Detecting Informal Settlements using VHR Satellite Imagery and Convolutional Neural Networks

Daniela de los Santos, Margherita Philipp

June 10, 2023

Contents

1 Abstract **1**

2 Introduction **1**

3 Background **2**

4 Data **3**

4.1 Data acquisition . . . 3

4.2 Preprocessing steps . . . 4

5 Model **5**

6 Results **6**

6.1 Within-slum predictions . . . 6

6.2 Across-slum predictions . . . 7

7 Discussion **8**

References **10**

1 Abstract

This document presents an exercise of mapping informal settlements using satellite imagery and deep learning models. Informal settlements, characterized by unplanned and often illegal housing, are a significant challenge in many regions of the world. Traditional methods of data collection and mapping in these areas are limited, time-consuming, and costly. Satellite imagery provides a unique opportunity to efficiently map and understand these settlements at a large scale. The study aims to answer two key questions: how helpful are very high-resolution (VHR) satellite images in predicting the location of informal settlements, and can a model trained in one location accurately map informal settlements in a different location? With a few promising results, we highlight and discuss the challenges associated with mapping informal settlements, including the limited availability of reliable ground truth data.

2 Introduction

Informal settlements, characterized by unplanned and often illegal housing, are prevalent in many regions of the world, posing significant social, economic, and environmental challenges.

Traditional methods of data collection and mapping in these areas are often limited, time-consuming, and costly. In developing countries, inadequate information on the identification of existing and potential informal settlements within cities remains a core concern to policymakers and urban planners (IBRAHIM et al., 2019).

In that framework, satellite imagery provides a unique vantage point, enabling comprehensive and efficient mapping of informal settlements at a large scale. By utilizing satellite imagery, we gain insights into the spatial extent, growth patterns, infrastructure deficiencies, and population dynamics of these settlements. Moreover, satellite-based mapping facilitates monitoring changes over time without depending on large data collection endeavors such as census, allowing for real-time evaluation of interventions and the development of sustainable strategies for inclusive urban development. Thus, harnessing the power of satellite imagery for mapping informal settlements is promising for advancing our understanding and fostering positive change in urban contexts worldwide.

Several researchers have applied various deep learning models to address the task of informal settlement mapping. Convolutional Neural Networks (CNNs) have been identified as the most effective approach for this purpose. However, there are still numerous challenges associated with this type of exercise. As discussed in the next section, many of these challenges stem from the limited availability of reliable and comparable labeled data.

In that sense, our approach aims to answer the following questions: how helpful are Very High-Resolution (VHR) satellite images in predicting the location of informal settlements in a city? Secondly, can a model trained in a given location help map informal settlements in a different location altogether?

In practical terms, this translates into a twofold exercise: 1) predicting the location of slums[1] using a model trained in the same city or location (what we will refer to as *within-slum prediction*), and 2) predicting the location of slums with a model that has been trained in a different location (what we refer to as *across-slum prediction*). Even though we expect to have much better predictive power in the within-slum exercise, we also expect across-slum predictions to have some predictive power.

The document is structured as follows. We first provide a brief literature review of the state of the art about informal settlements prediction using satellite imagery. We then describe the data sources used in the exercise, the preprocessing steps we take, as well as define the model specifications. Lastly, we present results for within-slum and across-slum predictions and finalize with a discussion section.

3 Background

There are several background papers that explore machine and deep learning applications to detect or map informal settlements, particularly using remote sensing and satellite imagery as a main sources of data.

Some studies have used CNNs, machine learning models (such as Random Forest), deep convolutional neural networks, and artificial neural networks for this task (BHANGALE et al., 2016; GADIRAJU et al., 2018; MBOGA et al., 2017; PERSELLO; STEIN, 2017; PRABHU; PARVATHAVARTHINI; ALAGURAJA, 2021). Comparative studies point to the fact that CNNs perform better in informal settlements classification problems since they learn deep

[1] Conceptually, a slum is a specific type of informal settlement. Our data is optimized for slum prediction instead of a more general informal settlement prediction. However, we use these two terms interchangeably in this paper.

features that are key for improving accuracy (XU, 2021). Because CNNs work best with VHR images, some researchers have focused on developing models that can work with medium and low-resolution images, to avoid the prohibitive costs of both obtaining VHR images and training models with them (GRAM-HANSEN et al., 2019).

In general, the absence of a coherent definition of informal settlements, a high degree of heterogeneity in appearance, and the variation of spectral and spatial image resolution have been relevant challenges to identifying and predicting the location of informal settlements from satellite images (HOFMANN; BEKKARNAYEVA, 2017; AMBUGADU; HOSNI, 2022). It has been observed that false positives are an important problem, especially in high-density areas (XU, 2021). The task becomes even more challenging when we intend to make predictions across different locations, i.e. predicting a slum location in a given city with a model that was trained elsewhere.

Furthermore, the limited availability of ground-truth data poses a significant challenge for this task. Official census data is typically provided at aggregated national levels, such as census tracts, which are not suitable for training these types of models. Researchers often resort to generating their own ground truth labels through field surveys or collaborations with institutions. However, only a few researchers make their ground truth data openly available, which slows down research progress in this field (MATTOS; MCARDLE; BERTOLOTTO, 2021).

In addition, even when ground-truth data is available, it can present its own set of challenges. Firstly, there is noticeable heterogeneity in the structure of each ground label footprint, which is highly dependent on factors such as annotators' methodology, available resources, definitions, and goals. We delve further into this topic and provide examples in the Data section. Secondly, there is ambiguity surrounding the content of background pixels (XU, 2021). While annotators, such as researchers, governments, or NGOs, may have a clear understanding of the boundaries of a known slum, other areas labeled as non-slum may actually consist of a combination of genuine background regions and regions that may contain various types of informal settlements or potentially newer, less-known, rapidly expanding slums.

Given these challenges, our paper aims to offer a systematic comparative review of within-slum and across-slum predictions. To achieve this, we utilize the datasets provided by (GRAM-HANSEN et al., 2019), which have been used in various studies, e.g., (MATTOS; MCARDLE; BERTOLOTTO, 2021). The primary objective is to provide insights about the reliability of employing models trained in different areas for predicting and mapping slums in unseen regions.

4 Data

4.1 Data acquisition

As mentioned, the datasets used in this project, consisting of both VHR satellite imagery and ground truth masks, were made available by (GRAM-HANSEN et al., 2019). In our application, we select four territories: Medellín, Colombia; Makoko, Nigeria, Kibera-Nairobi and Lower Nairobi, Kenya. Because the datasets are specifically focused on slum areas, we do not face a severe imbalance problem while identifying the target class (see Table 1). An example of how the data looks like can be observed in Figure 1.

In terms of how this data was generated, the authors employed a fusion of satellite imagery, on-the-ground measurements and partnered with diverse organizations to create ground truth masks. It should be acknowledged that because the ground truth data was gathered from various sources, they have different characteristics across locations. In some locations, ground truth

	Kibera-Nairobi, Kenya	Lower Nairobi, Kenya	Makoko, Nigeria	Medellín, Colombia
Train set	36,2	14,0	31,0	41,4
Validation set	36,1	14,2	35,4	39,3
Test set	34,5	12,5	38,0	46,4

Table 1: Share of slum pixels in each split

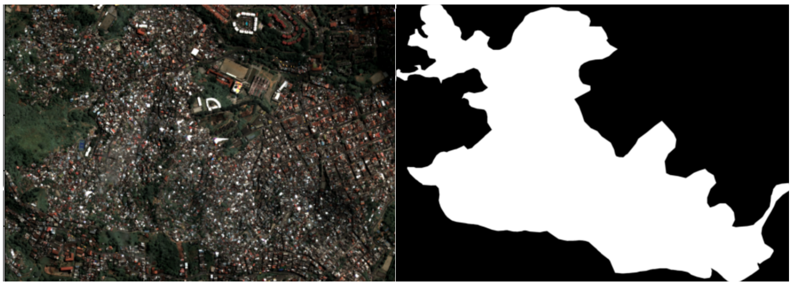

Figure 1: VHR satellite imagery and ground truth masks identifying slums in Medellín, Colombia.

labels provide very detailed information and prioritize the classification of specific features such as certain buildings or types of roofs. In contrast, others encompass larger areas and appear as more generalized representations. An example of this heterogeneity can be observed in Figure 2.

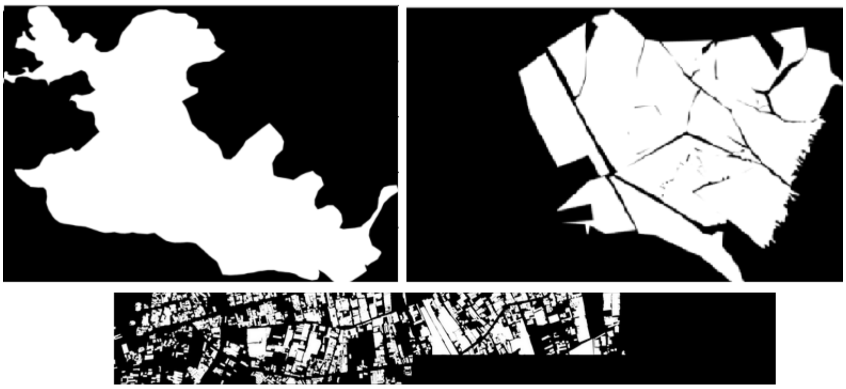

Figure 2: Ground truth masks for selected territories. Left: Medellín. Right: Makoko. Down: Lower Nairobi.

Datasets are available for download here: https:// frontierdevelopmentlab.github.io/informal-settlements/. It is worth noting that we are using VHR imagery in our exercise, whereas the repository also contains low-resolution images that we do not use. The VHR images come from DigitalGlobe [2], and were provided to the authors through Satellite Applications Catapult.

4.2 Preprocessing steps

As a first step, we partition both the satellite images and ground truth masks into tiles. This is a necessary step for several reasons. Firstly, working with a large image can be compu-

[2]See: https://www.digitalglobe.com

tationally intensive and may pose challenges in terms of memory requirements and processing time. By dividing the image into smaller tiles, the data becomes more manageable and can be processed efficiently. Secondly, tiling helps in ensuring that the model trained on the data does not solely rely on learning patterns specific to the entire image. By including diverse tiles from different parts of the image in the training, validation, and testing sets, the model can learn a more generalized representation of the features and patterns present in the data. Furthermore, tiling allows for better control over the distribution of the data across the different sets, ensuring that there is no bias introduced due to the spatial arrangement of the tiles. Figure 3 provides examples of several pairs of images and mask tiles.

As a second preprocessing step, we perform image augmentation. By applying a variety of transformations and modifications to the original images, such as rotations, translations, flips, zooms, and color manipulations, image augmentation introduces variations that can improve the performance and generalization ability of our model.

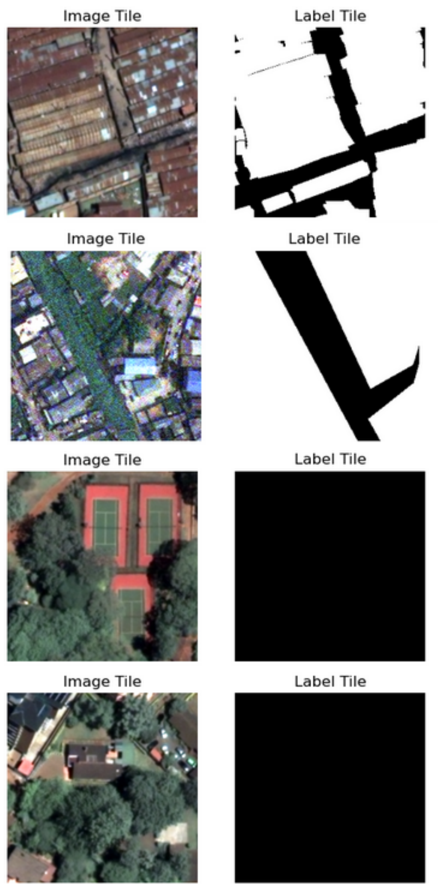

Figure 3: Examples of randomly chosen tiles for the case of Lower Nairobi.

5 Model

The model implemented in this study is based on the U-Net architecture, a popular and effective model for semantic segmentation tasks. The U-Net architecture is a CNN that consists of an encoder and a decoder. The encoder part, in this case, utilizes the ResNet-34 backbone, which has shown excellent performance in various computer vision tasks. The encoder weights are not initialized with pre-trained weights from ImageNet, allowing the model to learn from scratch based on the specific slum mapping task at hand. The model takes RGB images as input and predicts the class labels for each pixel, with a total of two classes, representing slum and non-slum regions. The model is trained using the Adam optimizer with a learning rate of 1e-4 and a batch size of 32.

In practical terms, we train a total of four models, one per study case. We then evaluate each model against a test set from the same territory (within-slum prediction), and also against the test sets of the other three territories under study (across-slum prediction). In the following section, we discuss the results and performance metrics for both exercises.

6 Results

6.1 Within-slum predictions

Table 2 presents the performance metrics for out-of-sample within-slum predictions in our four selected locations. The metrics evaluated include mIoU (mean Intersection over Union), precision, recall, and F1 score.

The mIoU measures the average overlap between the predicted and ground truth slum areas - higher mIoU values indicate better overall performance. Among the four locations, Medellín achieved the highest mIoU score of 0.63, indicating a relatively better alignment between the predicted slum areas and the ground truth. However, both Mokoko and Lower Nairobi present mIoU values around 0.5, showing a relatively good alignment between predictions and ground truth as well.

In terms of precision, which quantifies the accuracy of positive predictions (or, in other words, the ability of the model to avoid false positive predictions), Lower Nairobi achieved the highest precision score of 0.85, suggesting a relatively low rate of incorrectly identified slum areas. With the exception of Makoko, which scores a precision of 0.56, the results for the other areas are moderately good in this regard.

Recall measures the ability of the model to capture all actual positive instances, representing the proportion of correctly predicted slum areas out of all actual slum areas. Medellín attained the highest recall score of 0.91, indicating a better ability to detect slum areas accurately, followed by Makoko (0.85). Meanwhile, Lower Nairobi and Kibera show lower recall values.

Finally, we present the F1 score, which is the harmonic mean of precision and recall, providing a balanced measure of model performance. It considers both false positives and false negatives. We see that Medellín has the highest F1 scored, followed by Makoko and Lower Nairobi.

It is interesting to observe the heterogeneity in the results of applying the same CNN model to different VHR satellite imagery. The range of metrics obtained for the four territories shows that this is not an easy problem to solve and that more research is needed to optimize predictions. Furthermore, we go back to the problem of the difficulty to obtain accurate and comparable ground truth data, which clearly relates to the range of metrics obtained for the within-slum prediction exercise. For instance, if we inspect Figures 4 and 5, we could think that the CNN has the capacity to closely learn the deep features that characterize a slum. However, areas that look very similar to a slum from satellite images might either not be actual slums, or, alternatively, we might be facing the case of mislabeled ground truth data. We return to this in the Discussion section.

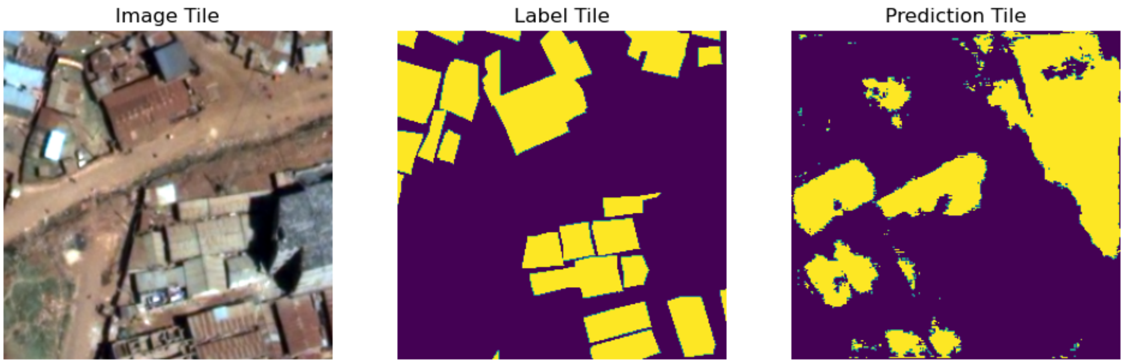

Figure 4: Examples of predictions for Lower Nairobi.

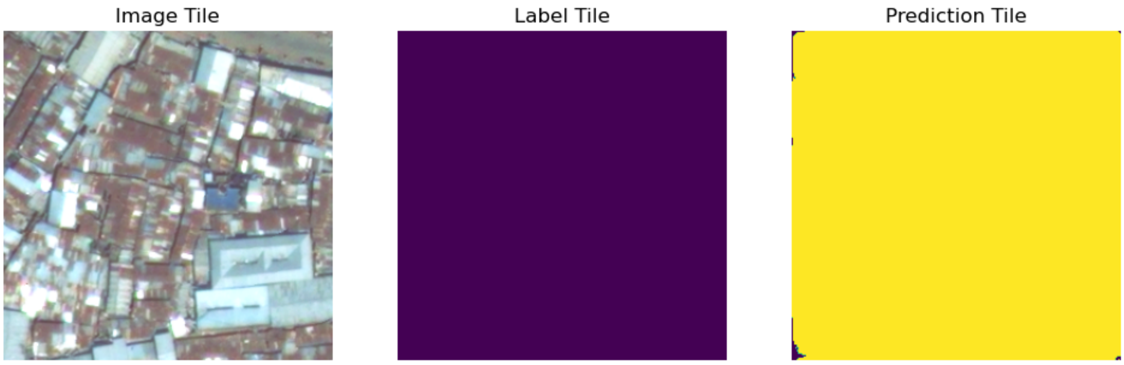

Figure 5: Examples of predictions for Kibera.

6.2 Across-slum predictions

Table 3 presents performance metrics for both across-slum and within-slum predictions. It is easy to see that within-slum predictions present (in general) much more accurate results, which is to be expected. This can be attributed to the familiarity of the models with the specific characteristics of the training area. However, the across-slum predictions also present a high degree of heterogeneity and some curious cases worth flagging.

For example, although Kibera and Lower Nairobi are both situated in the same city, trying to map the Lower Nairobi slum with a model trained in Kibera yields very inaccurate results. In fact, we might get better results when training in Makoko or even a non-African city like Medellín. Conversely, when trying to map the Kibera slum using the model trained in Lower Nairobi, precision values are very high, which seems promising, even though other values such as recall or mIoU are poor. In this case, the Medellín model seems to yield the better

	Lower-Nairobi, Kenya	Kibera, Nairobi, Kenya	Makoko, Nigeria	Medellín, Colombia
mIoU	0.4773	0.4110	0.5119	0.6252
Precision	0.8524	0.7256	0.5615	0.6639
Recall	0.5203	0.4867	0.8529	0.9146
F1 score	0.6462	0.5826	0.6772	0.7694

Table 2: Performance metrics for out-of-sample within-slum predictions.

Slum to predict	Precision Metric	Model trained in:			
		Lower Nairobi	*Kibera*	*Makoko*	*Medellín*
Lower Nairobi	mIoU	**0.4773**	0.0307	0.1459	0.1560
	Precision	**0.8524**	0.1165	0.2603	0.1638
	Recall	**0.5203**	0.0400	0.2494	0.7669
	F1-Score	**0.6462**	0.0595	0.2547	0.2699
Kibera	mIoU	0.1023	**0.4110**	0.0915	0.3680
	Precision	0.8258	**0.7256**	0.4594	0.4372
	Recall	0.1045	**0.4867**	0.1025	0.6993
	F1-Score	0.1855	**0.5826**	0.1676	0.5381
Makoko	mIoU	0.0003	0.0000	**0.5119**	0.2852
	Precision	0.0724	0.0000	**0.5615**	0.2859
	Recall	0.0003	0.0000	**0.8529**	0.9915
	F1-Score	0.0006	0.0000	**0.6772**	0.4438
Medellín	mIoU	0.0000	0.0139	0.2686	**0.6252**
	Precision	0.4105	0.1485	0.5612	**0.6639**
	Recall	0.0000	0.0152	0.3400	**0.9146**
	F1-Score	0.0001	0.0275	0.4234	**0.7694**

Table 3: Performance metrics for across-slum and within-slum predictions. **In bold:** within-slum predictions.

across-slum predictions.

For the case of Makoko, it is perhaps most surprising that neither of the Nairobi models work, with metrics approximating 0 -that is, almost all predictions, positive and negative, are wrong. Nevertheless, the Medellín model has an impresive recall value (there are virtually no false negatives amongst the predictions), while precision and mIoU metrics are not as low as other across-slum exercises. Finally, models trained in Makoko and Lower Nairobi show moderately good precision when mapping the slum in Medellín.

As a general takeaway, even though most across-slum predictions yield inaccurate results, the model trained in Medellín seems to be useful from predicting outside its original location. In many cases, it returns impressively high recall values. In the Discussion section, we analyze why this fact is very relevant to our problem.

7 Discussion

This study aimed to investigate the effectiveness of VHR satellite imagery in predicting the location of informal settlements within cities. Specifically, we focused on two prediction scenarios: within-slum prediction, where the model is trained and tested on the same location, and across-slum prediction, where the model is trained in one location and tested in a different location. In this section, we will discuss our findings, address the challenges encountered, and provide some insights for future research.

Firstly, our results showed that within-slum predictions yielded relatively better performance compared to across-slum predictions, as was expected. This suggests that the model trained on a specific location was able to capture the unique features and patterns of slums in that area, resulting in more accurate predictions. On the other hand, across-slum predictions exhibited lower performance metrics, indicating that the model trained in one location did not generalize well to a different location. This finding highlights the importance of local context and spatial variations in informal settlements, which pose challenges for models trained on one dataset to be directly applicable to other locations.

The variations in performance across different locations further emphasize the heterogeneity and complexity of informal settlements. Different regions exhibit diverse characteristics, such as architectural styles, materials used, and population density, which can significantly impact the appearance of slum areas in satellite imagery. This variation makes it challenging to develop a one-size-fits-all model that can accurately predict slum locations in different contexts. Future research should explore strategies to adapt models to local conditions and incorporate contextual information to improve prediction accuracy.

Nevertheless, it is worth noting that some performance metrics were sometimes surprisingly high in across-slum exercises, which shows promise. When analyzing these metrics, it is crucial to consider the inherent tradeoff between precision and recall and evaluate it from a substantive perspective. In our specific case, our primary concern lies with minimizing false negatives rather than false positives, making recall a priority. While false positives can be investigated and researched further on the ground, false negatives deprive us of valuable information about newly emerging or expanding slums. This absence of information can have significant implications, as it may result in people residing in these areas being deprived of essential services and resources. Therefore, prioritizing recall allows us to capture a higher proportion of positive instances, reducing the chances of missing critical developments in slum areas. For all four studied territories, predictions based on the Medellín model yielded very high recall values. In that sense, even though more research is needed, using this model for across-slum predictions even in distant regions might be of use to urban planners and policymakers.

One of the main challenges we encountered in this study was the limited availability of reliable and comparable ground truth data. Ground truth labels for informal settlements are often generated through field surveys or collaborations with organizations, resulting in heterogeneous and potentially inconsistent annotations. This heterogeneity can introduce uncertainties and affect the performance of the models. If the ground truth masks for each location have varying levels of detail, complexity, or subjective interpretation, it can impact the model's ability to learn and generalize effectively. In the case of Medellín, which seems to be the model with more 'generalization' power, the ground truth mask appears as a big blob, which suggests that the slum areas are relatively large and contiguous. This type of ground truth mask may make it easier for the model to identify and predict slum areas accurately, leading to higher performance metrics. On the other hand, cases like Lower Nairobi that have a more detailed ground truth mask present a greater challenge for the model, as it requires it to discern subtle differences between slum and non-slum areas, which can be more prone to errors and inconsistencies. Consequently, this could lead to lower accuracy metrics compared to locations with less detailed ground truth masks. These differences highlight the importance of having consistent and reliable ground truth data for training and evaluating models, as it directly influences their performance in different locations.

In addition to this, during our analysis (which included random visual checks), we observed discrepancies between the appearance of areas labeled as slums in the ground truth masks and their corresponding predictions. Some areas that visually resembled slums were not classified as such, suggesting potential mislabeled or ambiguous ground truth data. Addressing this challenge requires efforts to establish standardized annotation protocols, improve data-sharing practices, and incorporate expert knowledge to ensure the accuracy and consistency of ground truth labels.

Another important aspect to consider is the availability and quality of satellite imagery.

While VHR imagery provides valuable insights due to its high level of detail, obtaining and processing such imagery can be costly and resource-intensive. In our study, we utilized VHR images made available by other researchers, but we acknowledge the limitations of relying solely on this data source. Exploring alternative data acquisition methods, such as crowd-sourcing or drone imagery, could offer more accessible and up-to-date information for mapping informal settlements. Additionally, research should focus on developing models that can effectively utilize lower-resolution imagery, as it is more widely available and can provide valuable insights in resource-constrained settings.

Our findings indicate the inherent challenges associated with making accurate predictions for slum areas that differ significantly from the training data. The variations in the metrics underscore the need for further research and model optimization to improve the performance of across-slum predictions. Strategies such as incorporating more diverse training data or adapting models to specific characteristics of different slum areas may help address these challenges and enhance the accuracy of predictions in unfamiliar territories.

References

AMBUGADU, Amadu Monday; HOSNI, Nafisa. Identifying informal settlements in satellite images for sustainable urban planning: A systematic review of methods available. **Journal of Information System and Technology Management**, v. 7, n. 25, p. 102–119, 7 Mar. 2022. ISSN 0128-1666. DOI: 10.35631/JISTM.725008. Available from: <http://www.jistm.com/PDF/JISTM-2022-25-03-08.pdf>. Visited on: 9 June 2023.

BHANGALE, Ujwala M. et al. Identification of informal settlement using Remote Sensing Images. **Proceedings of the International Conference on Advances in Information Communication Technology & Computing**, 2016.

GADIRAJU, K. K. et al. Machine Learning Approaches for Slum Detection Using Very High Resolution Satellite Images. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW). [S.l.: s.n.], Nov. 2018. P. 1397–1404. DOI: 10.1109/ICDMW.2018.00198.

GRAM-HANSEN, Bradley et al. Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data. In: PROCEEDINGS of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. [S.l.: s.n.], 27 Jan. 2019. P. 361–368. DOI: 10.1145/3306618.3314253. arXiv: 1901.00861[cs,stat]. Available from: <http://arxiv.org/abs/1901.00861>. Visited on: 9 June 2023.

HOFMANN, Peter; BEKKARNAYEVA, Gulnaz. Object-based change detection of informal settlements. **2017 Joint Urban Remote Sensing Event (JURSE)**, p. 1–4, 2017.

IBRAHIM, Mohamed R. et al. 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**, v. 76, p. 31–56, 2019. ISSN 0198-9715. DOI: https://doi.org/10.1016/j.compenvurbsys.2019.03.005. Available from: <https://www.sciencedirect.com/science/article/pii/S0198971518303545>.

MATTOS, Agatha C. H. de; MCARDLE, Gavin; BERTOLOTTO, Michela. **Mapping Slums with Medium Resolution Satellite Imagery: a Comparative Analysis of Multi-Spectral Data and Grey-level Co-occurrence Matrix Techniques**. [S.l.]: arXiv, 21 June 2021. DOI: 10.48550/arXiv.2106.11395. arXiv: 2106.11395[cs]. Available from: <http://arxiv.org/abs/2106.11395>. Visited on: 9 June 2023.

MBOGA, Nicholus et al. Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. **Remote Sensing**, v. 9, n. 11, 2017. ISSN 2072-4292. DOI: 10.3390/rs9111106. Available from: <https://www.mdpi.com/2072-4292/9/11/1106>.

PERSELLO, Claudio; STEIN, Alfred. Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. **IEEE Geoscience and Remote Sensing Letters**, v. 14, p. 2325–2329, 2017.

PRABHU, Ravi; PARVATHAVARTHINI, Balasubramanian; ALAGURAJA, Alagumalai R. Integration of deep convolutional neural networks and mathematical morphology-based postclassification framework for urban slum mapping. **Journal of Applied Remote Sensing**, v. 15, 014515, p. 014515, Jan. 2021. DOI: 10.1117/1.JRS.15.014515.

XU, Yajin. Informal settlement classification using remote sensing: combining multi-level image features. **Technische Universität München - Master's Thesis**, 2021.