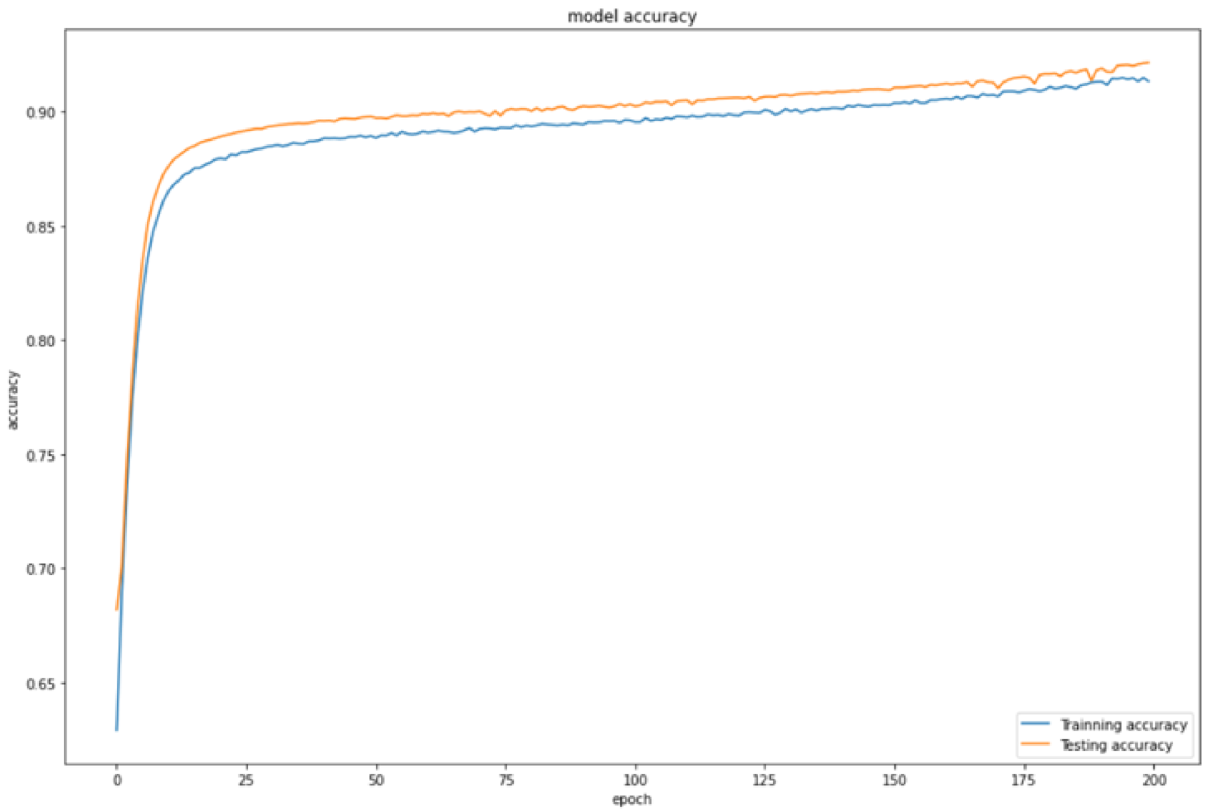


Figure 6: Training and testing accuracy for the dataset with 64 by 64 pixels size

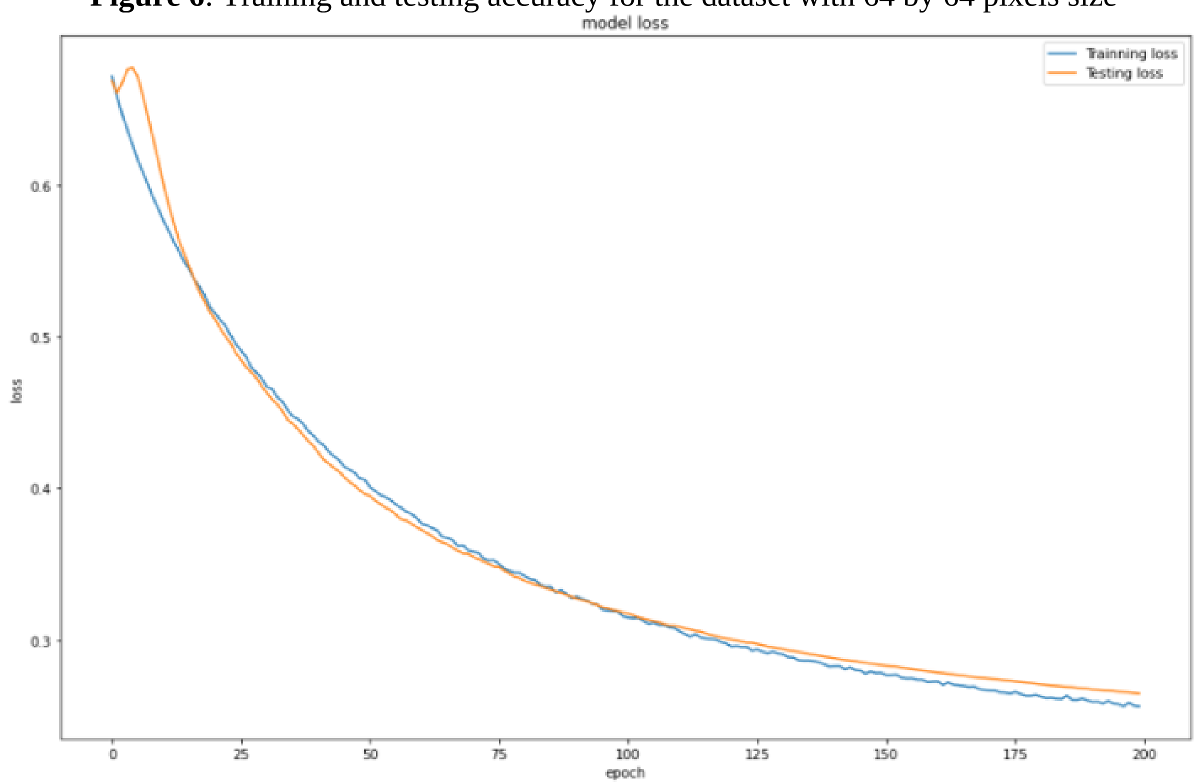


Figure 7: Training and testing loss for the dataset with 128 by 128 pixels size

https://dx.doi.org/10.4314/rjeste.v5i1.2

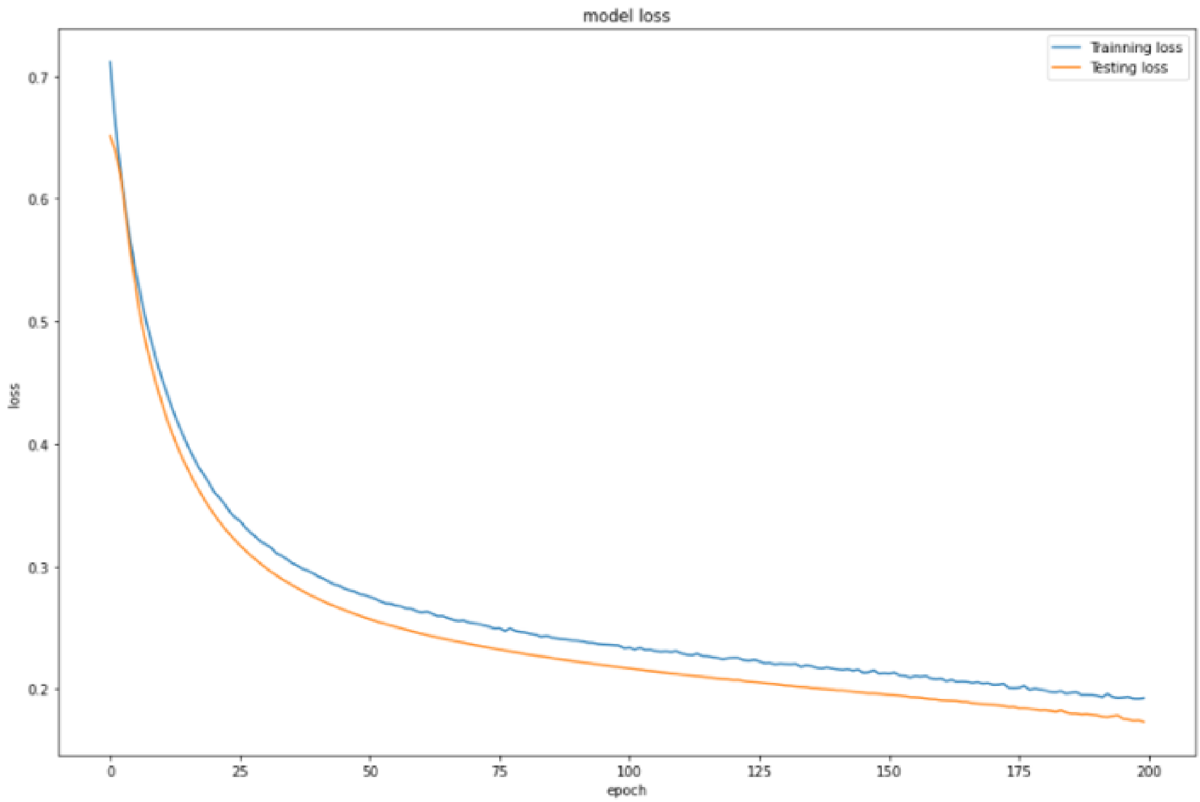


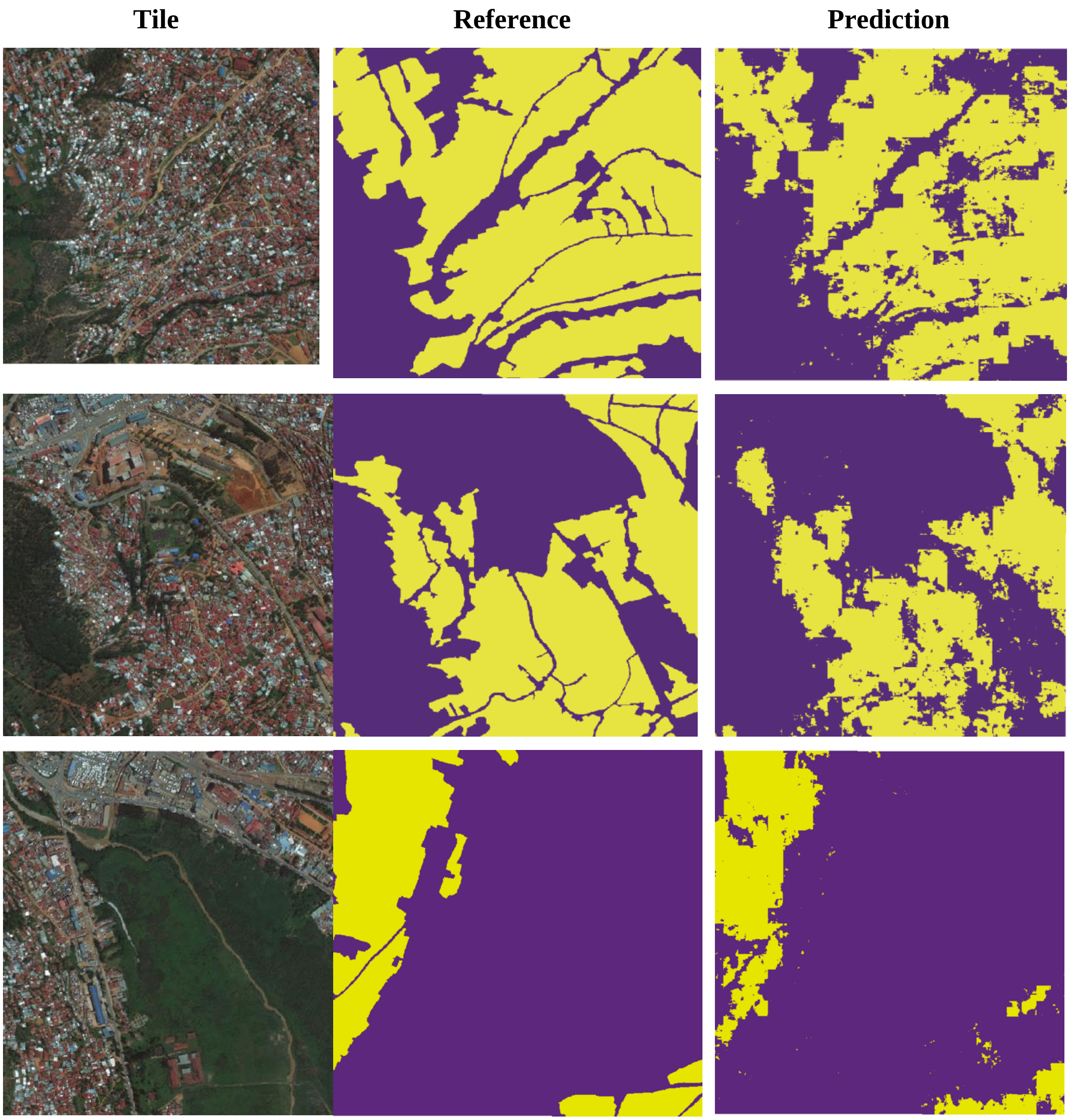
Figure 8: Training for the dataset with 64 by 64 pixels size

The graphs in figure 7 and 8 above presenting training and testing loss show a decreasing trend for the training loss during training process. The training loss, which is typically the mean squared error or cross-entropy loss, measures the difference between the predicted output of the model and the actual output. Training loss is the error that the model makes on the training data, while testing or validation loss is the error that the model makes on a separate test set of data. The test set is typically used to monitor the performance of the model during training and to avoid overfitting. Both the training loss and validation loss were decreasing, which means that the model was learning to generalize well and was not overfitting to the training data. The model trained on the patch size of dataset with 64 by 64 pixels size illustrated better training and loss accuracy compared to when trained on the image patches of 128 by 128 pixels size. The smaller patch size allowed the model to capture finer details and features in the images, resulting in better accuracy whereas the larger patch size led to more information loss and made it more difficult for the model to learn important features.

The model was evaluated on testing data. After tuning parameters for the model, the model trained on patch size of 64 by 64 pixels was found to have achieved more accuracy. The final accuracy obtained on the test set was 86.6%. The model achieved a recall of 0.862, precision of 0.810 and F1-Score of 0.809. The F1-Score of 0.809 suggests that the model was able to achieve a balance between precision and recall. The performance of the model was good, and the visual results are satisfactory given the quality of the input image and available data for training the model. Figure 8 presents the prediction results of the model four tiles that were reserved for

https://dx.doi.org/10.4314/rjeste.v5i1.2

prediction and were not used for neither training or testing the model. The results show that the model output prediction for classifying informal settlements in the study areas was quite good.



https://dx.doi.org/10.4314/rjeste.v5i1.2

Figure 8: Results of the model on prediction image tiles

4. Results Discussion

The proposed approach identified informal settlements in urban settings having spatial characteristics of slums. Thus, as shown by Lilford et al. (2019); Kuffer et al. (2018); Wurm & Taubenböck (2018), the CNN model applied in this study enabled to distinguish informal settlements from other land covers using VHR GE image. The model was able to learn, identify and segment the features of informal settlements in the input images. This is an important step towards building a reliable model for identifying and mapping informal settlements, which can have significant implications for urban planning and policy-making. The evaluation of the model on the testing data was a crucial step in assessing its accuracy and overall performance. After tuning the parameters for the model, it was found that the model trained on a patch size of 64 by 64 pixels achieved higher accuracy compared to other patch sizes tested. The final accuracy obtained on the test set was 86.6%. This was a significant achievement given the limitations of the input image used for the study. The model was able to accurately facilitate the classification of informal settlements in the study area, demonstrating its potential for application in slum mapping and deprivation research.

The recall of 0.862 indicates that the model was able to identify 86.2% of all positive cases correctly, while the precision of 0.810 suggests that the model was able to accurately identify 81% of positive cases. These results are promising and indicate that the model is robust and reliable in identifying informal settlements in the study area. The model's performance was considered good, and the visual results were satisfactory given the quality of the input image and the available data for training the model. The freely accessible VHR GE image used in this study has low quality compared to VHR images acquired from commercial providers. The limitations of the VHR GE image used in this study are worth exploring in more detail. While the image is freely accessible, it suffers from a number of quality issues that are not present in VHR images acquired from commercial providers. One major issue is the lack of spectral information, as the image only has visible bands (Red, Green and Blue), which can make it difficult for the CNN model to discriminate against certain features in the input image (Hu et al., 2013). In addition,

https://dx.doi.org/10.4314/rjeste.v5i1.2

the low quality of the image affects the accuracy of image classification, particularly in areas with shadow.

Despite these limitations, the CNN model still managed to achieve interesting results in the study. This is significant, given that most research on slum mapping and deprivation has relied on commercial VHR remote sensing images, which can be prohibitively expensive for many institutions and countries with limited financial resources (e.g., Wang et al., 2019; Kuffer et al., 2017; Tapiador et al., 2011; Bergado et al., 2016). The fact that the VHR GE image was freely available allowed this research to take advantage of an important resource that would otherwise have been inaccessible for institutions and countries with limited financial resources to use them. It is important to consider that the spatial characteristics of urban informal settlements are conceptualized depending on the place (Kuffer et al., 2020). Besides, the concepts of informal settlements and slums is linked to several events and regulations, which also differ from place to place. In this regard, the scalability and transferability of this research must focus on the overall method employed and considering spatial characteristics of informal settlements, which are location dependent. Therefore, methods used in this research CNN model can be transferred to other contexts by feeding them with new sample data. Though new CNN architecture or other types of machine learning models can be introduced, and tuning of their hyperparameters may differ from what is employed in this research, but the general analytical framework remains the same.

5. Conclusions

The research has applied the CNN model through U-Net architecture to detect informal settlements in the study area. The accuracy assessment of results needs to be related to the ability of CNN model used to distinguish area level features of informal settlements rather than pixel-based accuracy. Therefore, the results of the model are not pixel-based but is based on homogenous areas with similar spatial characteristics of slums and informal settlement often denoted as urban deprived areas. This study provides new evidence about the potential of Convolutional Neural Networks and earth observation to map informal settlements in urban settings. The study shows that application of CNN alongside satellite images can provide adequate information for supporting municipalities and stakeholders to address the issue of informal settlements. Furthermore, the research provides a basis for further researches concerning the spatial measurement of urban informal settlements toward the implementation of the SDGs, especially goal 11 for achieving safe, inclusive and sustainable cities and human settlements.

https://dx.doi.org/10.4314/rjeste.v5i1.2

References

Abascal, Á., Rothwell, N., Shonowo, A., Thomson, D. R., Elias, P., Elsey, H., Yeboah, G., & Kuffer, M. (2021). *"Domains of Deprivation Framework" for Mapping Slums, Informal Settlements, and other Deprived Areas in LMICs to improve urban planning and policy: A Scoping Review*. https://doi.org/10.20944/preprints202102.0242.v2

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). https://doi.org/10.1371/journal.pone.0176684

Baud, I., Sridharan, N., & Pfeffer, K. (2008). Mapping urban poverty for local governance in an Indian mega-city: The case of Delhi. *Urban Studies*, *45*(7), 1385–1412. https://doi.org/10.1177/0042098008090679

Bergado, J. R., Persello, C., & Gevaert, C. (2016). A deep learning approach to the classification of sub-decimetre resolution aerial images. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2016-Novem*, 1516–1519. https://doi.org/10.1109/IGARSS.2016.7729387

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep Learning-Based Classification of Hyperspectral Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2094–2107. https://doi.org/10.1109/JSTARS.2014.2329330

Chigbu, U., Alemayehu, Z., & Dachaga, W. (2019). Uncovering land tenure insecurities: tips for tenure responsive land-use planning in Ethiopia. In *Development in Practice* (Vol. 29, Issue 3). UN-Habitat. https://doi.org/10.1080/09614524.2019.1567688

City of Kigali. (2019). *Zoning regulations: Kigali Master Plan 2050*.

Duque, J., Patino, J., & Betancourt, A. (2017). Exploring the Potential of Machine Learning for Automatic Slum Identification from VHR Imagery. *Remote Sensing*, *9*(9), 895. https://doi.org/10.3390/rs9090895

Friesen, J., Friesen, V., Dietrich, I., & Pelz, P. F. (2020). Slums, Space, and State of Health—A Link between Settlement Morphology and Health Data. *International Journal of Environmental Research and Public Health*, *17*(6), 2022. https://doi.org/10.3390/ijerph17062022

Georganos, S., Gadiaga, A. N., Linard, C., Grippa, T., Vanhuysse, S., Mboga, N., Wolff, E., Dujardin, S., & Lennert, M. (2019). Modelling the Wealth Index of Demographic and Health Surveys Within Cities Using Very High-Resolution Remotely Sensed Information. *Remote Sensing*, *11*(21), 2543. https://doi.org/10.3390/rs11212543

Graesser, J., Cheriyadat, A., Vatsavai, R. R., Chandola, V., Long, J., & Bright, E. (2012). Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(4), 1164–1176. https://doi.org/10.1109/JSTARS.2012.2190383

Gunter, A. W. (2009). Getting it for free: Using Google Earth™ and Ilwis to map squatter settlements in Johannesburg. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *3*, 388–391. https://doi.org/10.1109/IGARSS.2009.5417784

Hitayezu, P., Rajashekar, A., & Stoelinga, D. (2018). *The dynamics of unplanned settlements in the City of Kigali* (Issue December 2018).

Hu, Q., Wu, W., Xia, T., Yu, Q., Yang, P., Li, Z., & Song, Q. (2013). Exploring the use of google earth imagery and object-based methods in land use/cover mapping. *Remote Sensing*, *5*(11), 6026–6042. https://doi.org/10.3390/rs5116026

Huang, X., & Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 257–272. https://doi.org/10.1109/TGRS.2012.2202912

Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. *Procedia Computer Science*, *132*, 679–688. https://doi.org/10.1016/j.procs.2018.05.069

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

Kohli, D., Sliuzas, R., & Stein, A. (2016). Urban slum detection using texture and spatial metrics derived from satellite imagery. In *Journal of Spatial Science* (Vol. 61, Issue 2, pp. 405–426). Mapping Sciences Institute Australia. https://doi.org/10.1080/14498596.2016.1138247

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

Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I., & van Maarseveen, M. (2017). Capturing the diversity of deprived areas with image-based features: The case of Mumbai. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040384

Kuffer, M., Thomson, D. R., Boo, G., Mahabir, R., Grippa, T., Vanhuysse, S., Engstrom, R., Ndugwa, R., Makau, J., Darin, E., de Albuquerque, J. P., & Kabaria, C. (2020). The role of earth observation in an integrated deprived area mapping "system" for low-to-middle income countries. *Remote Sensing*, *12*(6), 982. https://doi.org/10.3390/rs12060982

Kuffer, M., Wang, J., Nagenborg, M., Pfeffer, K., Kohli, D., Sliuzas, R., & Persello, C. (2018). The scope of earth-observation to improve the consistency of the SDG slum indicator. In *ISPRS International Journal of Geo-Information* (Vol. 7, Issue 11). MDPI AG. https://doi.org/10.3390/ijgi7110428

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. https://doi.org/10.1038/nature14539

Lilford, R., Kyobutungi, C., Ndugwa, R., Sartori, J., Watson, S. I., Sliuzas, R., Kuffer, M., Hofer, T., Porto De Albuquerque, J., & Ezeh, A. (2019). Because space matters: Conceptual framework to help distinguish slum from non-slum urban areas. *BMJ Global Health*, *4*(2), 1267. https://doi.org/10.1136/bmjgh-2018-001267

Mahabir, R., Crooks, A., Croitoru, A., & Agouris, P. (2016). The study of slums as social and physical constructs: challenges and emerging research opportunities. *Regional Studies, Regional Science*, *3*(1), 399–419. https://doi.org/10.1080/21681376.2016.1229130

Manirakiza, V., Mugabe, L., Nsabimana, A., & Nzayirambaho, M. (2019). *City Profile: Kigali, Rwanda*. https://doi.org/10.1177/0975425319867485

Mboga, N., Persello, C., Bergado, J. R., & Stein, A. (2017). Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sensing*, *9*(11). https://doi.org/10.3390/rs9111106

Moroni, S. (2018). Property as a human right and property as a special title. Rediscussing private ownership of land. *Land Use Policy*, *70*, 273–280. https://doi.org/10.1016/j.landusepol.2017.10.037

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

Pham, T. M., & Nguyen, T. T. L. (2020). Optimization of resource management for NFV-enabled IoT systems in edge cloud computing. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2017.DOI

Sameen, M. I., Pradhan, B., & Aziz, O. S. (2018). Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks. *Journal of Sensors*, *2018*. https://doi.org/10.1155/2018/7195432

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.1109/TPAMI.2016.2572683

Sun, X., Lin, X., Shen, S., & Hu, Z. (2017). High-Resolution Remote Sensing Data Classification over Urban Areas Using Random Forest Ensemble and Fully Connected Conditional Random Field. *ISPRS International Journal of Geo-Information*, *6*(8), 245. https://doi.org/10.3390/ijgi6080245

Tapiador, F. J., Avelar, S., Tavares-Corrêa, C., & Zah, R. (2011). Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. *International Journal of Remote Sensing*, *32*(21), 6437–6456. https://doi.org/10.1080/01431161.2010.512928

Taubenböck, H., & Kraff, N. J. (2014). The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment*, *29*(1), 15–38. https://doi.org/10.1007/s10901-013-9333-x

Thomson, D. R., Linard, C., Vanhuysse, S., Steele, J. E., Shimoni, M., Siri, J., Caiaffa, W. T., Rosenberg, M., Wolff, E., Grippa, T., Georganos, S., & Elsey, H. (2019). Extending Data for Urban Health Decision-Making: a Menu of New and Potential Neighborhood-Level Health Determinants Datasets in LMICs. *Journal of Urban Health*, *96*(4), 514–536. https://doi.org/10.1007/s11524-019-00363-3

UN-Habitat. (2003). The Challenge of Slums. In *The Challenge of Slums*. https://doi.org/10.4324/9781849772907

UN-Habitat. (2015). *Tenure Security, Land and Property Rights in an Urban Context*.

UN-Habitat. (2016). Urbanization and Development: Emerging Futures. In *World Cities Report 2016*. http://wcr.unhabitat.org/main-report/

UN-Habitat. (2018). *SDG Indicator 1.4.2 Training Module: Secure Tenure Rights to Land*. *March*.

United Nations. (2015). Transforming Our World: the 2030 Agenda for Sustainable Development United Nations United Nations Transforming Our World: the 2030 Agenda for Sustainable Development. A/RES/70/1. In *United Nations* (Vol. 16301, Issue October).

https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf

United Nations. (2018). The World 's Cities in 2018. *The World's Cities in 2018 - Data Booklet (ST/ESA/ SER.A/417)*, 34. United Nations: Department of Economic and Social Affairs

Uwayezu, E., and de Vries, W.T (2020), Access to Affordable Houses for the Low-Income Urban Dwellers in Kigali: Analysis Based on Sale Prices. Land 2020, 9(3), 85; https://doi.org/10.3390/land9030085

Wang, J., Kuffer, M., Roy, D., & Pfeffer, K. (2019). Deprivation pockets through the lens of convolutional neural networks. *Remote Sensing of Environment*, *234*, 111448. https://doi.org/10.1016/j.rse.2019.111448

Wurm, M., & Taubenböck, H. (2018). Detecting social groups from space – assessment of remote sensing-based mapped morphological slums using income data. *Remote Sensing Letters*, *9*(1), 41–50. https://doi.org/10.1080/2150704X.2017.1384586

https://dx.doi.org/10.4314/rjeste.v5i1.2