

Article

Machine Learning-Based Slum Mapping in Support of Slum Upgrading Programs: The Case of Bandung City, Indonesia

Gina Leonita ^{1,2}, Monika Kuffer ^{1,*} , Richard Sliuzas ¹  and Claudio Persello ¹ 

¹ Faculty of Geo-Information Science & Earth Observation (ITC), University of Twente, 7514 AE Enschede, The Netherlands; g.leonita2901@gmail.com (G.L.); r.sliuzas@utwente.nl (R.S.); c.persello@utwente.nl (C.P.)

² Ministry of Public Works and Housing of Indonesia, Jalan Pattimura No.20, Jakarta Selatan 12110, DKI Jakarta, Indonesia

* Correspondence: m.kuffer@utwente.nl; Tel.: +31-534-874-301

Received: 31 July 2018; Accepted: 19 September 2018; Published: 22 September 2018



Abstract: The survey-based slum mapping (SBSM) program conducted by the Indonesian government to reach the national target of “cities without slums” by 2019 shows mapping inconsistencies due to several reasons, e.g., the dependency on the surveyor’s experiences and the complexity of the slum indicators set. By relying on such inconsistent maps, it will be difficult to monitor the national slum upgrading program’s progress. Remote sensing imagery combined with machine learning algorithms could support the reduction of these inconsistencies. This study evaluates the performance of two machine learning algorithms, i.e., support vector machine (SVM) and random forest (RF), for slum mapping in support of the slum mapping campaign in Bandung, Indonesia. Recognizing the complexity in differentiating slum and formal areas in Indonesia, the study used a combination of spectral, contextual, and morphological features. In addition, sequential feature selection (SFS) combined with the Hilbert–Schmidt independence criterion (HSIC) was used to select significant features for classifying slums. Overall, the highest accuracy (88.5%) was achieved by the SVM with SFS using contextual, morphological, and spectral features, which is higher than the estimated accuracy of the SBSM. To evaluate the potential of machine learning-based slum mapping (MLBSM) in support of slum upgrading programs, interviews were conducted with several local and national stakeholders. Results show that local acceptance for a remote sensing-based slum mapping approach varies among stakeholder groups. Therefore, a locally adapted framework is required to combine ground surveys with robust and consistent machine learning methods, for being able to deal with big data, and to allow the rapid extraction of consistent information on the dynamics of slums at a large scale.

Keywords: machine learning; slums; slum upgrading programs; Bandung; Indonesia

1. Introduction

1.1. Background

Slum upgrading has become an international concern and agenda promoted by the Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs). The Government of Indonesia has committed to reducing slums and released a new national policy, called the Sustainable Housing Programs 100-0-100, aiming at achieving cities without slums by 2019 [1]. The lack of accurate baseline data of slum areas is one of the challenges in achieving this target. Such data are required to support the government in the selection of priority areas, monitoring the implementation,

and calculating areas before and after upgrading programs. In 2015, a total of 38,431 ha of slum areas were reported in 390 cities and districts of Indonesia using survey-based slum mapping (SBSM) [2]. Slum mapping is based on physical and social criteria [3]. However, SBSM is labor-intensive and time- and cost-consuming, particularly when frequent updating is required. A major shortcoming of SBSM is inconsistencies in the results due to different interpretations of slum indicators by surveyors in the field and differences in their experiences. Figure 1 depicts such inconsistencies from the report on “Strategy for achieving the target of the Medium-Term Development Plan in 2015–2019” [2] for the cities of Sorong and Samarinda, where a river, pond, and green areas are delineated as slums.

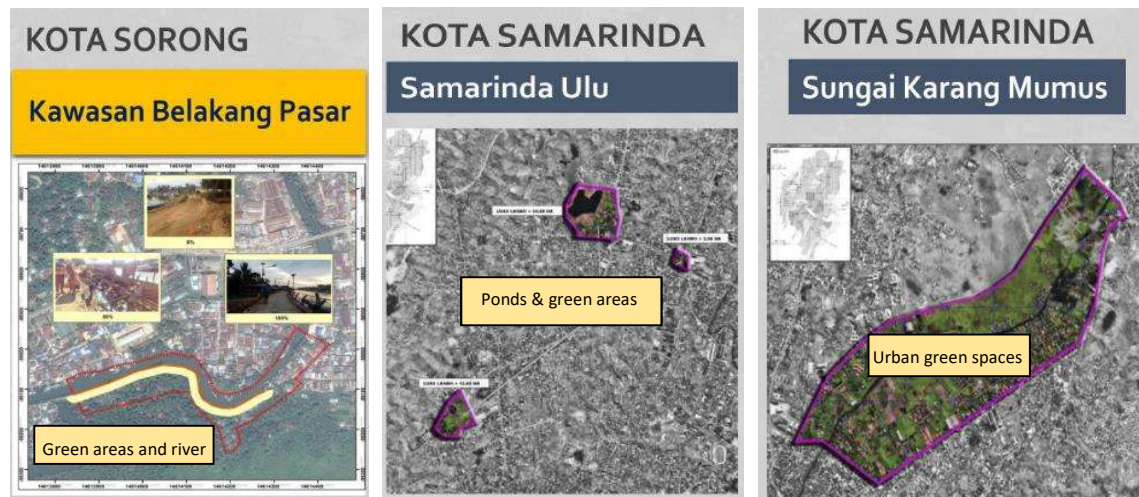


Figure 1. Example of inconsistencies in survey-based mapping, adapted from [2].

To tackle these issues, remote sensing-based slum identification is proposed. Several slum mapping studies have used VHR images (e.g., [4,5]), showing the scope of remote sensing, but also the inherent uncertainties [6]. Recently, several studies stressed the capacity of machine learning (ML) for slum identification, including, beyond spectral, also features of texture, geometry, and structure [7]. However, those studies did not analyze how the derived information from ML could be used to support slum upgrading programs; most studies do not consider this aspect and the political context of their mapping results.

In general, there are two essential elements that influence a successful slum mapping method: first, the conceptualization of real-world slum characteristics, which allows local slum characteristics to be translated into image features; second, classifiers must be fed with predefined contextual features of slum characteristics of the specific region. Thus, to perform slum identification by ML, slum characteristics need to be well understood. For this purpose, a generic ontological framework for slums has been developed by Kohli et al. [8], as slums vary across cities. Kohli et al. [8] stressed that a local adaptation of the generic slum ontology (GSO) is required, incorporating local expert knowledge, referred to as the local slum ontology (LSO).

Using VHR images, the LSO can guide the feature selection for slum detection with ML. It has the capability of operating with large sets of features with efficient computation [4]. A recent study [7] examining several ML approaches for slum classification using spectral, textural, and structural features within VHR imagery showed that the support vector machine (SVM) outperformed other ML methods for mapping slums at the city scale.

The aim of this study is to explore the potential of ML algorithms for slum mapping in support of the Indonesian national target of “cities without slums”. The performance of two popular ML algorithms [4,9], i.e., RF and SVM, is assessed for slum mapping, using the example of Bandung City. We analyze whether a ML-based slum mapping approach could be an alternative for the presently conducted survey-based approach. Thus, we want to understand the views of local stakeholders.

Therefore, we first mapped slums to discuss them with local stakeholders. For the methods, we select standard methods in machine learning that would allow the mapping of slums at the city scale. However, we want to go one step further. The qualitative analysis from stakeholder interviews is very useful to understand what is still missing for supporting local planning and decision-making. Thus, we can better understand which future developments are necessary.

SVM and RF are selected, from among other recent developments in the field of ML (e.g., artificial neural networks or deep learning), as they are available in standard, relatively user-friendly, open-access software to support easy access also in resource-constrained environments. Thus, we assess whether ML allows capturing of the unique and complex slum characteristics in an Indonesian city. Mapping slums in Indonesia is rather complex, as slum and nonslum Kampung (informally developed areas) commonly share similar morphological characteristics (many nonslum Kampung are, in fact, mid-income housing areas).

For SVM, the radial basis function (RBF) kernel is used. There are several SVM kernels, such as linear, polynomial, and sigmoid. In general, a linear kernel can also have a good performance for a binary problem and has advantages in terms of computational costs [10,11]. However, based on recent publications (e.g., [12,13]), the popular RBF kernel is selected as it generally produces state-of-the-art results in a variety of applications. Furthermore, RF and SVM RBFs show good performance in terms of computational time and classification accuracy [14], which is very relevant to upscale methods for city or national slum mapping. In general, RF is efficient in parameter selection and is computationally fast, while SVM commonly performs better with multidimensional features [15,16]. Many other prominent ML algorithms are found these days, such as convolutional neural networks (CNNs) [17]. However, those algorithms typically need large training datasets and are computationally more costly.

1.2. Conceptual Framework

To upgrade slum areas, the Indonesian government requires a consistent, detailed, correct, and timely method that meets the requirements specified in planning documents. Inconsistencies and temporal delays are shortcomings of the SBSM undertaken by the Indonesian government. Therefore, this study evaluates the utility of ML-based slum mapping to support stakeholders with consistent baseline data for planning processes and slum upgrading programs. Consistent data in this study refers to data generated using the same principles and which are replicable.

As mentioned in Section 1.1, local slum characteristics (LSO) are the basis for slum classifications using satellite imagery. The LSO is a local adaption of the GSO framework that covers the environs, settlements, and object dimensions of slums. Based on expert interviews and visual image inspection, our LSO only includes settlements and object-level image features. The environs level (the location or neighborhood) could be included by GIS layers (e.g., land use and hazard maps); however, to avoid introducing uncertainties (local maps can be dated and of varying scales), we omitted this level. The settlement level can be depicted by morphological, textural, and spectral features. The shape of slum settlements (such as irregular) can be determined by morphological features, while built-up densities, being usually high in slums, can be captured by contextual features and spectral features, such as low normalized difference vegetation index (NDVI) values, which indicate the absence of vegetation due to high built-up densities. The object level, referring to building and road characteristics, is specified by contextual, spectral, and morphological features. The roof material and unpaved streets in slums can be explained by spectral features; object (roof) shapes can be described by morphological features, while irregular-access networks can be described by contextual features. The relationship between image features and LSO is not simple: It can be one to many; one image feature can describe several LSO. The relationship can also be many to one, where many image features describe one LSO component, or many to many, where many image features describe many components (Figure 2).

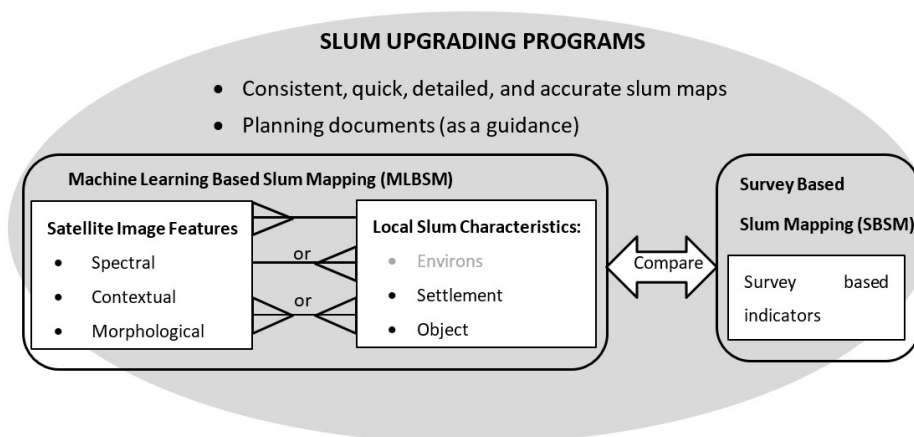


Figure 2. Conceptual framework.

1.3. Study Area

This study was conducted in Bandung, the capital city of West Java Province in Indonesia. The city is attracting many immigrants because of employment and educational opportunities. Its population is 2,481,500 persons, with a density of 14,831 people per km² in 2016 [18]. The city is subdivided into 30 kecamatan (districts) with 151 kelurahan (urban villages) [19]. The backlog of housing provision [20] and the immigration flow are the main reasons for the slum existence in Bandung [21]. According to SBSM, there are 454 slum neighborhoods within the city, with a total area of 1457.45 ha [20].

2. Methodology

The methodology is split into four main steps (Figure 3), i.e., preprocessing, main process, comparing with SBSM result, and the evaluation in the context of the national target of “cities without slums”. In the first step, radiometric correction was conducted. Next, we selected several kelurahan (urban villages) from the city planning documents, based on slum location characteristics. By combining the LSO and government criteria for slum mapping, we analyzed the potential of image-based features to differentiate slum and nonslum areas. The second step included feature extraction, feature selection and classification. The extraction of contextual, spectral, and morphological features was followed by sequence forward selection (SFS) combined with the Hilbert–Schmidt independence criterion (HSIC). This produced an informative feature subset to be used as input for the classification, and then the classification was performed, next the accuracy was assessed using ground truth data (collected by the first author, guided by the local surveyor team). In the third step, the classification results were compared with the SBSM result. This allowed us to compare strengths and weaknesses of both approaches. Within the fourth step, we assessed the application potential of ML-based slum mapping in support of the national slum mapping campaign in Indonesia, focusing on the city of Bandung.

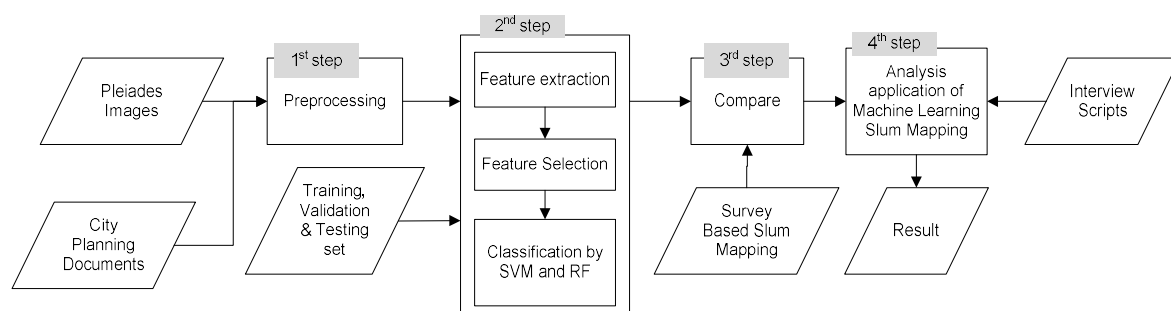


Figure 3. Research methodology.

2.1. Material

This study used primary and secondary data (Table 1), including pansharpened Pleiades imagery from 2016. To anticipate changes and to check the quality of slum boundaries from 2015, we used historical Google Earth images and ground truth data. For the ground truth data collection, one hundred random points were selected, and in addition, areas with doubtful cases during image interpretation (whether those areas were slums or not) were included. The primary data collection included also expert interviews and a local meeting with the surveyor team, in order to understand the SBSM and to evaluate the possibility of implementing a ML-based slum mapping approach. The respondents for the expert interviews included an urban planner from the Ministry of Public Works and Housings and another from the municipality who was organizing the slum upgrading program and the slum delineation process, a surveyor team experienced in survey-based mapping, and a professor at a local university with expertise in slum mapping.

Table 1. Primary and secondary data.

Data	Year	Data Sources	Category
Pleiades (pansharpened) images. Res: 0.5 m	2016 (July and August)	European Space Agency (ESA)	Primary
Slum boundaries	2015	Ministry of Public Works and Housing	Secondary
Administrative boundary of Bandung city	2015	Municipality of Bandung	Secondary
Historical Google Earth images	2013–2016	Online data	Secondary
Validated slum boundaries	October 2017	Ground truth checking	Primary
Expert interview scripts	October 2017	Interview	Primary

2.2. Bandung Slum Characteristics and Image Features

Based on the field observations, Table 2 presents the slum characteristics in Bandung city and relates them with contextual, spectral, and morphological image features, thus representing the local slum ontology.

Table 2. Slum characteristics: the local slum ontology.

GSO Dimension	Indicator	Local Indicator	Image Feature
Environs	Location	Hazardous areas, in between small alleys	No image feature was used explain the environs level
	Neighborhood Characteristics	Proximity to industrial, commercial, formal residential, bus stations, and smelly and dirty areas	
Settlement	Shape	Irregular pattern, elongated formation following the river or railway	Contextual (PanTex, LBP, GLCM) and morphological features (APPR)
	Density	High density (more than 250 unit/ha), high roof coverage, less vegetation	Contextual (PanTex, LBP, GLCM) and spectral features (NDVI)
Object	Access Network	Unpaved or poorly constructed streets, width ≤ 2.5 m, covered conduits or without conduits	Contextual features (PanTex, LBP, GLCM)
	Building Characteristics	Permanent and nonpermanent structures, with the roofs made from corrugated iron, asbestos, plastic, fiber, and clay tiles; building size from 10–60 m ² ; poor sanitation, using well water or bought water	Spectral (original band) and morphological features

In slum neighborhoods, not all slum dwellers are poor. We found several houses with solid structures, clean walls, and strong gates. The average density of slums in Bandung city is 260–285 units/ha. Several houses were occupied by many people (overcrowding); e.g., a house located in Babakan neighborhood having only 60 m² was populated by 24 people. The dwellers made two impermanent floors to make more space. Moreover, they arranged to take turns in sleeping. In some cases, slum dwellers made a bridge at the second floor to connect the house to another house across the alley to expand their house, still allowing passage along the path below. In addition, small open spaces in slum areas were found, such as cramped football/basketball fields, cemeteries, or waste dumps. Vegetation is rarely found in slums. A lot of houses did not have sanitary waste management, using (covered) conduits to control the flow of grey and black water. When flooding occurs, all the

waste comes to the surface. Sanitation is a critical issue in such neighborhoods; e.g., the children usually get sick after the flooding. In the context of Indonesia, Pratomo et al. [6] found, in general, high uncertainties on slum locations and boundaries (existential and extensional uncertainties), and often the higher accuracy, the lower the certainty of the mapping result. Thus, the existence of kampongs contributes to these uncertainties. To describe the complex morphology, a large feature set was employed, which included original bands, NDVI (normalized difference vegetation index), built-up presence index (PanTex), grey-level co-occurrence matrix (GLCM), local binary pattern (LBP), and morphological features. The NDVI was used for analyzing vegetation presence and its conditions, since Bandung slums are very dense (with absence of vegetation), make it a good indicator to distinguish slum and nonslum neighborhoods [22]. PanTex is a built-up presence index [23], providing the degree of confidence of the presence of man-made structures [24] (for more explanation and equations, refer to Appendix A). It uses the GLCM contrast and rotation-invariant anisotropic measurement in order to characterize built-up areas [23]. PanTex was extracted using the Massive Spatial Automatic Data Analytics (MASADA) tool [25]. We employed several window sizes, i.e., 13, 27, 53, and 105, for comparison. We extracted PanTex with enhancement by histogram standardization, since this feature is highly dependent on the contrast images. Beyond PanTex, we extracted GLCM [9,23] using several window sizes, namely 13, 27, 53, and 105, to examine which size has the best performance. In general, the larger the window size, the higher the computational cost. Thus, we limited the window size to max. 105. GLCM was calculated for all original bands, i.e., mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation. We have done several experiments with different directions, and 1,1 is the best direction according to the accuracy. We tested also the rotationally invariant GLCM. However, the process was very resource-consuming, yet the results were not significantly different [17]. Therefore, we decided to use 1,1 as the direction to save computational time, which also had the best accuracy.

LBP characterizes the spatial distribution of the local image texture as being rotation-invariant, making it robust against greyscale variation in the images [26]. This is important for the image classification of slum areas, since slums have irregular patterns. The parameters were selected based on a previous study [27]. In total, five LBPs were examined, which are the LBPs with radius of 1 and 8 neighbor points ($LBP_{8,1}^{riu2}$), radius of 2 and 8 neighbor points ($LBP_{8,2}^{riu2}$), radius of 3 and 8 neighbor points ($LBP_{8,3}^{riu2}$), radius of 2 and 16 neighbor points ($LBP_{16,2}^{riu2}$), and radius of 3 and 24 neighbor points ($LBP_{24,3}^{riu2}$). The histogram was extracted by a 105×105 window size. The window size was chosen based on the best GLCM window size. However, as input for the classification, we picked the best LBP feature to prevent unnecessarily high-dimensional feature vector. To capture the complexity of slum morphologies, a morphological feature was employed using attribute profiles partial reconstruction (APPR) [28]. The main advantage of partial reconstruction is that it only reconstructs the immediate surrounding area of larger areas [29], resulting in a better spatial model of the image and an improved classification performance [28]. For the input, we used the NIR band, since it has a high contrast between vegetation and built-up areas. Next, the intensity of the image was rescaled to the 0–10 grey level range to reduce computational cost [29]. We set three parameters, which were the area of the region, the standard deviation of grey levels in the region, and Hu's first moment invariant. For each parameter, we selected three values. The area parameter is $\lambda_a = [50, 200, 500]$, standard deviations is $\lambda_s = [0.1, 0.3, 0.5]$, and moment invariant is $\lambda_i = [0, 0.1, 0.3]$.

2.3. Feature Selection

After we extracted the features, they were normalized in the range [0, 1]. In total, we obtained 78 features for differentiating slum and nonslum areas as input for the feature selection. Table 3 presents the features and number of bands, and the suffix number shows the window size.

Table 3. Feature and number of bands.

Features	Number of Bands
Original band	4
NDVI	1
PanTex with contrast adjustment 13	1
PanTex with contrast adjustment 27	1
PanTex with contrast adjustment 53	1
PanTex with contrast adjustment 105	1
GLCM 105	32
LBP	19
APPR	18

Hence, we conducted feature selection to select only the most informative features and to reduce the data dimensionality [30]. From an application context, this is important, improving the accuracy, reducing computation time, increasing the simplicity [31], and preventing overfitting [32]. The simplest feature selection method is SFS [30,31]. This algorithm is commonly operated [33] and popular [34]. SFS is a greedy strategy that decreases the number of states to be searched by applying a local search [34]. It is the bottom-up approach, which starts with zero features and iteratively adds more features that have not been added to the feature set, and applies a selection function to assess whether the features are obtaining the best result [30,31]. The feature that has the maximum score is added to the set of the best features. The score is based on the HSIC score to measure the dependence of the input features and the label [13].

The HSIC score measures the resemblance of the kernel matrix K (the feature kernel) as the input with kernel matrix L (label) as the output. In the beginning, the HSIC criterion was calculated for all features. The feature that had the biggest HSIC score is added to the “set” and is excluded for the next calculation. Then, it will continue calculating the score without the prior selected feature until the HSIC score is stable or reduced. We randomly selected 75% (2440 pixels) as the training set for this process to reduce computation time. We set the maximum number of features to the 35 best features to avoid high computational costs. To compare, we examined the result without feature selection.

2.4. Classification

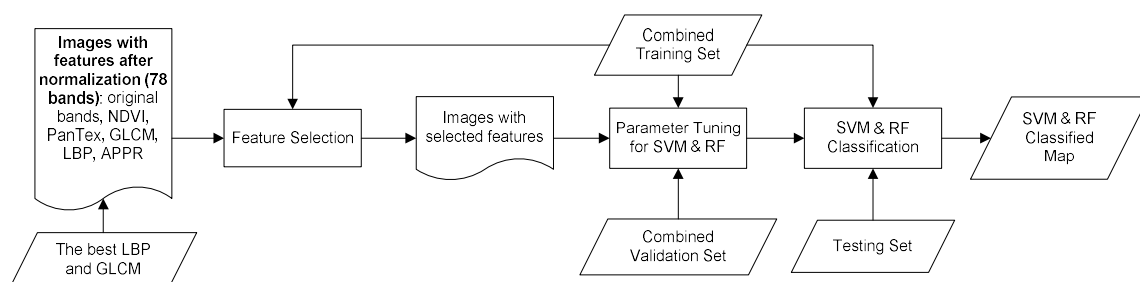
Classification using SVM and RF was done in R. We took 10 tiles of approximately 500×500 m from the Pleiades image based on city planning documents. Then, we generated approximately 100 random points in each tile. We used 30% of the set for training and validation and 70% for testing. We did this on purpose, as in a ‘real-word’ (urban planning) application, training data is scarce (high cost for collecting ground data), in particular when aiming to classify a large area (e.g., an entire city). However, most ML studies use a large amount of training data to obtain high accuracies, which is not realistic for slum mapping programs: if we already know the location of slums, we do not need to classify them.

Next, we randomly chose approximately 30 points that represent slum and nonslum characteristics in each tile. Then, we combined all the selected points from all tiles into one set. The rest of the points in the tiles were used for testing. The prior selection of 30% for training and validation were split into training (80%) and validation (20%). These sample were selected randomly. The validation set was used for tuning parameters of the classifiers. From the points, we made a 1-m buffer to generate polygons to increase the number of pixels for training and testing. Table 4 shows the training, validation, and testing set allocation.

Table 4. Training, validation, and testing set numbers.

Kelurahan	Area	Training and Validation Set	Training and Validation Pixel Number	Testing Set	Testing Pixel Number
Antapani	1003 × 1004	28 polygons	349	69 polygons	866
Babakan	1002 × 1002	31 polygons	385	50 polygons	635
Campaka 1	1004 × 1004	32 polygons	400	63 polygons	790
Campaka2	1002 × 1002	30 polygons	374	49 polygons	608
Cigondewah	1002 × 1004	36 polygons	455	69 polygons	866
Pasir Impun-1	1005 × 1006	36 polygons	453	41 polygons	519
Pasir Impun-2	1003 × 1003	29 polygons	350	44 polygons	557
Sekejati	1002 × 1007	35 polygons	434	61 polygons	753
Tamansari 1	1002 × 1009	32 polygons	398	54 polygons	679
Tamansari 2	1002 × 1001	36 polygons	450	59 polygons	741
Number of pixels in the training and validation sets: 4048; Number of pixels of all training sets (80%): 3238; and of all validation sets (20%): 810				Number of all testing sets: 7014 pixels	

Before the classification, we tuned the parameters by grid search to improve the classifiers. For the grid search, we used the validation set to inspect the best combination of C and γ for SVM and Mtry (number of features selected when generating a tree) and Ntree (is the number of trees generated) for RF. Furthermore, C is a regularization parameter to control the penalty between the errors and generalization capability [16]. If C is too small, it allows many errors and the classifier will not fit the data [16]. In contrast, SVM will overfit the data and have low generalization ability if C is too large [16]. The kernel width or γ is inversely proportional to the variance of the radial basis function (RBF) kernel [35]. It will determine the distance to select the support vectors. In SVM, we randomly set 900 combinations of C and γ for one-time tuning. The first tuning of C ranges from 10^{-1} – 10^5 and γ ranges from 10^{-1} – 10^5 . This allowed analyzing the trend of accuracy, optimizing the C and γ range, and selecting the best combination with the highest accuracy. For RF, we determined 400 combinations of Mtry and Ntree, where Mtry ranged from 1–78 for the model without SFS and 1–35 for the model with SFS, with an interval of 4, and Ntree ranged from 100 until 2000 with an interval of 100. After optimization, the classifiers were tested for each tile. Figure 4 shows the process for classification and feature selection.

**Figure 4.** The process of feature selection and classification.

2.5. Evaluation of Machine Learning Slum Mapping

The application potential of ML slum mapping is evaluated quantitatively and qualitatively. For the qualitative analysis, we compared the classified map, strengths and weaknesses, and the perception of stakeholders. Meanwhile, the quantitative analysis used several statistics, i.e., overall accuracy (OA), time, kappa, correctness, completeness, and F1 score based on the confusion matrix (CM). CM consists of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Figure 5 illustrates possible classification results. Figure 6 illustrates the evaluation framework of this study.

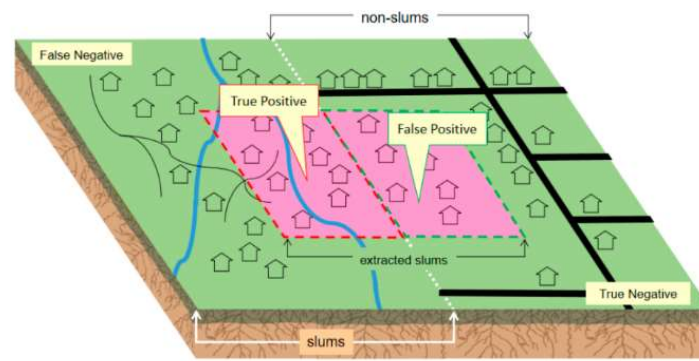


Figure 5. Confusion matrix illustration [6].

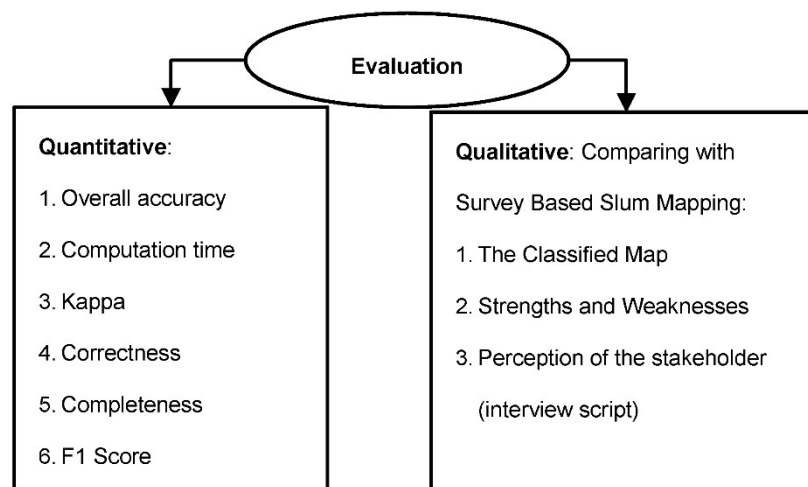


Figure 6. Evaluation framework of the application potential of machine learning-based slum mapping.

Overall accuracy is defined as:

$$\text{Overall Accuracy (OA)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

Kappa measures the overall agreement of a matrix [36], and it is defined as:

$$\text{Kappa} = (\text{observed accuracy} - \text{expected accuracy}) / (1 - \text{expected accuracy}) \quad (2)$$

Moreover, correctness (precision) and completeness (recall) are commonly used accuracy assessment measures [6,7,32]. Correctness measures the reliability of the slums detected, while completeness measures the ability of classifiers to retrieve the areas defined as slums [7]. Correctness and completeness are calculated as:

$$\text{Correctness} = \text{TP} / (\text{TP} + \text{FP}) \quad (3)$$

$$\text{Completeness} = \text{TP} / (\text{TP} + \text{FN}) \quad (4)$$

In addition, the F1 score (recurrent multiresolution convolutional networks for VHR image classification), another common accuracy measure [37], is measured as the harmonic mean of precision and recall, as follows:

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

2.6. Experimental Setup

To assess ML slum mapping in support of the national target, an experimental setup was designed to examine whether a methodology developed on 10 small tiles would allow to be transferred to a larger area (Figure 7—the larger area has the number 11). This scenario used tile 1 to 10 (Figure 7).

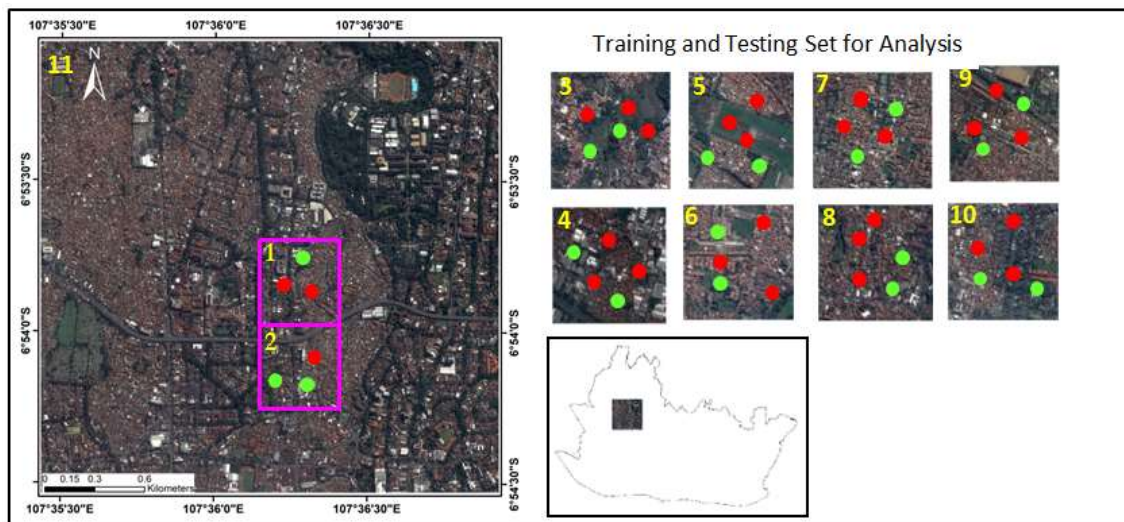


Figure 7. The setup. The analysis is conducted for tiles 1–10. Tile 11 is the larger image that we want to classify. The green and red dots illustrate the samples for non-slums and slums respectively used for the analysis.

3. Results

3.1. GLCM and LBP Assessment

Before we combine the features for classification, GLCM and LBP features that have many bands were analysed at the beginning to save computation time (Table 5 presents the accuracies based on GLCM features using RF for all images). The suffix of GLCM refers to the window size used. The accuracy increased with increasing window size. The GLCM with a window size of 105×105 pixels had the highest accuracy; thus, it was chosen to be combined with other features.

Table 5. Comparison of OA for GLCM features by RF in all tiles.

GLCM 13	GLCM 27	GLCM 53	GLCM 105
72.7%	77.7%	82.1%	83.8%

Table 6 provides the accuracy assessment for LBP features for several types of radii and neighbor points. The histogram LBP was calculated for the 105×105 window size (the best GLCM window size). $LBP_{16,2}^{riu2}$ obtains the highest accuracy. Thus, it was selected to be merged with other features.

Table 6. A comparison of the overall accuracy for LBP by RF.

$LBP_{8,1}^{riu2}$	$LBP_{8,2}^{riu2}$	$LBP_{8,3}^{riu2}$	$LBP_{16,2}^{riu2}$	$LBP_{24,3}^{riu2}$
81.3%	81.1%	81.2%	81.6%	80.7%

3.2. Sequential Feature Selection

This process evaluates the feature relevance to the label. It leads to better performance and saves time in classification. We set a maximum of 35 features to be selected from the total of 78 features.

However, after selecting the 32nd feature, the maximum HSIC score was obtained (Figure 8), so the process was stopped. Table 7 presents the best feature set, where Pantex, LBP, GLCM, APPR, and the green band were the most significant bands.

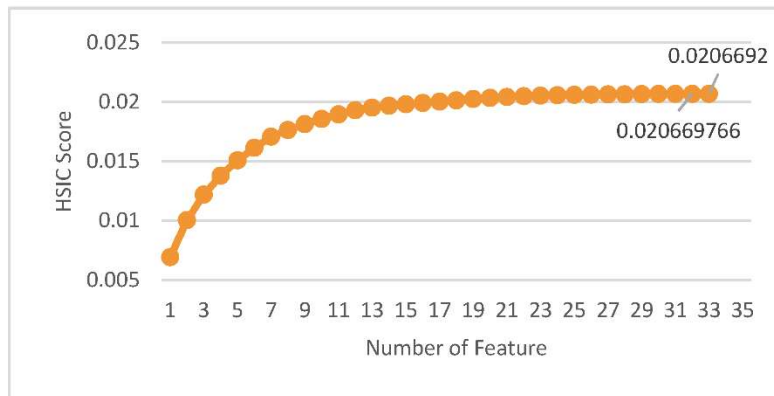


Figure 8. HSIC score against the number of features.

Table 7. The 32 selected features.

No.	Features	No.	Features	No.	Features
1	PanTex window size 105	12	GLCM Dissimilarity Band-1	23	GLCM Entropy Band-3
2	PanTex window size 53	13	LBP	24	GLCM Entropy Band-2
3	LBP	14	LBP	25	APPR area 200 opening
4	PanTex window size 27	15	GLCM Entropy Band-1	26	LBP
5	LBP	16	GLCM Dissimilarity Band-2	27	GLCM Correlation Band-2
6	LBP	17	GLCM Variance Band-1	28	GLCM Mean Band-1
7	GLCM Homogeneity Band-1	18	GLCM Variance Band-2	29	Green Band
8	GLCM Homogeneity Band-2	19	GLCM Dissimilarity Band-3	30	GLCM Second Moment Band-1
9	GLCM Homogeneity Band-3	20	GLCM Variance Band-4	31	LBP
10	PanTex window size 13	21	GLCM Correlation Band-1	32	GLCM Correlation Band-3
11	GLCM Correlation Band-4	22	GLCM Variance Band-3		

Moreover, RF provides an out-of-bag (OOB) error including the feature importance. The OOB error is 0.09%. Table 8 presents the Gini feature importance by the mean decrease.

Table 8. Feature importance with Gini index.

No.	Feature Type	Mean Decrease (Gini)	No	Feature Type	Mean Decrease (Gini)
1	PANTEX 53	57.998	18	GLCM—Variance band 1	25.043
2	GLCM—Correlation band 4	52.099	19	GLCM—Dissimilarity band 1	24.272
3	PANTEX 105	44.918	20	GLCM—Variance band 3	24.246
4	PANTEX 27	42.463	21	GLCM—Variance band 2	24.066
5	PANTEX 13	36.494	22	LBP	23.032
6	GLCM—Homogeneity band 1	32.559	23	GLCM—Homogeneity band 3	22.692
7	LBP	30.548	24	GLCM—Homogeneity band 4	22.519
8	GLCM—Correlation band 1	30.173	25	GLCM—Mean band 1	21.647
9	GLCM—Second moment band 3	29.326	26	NDVI	21.6
10	GLCM—Homogeneity band 2	29.014	27	LBP	21.351
11	LBP	27.781	28	LBP	21.181
12	GLCM—Second moment band 2	26.959	29	LBP	20.939
13	GLCM—Variance band 4	26.866	30	GLCM—Second moment band 4	20.866
14	GLCM—Correlation band 2	26.374	31	GLCM—Contrast band 3	20.676
15	GLCM—Second moment band 1	26.148	32	LBP	20.633
16	GLCM—Correlation band 3	26.137	33	GLCM—Entropy band 2	20.408
17	LBP	26.038			

3.3. Support Vector Machine and Random Forest

Because the sequential feature selection (SFS) process is very time-consuming, we compared the performance of SVM and RF with and without SFS (Table 9). The highest accuracy is obtained with

SVM with SFS. However, the results are not significantly different and RF has a stable result with SFS and without SFS.

Table 9. A comparison between SVM and RF overall accuracies with and without SFS.

Without SFS		With SFS	
SVM	RF	SVM	RF
86.5%	85.2%	88.5%	85.2%

Tables 10 and 11 (in bold are the highest and lowest accuracies across all tiles, and accuracy for all merge tiles) present the detailed results for SVM and RF with SFS. After we obtained the significant features, all tiles were classified. The best feature set is employed to tune the SVM parameters, which are $c = 3.16$ and $\gamma = 3.04$. In RF, the highest accuracy was achieved with Mtry and Ntree being 1 and 200, respectively. With those parameters, the RF and SVM was trained and tested for each testing set in each area/tile. For RF, the overall accuracy is 85.18%, ranging between 72.0–93.9%. For SVM, the overall accuracy is 88.5%, ranging from 72.6–92.4% for the different tiles.

Table 10. RF accuracy assessment results. In bold the highest and lowest overall accuracy (OA), and the OA for all merged tiles.

No.	Selected Area	Time (s)	OA	Kappa	Completeness	Correctness	F1 Score
1	Antapani	0.028	0.859	0.709	0.938	0.831	0.881
2	Babakan	0.023	0.938	0.861	0.876	0.941	0.907
3	Campaka-1	0.021	0.882	0.758	0.861	0.941	0.899
4	Campaka-2	0.019	0.799	0.599	0.721	0.8601	0.784
5	Cigondewah	0.022	0.869	0.730	0.804	0.878	0.839
6	Pasir Impun-1	0.020	0.720	0.033	0.176	0.228	0.199
7	Pasir Impun-2	0.020	0.863	0.704	0.815	0.807	0.811
8	Sekejati	0.025	0.806	0.588	0.911	0.789	0.846
9	Tamansari-1	0.023	0.873	0.746	0.845	0.888	0.866
10	Tamansari-2	0.021	0.869	0.738	0.880	0.859	0.869
	All	0.294	0.856	0.712	0.845	0.849	0.847
	Training Time	3.673					

Table 11. SVM RBF result. In bold the highest and lowest overall accuracy (OA), and the OA for all merged tiles.

No.	Selected Area	Time (s)	OA	Kappa	Completeness	Correctness	F1 Score
1	Antapani	0.075	0.895	0.784	0.956	0.868	0.91
2	Babakan	0.057	0.924	0.836	0.936	0.857	0.895
3	Campaka-1	0.067	0.918	0.826	0.950	0.918	0.934
4	Campaka-2	0.061	0.803	0.606	0.727	0.861	0.788
5	Cigondewah	0.072	0.908	0.813	0.935	0.856	0.894
6	Pasir Impun-1	0.054	0.726	0.127	0.294	0.3	0.297
7	Pasir Impun-2	0.053	0.856	0.698	0.875	0.761	0.815
8	Sekejati	0.064	0.908	0.811	0.929	0.914	0.921
9	Tamansari-1	0.066	0.908	0.816	0.891	0.921	0.906
10	Tamansari-2	0.057	0.903	0.806	0.939	0.871	0.904
	All	0.510	0.885	0.769	0.894	0.865	0.879
	Training Time	1.928					

3.4. Classified Slum Map

Figure 9 shows the classification results for each tile. In general, the SVM result is noisier than the RF result, and the highest accuracy (93.8%) is achieved for Babakan by RF; however, some misclassifications still occurred (shown in blue circles).

Tile	Raw Image	Ground Truth	SVM Classification	RF Classification
1. Antapani (TS1)				
2. Babakan (TS2)				
3. Campaka-1 (TS3)				
4. Campaka-2 (TS4)				
5. Cigondewah (TS5)				
6. Pasir Impun-1 (TS6)				
7. Pasir Impun-2 (TS7)				

Figure 9. Cont.

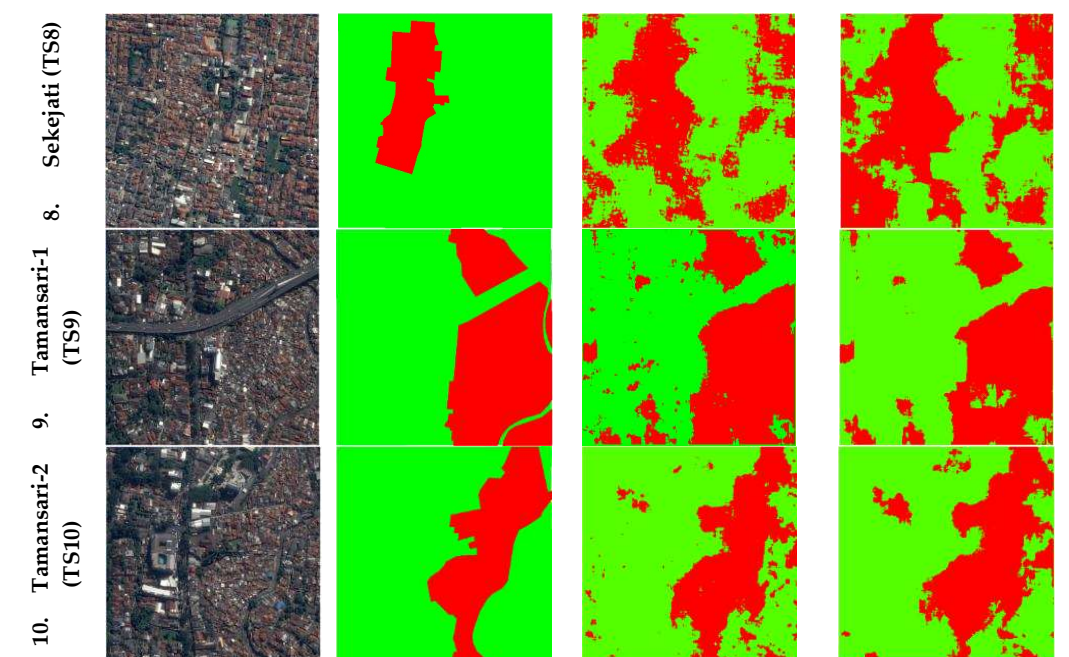


Figure 9. Comparison of classification results and ground truth; slums are in the red and green are non-slums. Blue circles show an example of misclassification in the tile with the highest accuracy.

3.5. Extending the Approach to a Larger Area

Although the overall accuracy of SVM is higher than RF, the classified map of SVM is noisier. Therefore, we selected the RF-classified map with the feature selection method (Figure 10). Moreover, we also did postprocessing to remove salt-and-pepper noise; we set the threshold as 0.135 ha, as the minimum size of slum areas as stated by the Ministry of Public Works and Housing in the interview. Hence, the slums smaller than 0.135 ha were removed.

It was difficult to assess the accuracy, since we do not have ground truth points for the entire area except for the testing set (a small part of this image). Moreover, Google Street View in Bandung city only covers the main roads, with mainly shops and offices. Slums in Bandung are mostly adjacent to formal areas and are usually located behind main roads, and are therefore not shown on Google Street View. In addition, the morphological similarity of slum and non-slum kampungs (in an image) introduces uncertainties for generating reference data. As we can see in the blue circle of Figure 10 (below left), the morphological structures of the building are relatively small and very dense. Thus, such areas are classified as slums. However, in the yellow circle in Figure 10, the public cemetery is also classified as a slum, because its patterns and small structures are similar to those in slums. However, success was achieved in classifying formal residential areas as non-slums (pink circle in Figure 10). Nevertheless, to evaluate the results for the larger area, we used visual interpretation, while being aware of the uncertainties described above. Overall accuracy reached 87.5%. To obtain the broader view of algorithm performance, Kappa, completeness, correctness and F1 score values were used, indicating in general lower performance and pointing to the fact that several slums were wrongly classified. However, there is a high uncertainty as to whether the visual image interpretation is correctly labeling these areas. Table 12 presents the confusion matrix of the result.

Table 12. Confusion matrix.

	Actual	
	Slums	Non-slums
Predicted Slums	18	16
Predicted Non-slums	9	157

From the confusion matrix, RF predicted nonslum better than slums. From 27 slum and 173 nonslums, RF predicted 18 slums and 155 nonslums correctly; thus overall, giving an accuracy of 87.5%. Moreover, Table 13 presents the complete accuracy assessment for this area.

Table 13. Accuracy assessment of the larger area.

Overall Accuracy	Kappa	Completeness	Correctness	F1 Score
87.5%	0.518	0.667	0.529	0.59

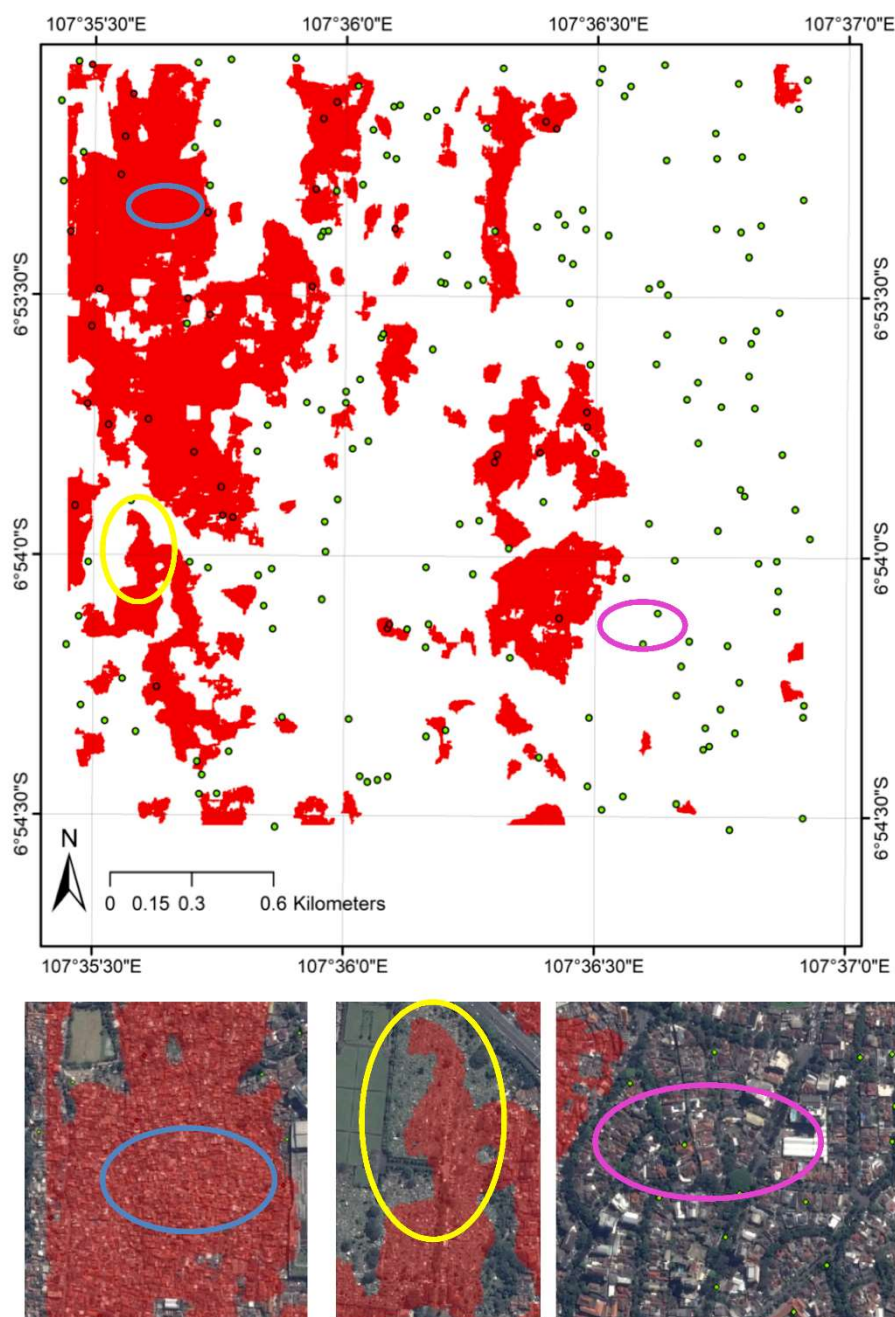


Figure 10. RF-classified map of the larger images with 200 random points and overlaid with the original images (below). The different color circles on the map (upper) correspond to the different circle on top of the satellite images (lower), showing the real condition on the ground.

3.6. Comparing the Classified Map with the Survey-Based Slum Mapping Map

To assess the potential of ML-based slum mapping for slum upgrading programs, we compared the result of this approach with the survey-based slum mapping (SBSM) result (Figure 10).

Figure 11 shows differences between the two mapping products. Areas of small buildings are classified as slums by RF (see circles 1, 2, 4), while SBSM excludes them. Moreover, vegetation and large formal buildings in circle 3 are classified as slums by the surveyor, while RF does not include them. In addition, in circle 5, the surveyors generalized the slum area, while RF resulted in a more detailed and accurate slum map.

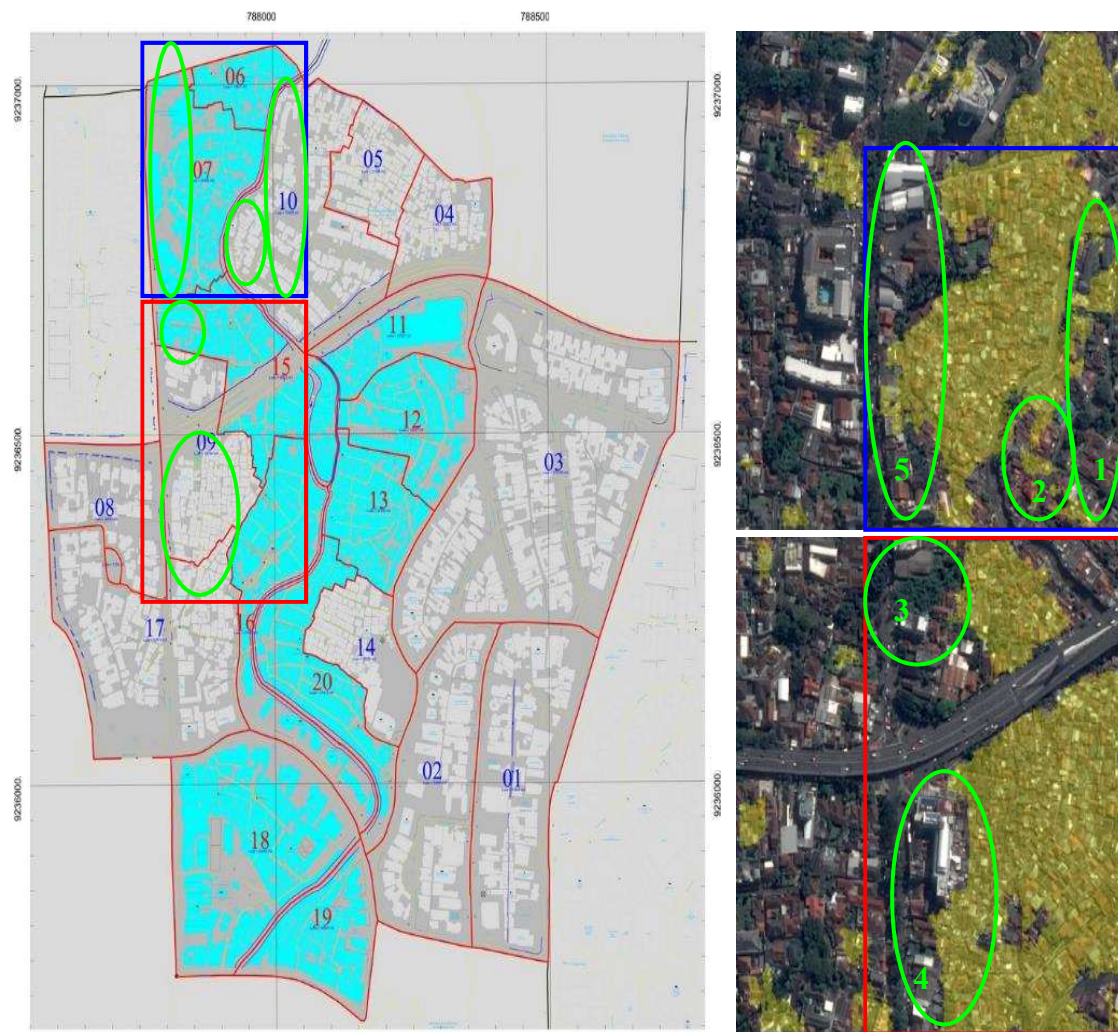


Figure 11. Comparison of the SBSM (left) and RF-classified image (right top and below). The red and blue squares show the same location, and the green circles show the differences [20].

3.7. Strengths and Weaknesses

Table 14 analyses the utility of ML-based slum mapping compared to survey-based slum mapping in support of slum upgrading programs.

Table 14. Comparison of machine learning-based slum mapping and survey-based slum mapping (Currency: 1 Euro = 17,024.06 IDR at 26 August 2018).

Factors	Machine Learning-Based Slum Mapping (MLBSM)	Survey-Based Slum Mapping (SBSM)
Cost	<ol style="list-style-type: none"> Human resources: Planning expert = 1 person Infrastructure expert = 1 person GIS expert = 1 person Remote sensing expert = 1 person Programming expert = 1 person Surveyors = 20 persons Total estimate for one year = 480,000,000 IDR = 28,195.38 EUR (based on [38]) Infrastructures Computer = 127,680,427.95 IDR Images = 28,552,821.53 IDR Software: e.g., QGIS and SagaGIS MASADA, R Total budget = 156,258,055.77 IDR 9177 EUR Time = 3 months 	<ol style="list-style-type: none"> Human resources: Team Leader = 1 person Infrastructure expert = 1 person Planning expert = 1 person Community development expert = 1 person Economic development expert = 1 person Safe guard expert = 1 person Data officer = 1 person Surveyor: 130 persons Total estimate for one year = 5,922,000,000 IDR = 347,925.22 EUR (based on [39]) Infrastructures Computer = 10,000,000 × 3 = 30,000,000 IDR GIS Software = QGIS and SagaGIS Total budget = 30,000,000 IDR = 1762.21 EUR Time = 12 months
Human resources	<ol style="list-style-type: none"> Remote sensing expert GIS expert Programming expert Urban planning expert Infrastructure expert Surveyors 	<ol style="list-style-type: none"> Team leader Surveyors (for Bandung city, there are 1620 surveyors) Urban planning expert Infrastructure expert Community development expert Economic development expert Safe guard expert Data officer
Infrastructures	<ol style="list-style-type: none"> High specification computer Very high-resolution satellite images 2.5–0.5 m (such as Pleiades, SPOT) Processing software (GIS, advanced remote sensing software, e.g., Matlab) 	<ol style="list-style-type: none"> Lower specification memory computer than MLBSM method (such as 4 GB RAM) Processing software (GIS, QGIS)
Processing Time	Approximately one month depending on the capacity of the computer, as well as surveys on the field to get the training set.	Approximately six months depending on the capacity of surveyors and participatory process with the community.
Spatial Coverage	With one set of the resources (human, and infrastructures) in 2 months, it possibly produces one city	With one set of the resources (human, and infrastructures) in 2 months, it possibly produces only some parts of the city depending on how large the city is.
Accuracy	88.5% of the reference (ground truth data) by the highest accuracy result from SVM	80% (claimed by ministry); However, it is only an assumption, because they do not have a mechanism for the accuracy assessment. They realized results depend on surveyor's understanding. Limitations are also caused by time and geographic barriers to collect data on the ground, meaning sometimes the surveyor only estimates the data.
Degree of automation	33.33% From the three steps (surveying, making the slum maps, validating), one step (making the slum maps) is automated	0%
Maintenance	The parameter should be adjusted for another city according to the local slum characteristics	Not relevant

4. Discussion

4.1. Quantitative Analysis

The feature extraction and parameter settings are important in MLBSM. In the assessment of the GLCM (Table 6), the largest window size was selected. In general, the larger the window size, the more stable the patterns and the more contextual information is used. This was also confirmed by Wurm et al. [9], emphasizing that a very large kernel size of GLCM has a smoothing effect on the image content, which is very useful for mapping slums (being very heterogeneous on a large scale and rather homogeneous on a small scale) [9]. An increasing accuracy trend along with an increasing window size were also found in [17]. LBP results (Table 7) show that they are not sensitive to the radius and interpolation points.

For the classification results (Table 9), RF had a stable accuracy with and without SFS. This indicates that RF is robust to the Hughes phenomenon, where each decision tree has a random method to select data and features to be classified using the Gini index [40]. Moreover, RF can reduce the required computational resources, since SFS is computationally costly. From Table 8, features that had the highest mean decrease (Gini) are similar to the selected features by SFS, except for the green band and APPR. SVM and RF did not have a significant accuracy gap. Moreover, the tuning of parameters in SVM is more complex than in RF. In addition, to get the best accuracy, computationally

costly feature selection was needed by SVM. This was also confirmed the finding of Abe et al. [41], in that those algorithms can reach similar accuracies, but RF is less computationally expensive. Further studies should explore other computational feasible methods, e.g., Rahmati et al. [12] added boosted regression trees (BRT) as they are capable of rapidly producing accurate results.

PanTex (window size 105) was the most important feature in the set. This confirms the findings of [42]. However, PanTex strongly depends on the contrast level, thus contrast enhancement is important to distinguish slums. From the 18 bands of APPR, only an area of 200 pixels with an opening operator is useful to distinguish slums. This might be caused by the simple rescaling (0–10) of the pixel input. Thus, the result was not significant to characterize the morphology of slums. Moreover, only 18 attribute profiles were evaluated; further analysis could explore more morphological profiles for slum mapping. In addition, the green band (original spectral bands) is important, which might relate to the potential of characterizing vegetation besides other land cover types. Furthermore, several GLCM bands (dissimilarity, homogeneity, entropy, and second moment and variance) and LBP histograms have a significant contribution to distinguish slums and non-slums. GLCM was restricted to a window size of maximum 105 to reduce computation time. Thus, larger window sizes could be beneficial for improving the mapping accuracies.

The tuning parameter of SVM RBF is complex due to the absence of a clear rule to determine the range of C and γ . This problem was also stressed by Adiningrat [43]; the common approach is trial-and-error for defining the range. Regarding RF, the process is quite simple and resulted in small number of features and trees. Thus, in the training and testing processes, the model is computationally efficient. In the validation process, the best parameter reached up to 100% accuracy, while in the testing set, the maximum accuracy achieved was 88.5% and 85.6% for SVM and RF, respectively. It is a common condition in ML that the accuracy based on the test data is lower than that of the training data. Moreover, the uncertainty and inconsistency in slum characteristics between the training and testing set added to the problem, since the experiment only used 30% of the data for the training. Moreover, there were uncertainties in exacting slum boundaries in several tiles, as boundaries tend to be fuzzy. Uncertainties are inevitably happening in assessing the accuracy [6] and further increasing when aiming for change detection (e.g., in the context of long-term slum monitoring programs [44]). For tuning parameters, a grid search was used, causing difficulties to obtain the best parameter. Therefore, there is a need to use better techniques such as k-fold cross validation to optimize parameters.

4.2. Qualitative Analysis

4.2.1. Classified Map

Due to working with a rather standard computer (16 GB RAM, Intel core i7 2.6 GHz, and 230 GB hard disk), we limited the larger subset to only 5500×5000 pixels or 2.25×2.25 km, which reduced the possible variation in slum characteristics. Extending this work to city scale would require big data techniques and additional computing power.

Both SVM and RF classification results show misclassifications, particularly for small formal structures. This is due to similar morphological characteristics and roof material of both categories, thus with an image, we can only capture morphological slums [45,46]. Furthermore, the uncertainty of slum boundaries plays a role. In Pasir Impun-1 (Figure 12, right), slums and non-slums have fuzzy boundaries. Figure 12 (left) shows the ground truth (identified by surveyors in the fields). This uncertainty was also reported in the literature as influencing the accuracy [47]. The surveyors affirmed that in some areas, they were in doubt to determine the slum boundary due to mixed condition within the area (mix of slums and non-slums), yet all delineated polygons have crisp boundaries.

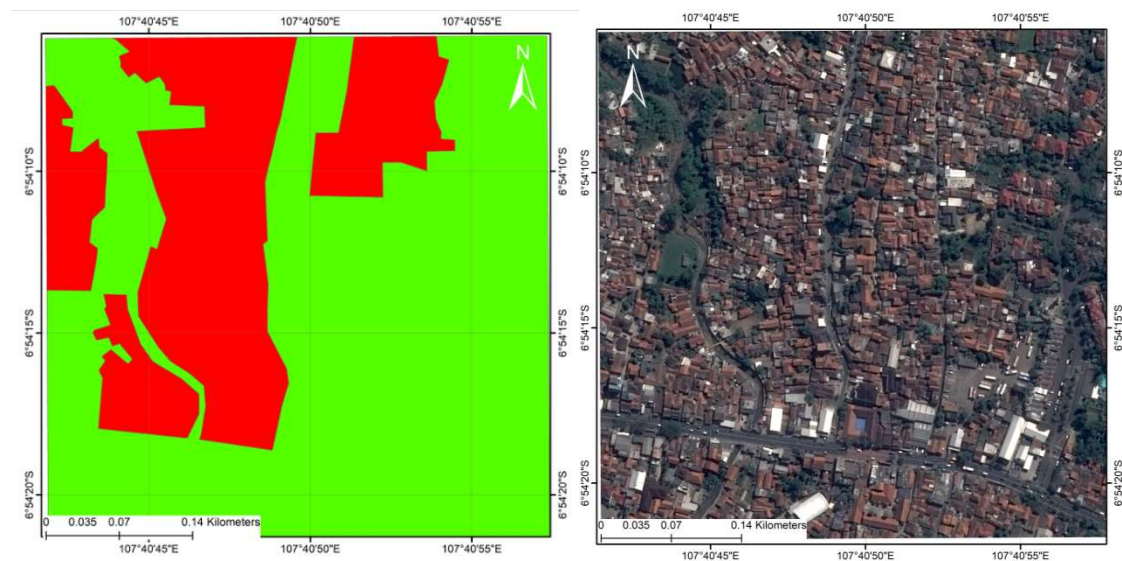


Figure 12. Uncertainty of slum boundaries in Pasir Impun. Image (right), ground truth survey map (left). Slums are in the red, green represents non-slums.

For the ground survey, no clear rule exists to determine the size and arrangement of non-slum areas within slums. This is an important issue in generating ground truth data, since slums are defined at the settlement level that includes also infrastructure and facilities. Thus, only if an area of non-slum within a slum area is more than 500 m², it is labelled as non-slum. Also, we determined 6.5 m as the maximum threshold for the road width to be considered as a slum, as stated by the Ministry of Public Works and Housing [48] (also see Section 2.2). Overestimation also happened due to the large window size used for feature extraction (i.e., GLCM, LBP, PanTex), as was also stressed by Sliuzas et al. [49].

Our work only included the settlement and object levels of the GSO [50], because these can be described by image features. To implement the environs level, we would need to include additional data such as hazard and land use maps as features to explain location and neighborhood characteristics. In a recent study, Jochem et al. [51] used vector features such as points and polygons as features that could add information which is not available in the images. However, doing this might also increase uncertainties due to quality issues with such data.

SBSM shows inconsistencies (Figure 9), e.g., vegetation and large formal buildings are included in slum areas. The generalization of SBSM maps omits details and results in inaccurate delineations for some areas (also depending on the surveyor's experience). However, based on surveyor experience, SBSM could distinguish slum and non-slum small buildings in the field, while ML identified small structures as slums. Therefore, we conclude that both methods have shortcomings. Thus, a combination of both ML-based slum mapping and SBSM may be the best solution for supporting slum upgrading programs. ML, combined with other advanced remote sensing technology (e.g., working with large image-based feature sets), is a promising development. Moreover, in slum mapping, the employment of ML is becoming popular [9,17,32,52,53].

Apart from the spatial resolution, the temporal resolution of the sensor is very important [54] to regularly evaluate the planning strategies and to avoid time- and cost-consuming ground data collection. Recent advances in remote sensing have increased the opportunity to monitor urban change and its consequences on complex urban sociotechnical systems [55]. Therefore, such information would enable stakeholders to make more informed decisions and to reduce negative impacts on the environment (ibid). Particularly in a developing country, a lack of finances is a main limitation to gaining complete and up-to-date base data, even for major cities. Moreover, monitoring and comparisons across a city or country are easier to realize using remote sensing methods [54]. Although the accuracy of information extraction by remote sensing images has generally improved, there are limitations for using remote sensing in analyzing urban sustainability due to the complexity of

the urban landscape, limited computer capacity, shortcomings in the methods, and complexities in integrating multisource data [55]. Hence, to take the full benefit of the diversity and the potential of remote sensing data, there is a need to establish better strategies and approaches and improve the hardware and algorithms. Moreover, object-based image analysis (OBIA) could provide suitable aggregation levels for slum mapping [56]. OBIA has been criticized for its complexity in selecting the rules and parameters [57]; however, besides producing data at a suitable aggregation level (segments, not pixels), OBIA postclassification processing could be beneficial. We applied postprocessing, using a specific threshold to delete the ‘salt-and-pepper’ noise in the end product. This calls for a possible combination of OBIA and ML approaches, which could produce outputs which are more similar to human interpretations (better fulfilling the demands of stakeholders). Furthermore, the information from OBIA’s segmentation is more contextual and time-saving in processing.

The MLBSM can only examine slum appearance from an aerial perspective. Therefore, it produces maps that indicate the possible presence of a slum. Ground truth surveys are needed to validate the slum areas. The slum upgrading programs require iterative data collection process such as multiple building level surveys throughout the implementation phases of a project. Thus, such surveys will subsequently improve the initial slum boundaries from MLBSM.

4.2.2. Strengths and Weaknesses

Strengths and weaknesses (Table 12) were analyzed in several dimensions. The analysis shows clear tradeoffs between human and technical resources. MLBSM requires much fewer and a different type of experts from SBSM. In terms of maintenance and transferability to other regions in the country, MLBSM needs to be optimized for each new context. Feature selection or parameter tuning needs to be conducted again to get optimal results, particularly if regions have different slum characteristics, such as in the Eastern region of Indonesia, with its lower population density. For SBSM, optimization is not relevant, since surveyors from local people in the region should be familiar with the condition of the slums. However, surveyors need to be trained to improve the consistency of their mapping.

Although SBSM resulted in very detailed data, this method is extremely time- and effort-intensive [58] and may also be inaccurate. Meanwhile, MLBSM can produce fast slum ‘indication maps’ for the city and would allow monitoring of the slum developments in the following years. As stated by Patino and Duque [59], remote sensing images are essential and capable sources of information on the urban morphology and changes over time.

An MLBSM map is useful to obtain initial data of slums. Ground surveys can further refine the initial map to improve its accuracy and consistency in support of upgrading programs. However, for the implementation in a large region such as Indonesia, MLBSM needs to be adjusted for different contexts in correspondence with local urban and slum characteristics.

4.2.3. Perception of the Stakeholders

The final goal of Indonesia’s slum upgrading program is to develop livable cities; specifically, to fulfill the target that has been set to have cities without slums in 2019. The participatory slum mapping process involves the community in the neighborhoods, facilitators, and the local government in a forum, where the information of local conditions is gathered, discussed, and measured based on slum indicators. Finally, all information is arranged as base data for a neighborhood plan and a detailed engineering design document.

However, all stakeholders criticized the indicators. For example, the Ministry staff are not satisfied with the inundation indicator due to its complexity. The municipal staff thought that the absence of green space might be good to be included as slum indicator. The academics criticized that several indicators are meaningless, e.g., safe drinking water, drainage system connection, fire protection, and building permits (they do not distinguish slums from non-slums). Such critiques point to a need to review the SBSM indicators. In this review, the MLBSM-classified map could be used as input, since it is based on a conceptual definition of slums in the field (LSO).

Time is also a main issue. The SBSM depends on the amount of slum areas. Commonly, the survey area is much larger than the capacity of the surveyors. This affects the quality of the planning document for upgrading programs. For example, some boundaries in the SBSM do not follow physical boundaries such as roads, buildings, or rivers. All respondents agreed that such a map might cause problems for slum upgrading. Meanwhile, the Ministry does not have a process to validate the slum maps. However, they commonly check the slum areas before upgrading. In the validation process, the municipality was asked to make both aerial (drone) and terrestrial videos. The validation is done to prevent the overreporting of slums to get more funds. It allows for more accurate calculations of the required infrastructure to be upgraded as related to the allocated funds. Until 2017, after three years of implementation, the achievement of slum upgrading programs was 11,565 ha out of the target 38,431 ha, or 30.1% [60]. Several problems were identified, such as incorrect delineation, misunderstanding between stakeholders in the implementation, social problems, technical mistakes, and misallocation of the budget. However, the Ministry remains optimistic about reaching the ultimate target by 2019.

By the end of 2017, the slum mapping in all urban areas in Indonesia was completed through SBSM. The Indonesian government is now focusing on upgrading these areas. However, empowerment of the local governments in Indonesia through training, with a focus on prevention and improvement of slum areas, is still required. Considering the required accuracies, the municipality stated that an accuracy of 88.5% of MLBSM is adequate to identify slums, since field validation will be conducted. By contrast, the Ministry expects that the results can be directly used without field checking (to avoid additional budget). However, the level of noise in the MLBSM maps results in some potential users being reluctant to adopt them. In addition, as the SBSM data is complete, the government is currently not considering alternative approaches such as MLBSM. Besides, the development of MLBSM would require an extensive effort and budget, since such a system would be developed from scratch, requiring substantial investments in geospatial infrastructure and capacity. Yet, the Ministry is not certain about the long-term utilization and capacity of this system, being unfamiliar with machine learning and remote sensing.

Slum data is sensitive data, and the use of nonvalidated data would reduce the acceptance by different stakeholders. There is a need include good metadata to explain the data, concerning their limitations and an explanation how to interpret the data. An initial higher investment for MLBSM could produce more consistent and timely data and would allow future monitoring. However, the MLBSM could not use all SBSM indicators. Thus, as mentioned by Kuffer et al. [61], the combination of community-driven data and spatial information from remote sensing imagery is most optimal in support of pro-poor policy.

To promote MLBSM, more user-friendly software interfaces are required that allow local geospatial experts to run such systems and combine them with community-based information. This would allow monitoring changes after implementing upgrading programs. However, for a national implementation, the MLBSM needs to be adjusted for different contexts [54]. Figure 13 illustrates the workflow of the MLBSM approach prior to implementation for slum upgrading programs.

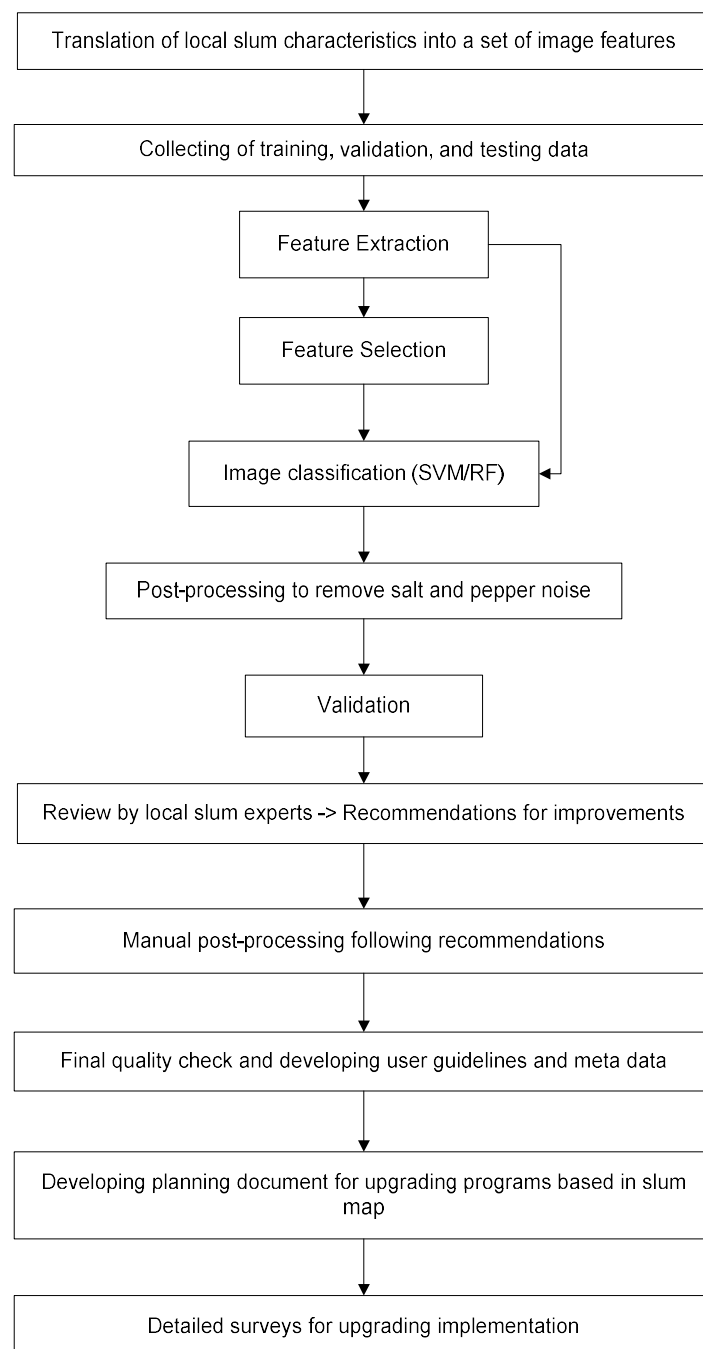


Figure 13. Machine learning-based slum mapping (MLBSM) workflow.

5. Conclusions

Developing a contextual, machine learning-based slum mapping (MLBSM) approach requires a good understanding of the specific context. Based on such a conceptualization, image-based features are proxies to slum maps made by remote sensing imagery and machine learning. Feature selection is an important step to ensure working with the best set and achieving high accuracies; however, it is computationally costly. From the selected features, contextual features are the most significant for slum mapping. For the case of Bandung, the highest accuracy (88.5%) was obtained with SVM. However, the classified map is noisier than the RF map. To implement MLBSM, we need to consider the cost for providing all the requirement of infrastructure and human resources. MLBSM has a high cost in infrastructure, while survey-based slum mapping (SBSM) has high costs in human resources

and is very time-consuming. Both MLBSM and SBSM require validation before implementation in slum upgrading programs. In combining MLBSM and SBSM in support of slum upgrading programs, MLBSM could help the government to produce consistent maps, using SBSM for training and validation. A fundamental prerequisite for MLBSM is the involvement of stakeholders, in particular the local communities, to build local knowledge and local acceptance.

Author Contributions: Conceptualization, G.L., M.K. and R.S.; Methodology, G.L., C.P. and M.K.; Software, G.L.; Validation, G.L.; Formal Analysis, G.L.; Investigation, G.L.; Resources, C.P.; Data Curation, G.L. and C.P.; Writing-Original Draft Preparation, G.L., M.K., R.S. and C.P.; Writing-Review & Editing, G.L., M.K., R.S. and C.P.; Visualization, G.L.; Supervision, M.K., R.S. and C.P.; Project Administration, M.K.; Funding Acquisition, G.L. and M.K.

Funding: Nuffic Neso Indonesia supported this research. We acknowledge the European Space Agency (ESA) for providing the image data through its Third Party Missions.

Acknowledgments: We thank the Ministry of Public Works and Housing, the Municipality of Bandung, an academic from a local university, and a surveyor from the survey-based slum mapping program in Indonesia for all their support for this study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The Pantex is able to extract structural characteristics of buildings, by the fact that buildings possess shadows that lead to high local contrast [23].

The brightness band is used for the input due to the absence of a panchromatic band, according to the assumption that the built-up structures are the most shining features in the optical bands [25], with the formula:

$$\text{Brightness} = \text{Max}(\text{Visible Bands}) \quad (\text{A1})$$

Anisotropic rotation-invariant texture measure is employed by PanTex to resolve the suboptimal displacement problem in GLCM [62]. PanTex uses the min or fuzzy \cap to replace averaging, as defined in the formula:

$$\text{tx}(\text{built up}) = \cap_i \text{tx}_i; i \in [1 \dots n] \quad (\text{A2})$$

where n is the number of displacement vectors (distance and angle combinations) [62]. In additions, the intersection operator (min) between the textural measure from different displacement vectors is used [23] as formulated below:

$$f(\text{built up}) = \wedge \{ \text{tx}_i, \text{tx}_2, \dots, \text{tx}_n \}; i \in [1] \quad (\text{A3})$$

References

1. Ministry of Public Works and Housing (Kemen PUPR). *Slum Upgrading Programs*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2015; pp. 6–39.
2. Ministry of Public Works and Housing (Kemen PUPR). *Strategy for Achieving the Target of the Medium-Term Development Plan in 2015–2019*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2015.
3. Ministry of Public Works and Housing (Kemen PUPR). *City Without Slums (KOTAKU) Program Guide*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2016; pp. 1–148.
4. Kuffer, M.; Pfeffer, K.; Sliuzas, R.; Baud, I. Extraction of slum areas From VHR imagery using GLCM variance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1830–1840. [[CrossRef](#)]
5. Kohli, D.; Sliuzas, R.; Stein, A. Urban slum detection using texture and spatial metrics derived from satellite imagery. *J. Spat. Sci.* **2016**, *61*, 405–426. [[CrossRef](#)]
6. Pratomio, J.; Kuffer, M.; Martinez, J.; Kohli, D. Coupling uncertainties with accuracy assessment in object-based slum detections, case study: Jakarta, Indonesia. *Remote Sens.* **2017**, *9*, 1164. [[CrossRef](#)]
7. Duque, J.C.; Patino, J.E.; Betancourt, A. Exploring the Potential of Machine Learning for Automatic Slum Identification from VHR Imagery. *Remote Sens.* **2017**, *9*, 895. [[CrossRef](#)]

8. Kohli, D.; Kerle, N.; Sliuzas, R. Local ontologies for object-based slum identification and classification. In Proceedings of the 4th GEOBIA, Rio de Janeiro, Brazil, 7–9 May 2012; pp. 1–4.
9. Wurm, M.; Weigand, M.; Schmitt, A.; Geiß, C.; Taubenböck, H. Exploitation of textural and morphological image features in sentinel-2a data for slum mapping. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017; pp. 1–4.
10. Sharma, V.; Baruah, D.; Chutia, D.; Raju, P.; Bhattacharya, D.K. An assessment of support vector machine kernel parameters using remotely sensed satellite data. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; pp. 1567–1570.
11. Bachofer, F.; Quénéhervé, G.; Märker, M.; Hochschild, V. Comparison of SVM and boosted regression trees for the delineation of lacustrine sediments using multispectral ASTER data and topographic indices in the Lake Manyara Basin. *Photogramm. Fernerkund. Geoinf.* **2015**, *2015*, 81–94. [[CrossRef](#)]
12. Rahmati, O.; Tahmasebipour, N.; Haghizadeh, A.; Pourghasemi, H.R.; Feizizadeh, B. Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* **2017**, *298*, 118–137. [[CrossRef](#)]
13. Persello, C.; Bruzzone, L. Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2615–2626. [[CrossRef](#)]
14. Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; Schmidt, M.G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* **2016**, *265*, 62–77. [[CrossRef](#)]
15. Kotsiantis, S.B. Supervised machine learning: A review of classification techniques. *Informatica* **2007**, *31*, 249–268.
16. Bruzzone, L.; Persello, C. Approaches Based on Support Vector Machine To Classification of Remote Sensing Data. In *Handbook of Pattern Recognition and Computer Vision*; Chen, C.H., Ed.; World Scientific: Singapore, 2009; pp. 329–352.
17. Mboga, N.O. Detection of Informal Settlements from VHR Satellite Images Using Convolutional Neural Networks. Master's Thesis, University of Twente, Enschede, The Netherlands, 2017.
18. Statistical Agency of Bandung City. *Bandung City in Figures*; 9789791426442; Statistical Agency of Bandung City: Bandung, Indonesia, 2016; p. 350.
19. Municipality of Bandung City. *Areas in Bandung City*; Statistical Agency of Bandung City: Bandung, Indonesia, 2008.
20. Municipality of Bandung City. *Planning Document: Slum Upgrading Programs in Bandung City in Press*; Statistical Agency of Bandung City: Bandung, Indonesia, 2015; p. 282.
21. Tarigan, A.K.M.; Sagala, S.; Samsura, D.A.A.; Fiisabilillah, D.F.; Simarmata, H.A.; Nababan, M. Bandung City, Indonesia. *Cities* **2016**, *50*, 100–110. [[CrossRef](#)]
22. Owen, K.K.; Wong, D.W. An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics. *Appl. Geogr.* **2013**, *38*, 107–118. [[CrossRef](#)]
23. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [[CrossRef](#)]
24. Pesaresi, M.; Ouzounis, G.K.; Gueguen, L. A new compact representation of morphological profiles: Report on first massive VHR image processing at the JRC. In Proceedings of the SPIE Defense, Security, and Sensing, Baltimore, MD, USA, 23–27 April 2012; p. 6.
25. Politis, P.; Corbane, C.; Maffenini, L.; Kemper, T.; Pesaresi, M. *Masada User Guide*; Publications Office of the European Union: Luxembourg, 2017; Volume JRC106667.
26. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
27. Ella, L.P.A.; van den Bergh, F.; van Wyk, B.J.; van Wyk, B.J. A comparison of texture feature algorithms for urban settlement classification. In Proceedings of the IGARSS 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 7–11 July 2008; pp. III-1308–III-1311.
28. Liao, W.; Dalla Mura, M.; Chanussot, J.; Bellens, R.; Philips, W. Morphological attribute profiles with partial reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1738–1756. [[CrossRef](#)]

29. Liao, W.; Chanussot, J.; Dalla Mura, M.; Huang, X.; Bellens, R.; Gautama, S.; Philips, W. Taking optimal advantage of fine spatial resolution: Promoting partial image reconstruction for the morphological analysis of very-high-resolution images. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–28. [[CrossRef](#)]
30. Gutierrez-Osuna, R. Pattern analysis for machine olfaction: A review. *IEEE Sens.* **2002**, *2*, 189–202. [[CrossRef](#)]
31. Marcano-Cedeño, A.; Quintanilla-Domínguez, J.; Cortina-Januch, M.G.; Andina, D. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In Proceedings of the IECON 2010 36th Annual Conference on IEEE Industrial Electronics Society, Glendale, AZ, USA, 7–10 November 2010; pp. 2845–2850.
32. Gevaert, C.M.; Persello, C.; Sliuzas, R.; Vosselman, G. Informal settlement classification using point-cloud and image-based features from UAV data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 225–236. [[CrossRef](#)]
33. Mao, K.Z. Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **2004**, *34*, 629–634. [[CrossRef](#)]
34. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [[CrossRef](#)]
35. Kavzoglu, T.; Colkesen, I. A kernel functions analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 352–359. [[CrossRef](#)]
36. Banko, G. *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory*; International Institute for Applied Systems Analysis: Laxenburg, Austria, 1998; p. 44.
37. Bergado, J.R.; Persello, C.; Stein, A. Recurrent multiresolution convolutional networks for VHR mage Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, 1–14. [[CrossRef](#)]
38. Ministry of Public Works and Housing. *The Minimum Amount of Remuneration for the Construction Workforce in Expert Positions for Construction Consultancy Services*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2018.
39. Directorate Settlement Development. *Salary Adjustment for an Independent Consultant in the City without Slum Programs*; Directorate Settlement Development: Jakarta, Indonesia, 2018.
40. Leichtle, T.; Geiß, C.; Wurm, M.; Lakes, T.; Taubenböck, H. Evaluation of clustering algorithms for unsupervised change detection in VHR remote sensing imagery. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017; pp. 1–4.
41. Abe, B.T.; Olugbara, O.O.; Marwala, T. Experimental comparison of support vector machines with random forests for hyperspectral image land cover classification. *J. Earth Syst. Sci.* **2014**, *123*, 779–790. [[CrossRef](#)]
42. Syrris, V.; Ferri, S.; Ehrlich, D.; Pesaresi, M. Image enhancement and feature extraction based on low-resolution satellite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1986–1995. [[CrossRef](#)]
43. Adiningrat, D.P. *Mapping Dominant Tree Species from Remotely Sensed Image Using Machine Learning Algorithms*; ITC-University of Twente: Enschede, The Netherlands, 2017.
44. Pratomo, J.; Kuffer, M.; Kohli, D.; Martinez, J. Application of the trajectory error matrix for assessing the temporal transferability of OBIA for slum detection. *Eur. J. Remote Sens.* **2018**, *51*, 838–849. [[CrossRef](#)]
45. Wurm, M.; Taubenböck, H. Detecting social groups from space—Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sens. Lett.* **2018**, *9*, 41–50. [[CrossRef](#)]
46. Taubenböck, H.; Kraff, N.J.; Wurm, M. The morphology of the Arrival City—A global categorization based on literature surveys and remotely sensed data. *Appl. Geogr.* **2018**, *92*, 150–167. [[CrossRef](#)]
47. Mboga, N.O.; Persello, C.; Bergado, J.R.; Stein, A. Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sens.* **2017**, *9*, 1106. [[CrossRef](#)]
48. Ministry of Public Works and Housing. *Roads*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2006.
49. Sliuzas, R.; Kuffer, M.; Kemper, T. Assessing the quality of Global Human Settlement Layer products for Kampala, Uganda. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, United Arab Emirates, 6–8 March 2017; pp. 1–4.
50. Kohli, D.; Sliuzas, R.; Kerle, N.; Stein, A. An ontology of slums for image-based classification. *Comput. Environ. Urban Syst.* **2012**, *36*, 154–163. [[CrossRef](#)]
51. Jochem, W.C.; Bird, T.J.; Tatem, A.J. Identifying residential neighbourhood types from settlement points in a machine learning approach. *Comput. Environ. Urban Syst.* **2018**, *69*, 104–113. [[CrossRef](#)] [[PubMed](#)]
52. Vatsavai, R.R. Gaussian multiple instance learning approach for mapping the slums of the world using very high resolution imagery. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1419–1426.

53. Persello, C.; Stein, A. Deep fully convolutional networks for the detection of informal settlements in VHR images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2325–2329. [[CrossRef](#)]
54. Sliuzas, R.; Kuffer, M.; Masser, I. The spatial and temporal nature of urban objects. In *Remote Sensing of Urban and Suburban Areas*; Rashed, T., Jürgens, C., Eds.; Springer: Dordrecht, The Netherlands, 2010; Volume 10, pp. 67–84.
55. Kadhim, N.; Mourshed, M.; Bray, M. Advances in remote sensing applications for urban sustainability. *Euro-Mediterr. J. Environ. Integr.* **2016**, *1*, 1–22. [[CrossRef](#)]
56. Pratomo, J. Transferability of The Generic and Local Ontology of Slum in Multi-temporal Imagery, Case Study: Jakarta. Master's Thesis, University of Twente, Enschede, The Netherlands, 2016.
57. Bialas, J.; Oommen, T.; Rebbapragada, U.; Levin, E. Object-based classification of earthquake damage from high-resolution optical imagery using machine learning. *J. Appl. Remote Sens.* **2016**, *10*, 16. [[CrossRef](#)]
58. Kit, O.; Lüdeke, M. Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2013**, *83*, 130–137. [[CrossRef](#)]
59. Patino, J.E.; Duque, J.C. A review of regional science applications of satellite remote sensing in urban settings. *Comput. Environ. Urban Syst.* **2013**, *37*, 1–17. [[CrossRef](#)]
60. Ministry of Public Works and Housing (Kemen PUPR). *Achievement of Slums Upgrading Programs in 2017*; Ministry of Public Works and Housing: Jakarta, Indonesia, 2017; pp. 1–24.
61. Kuffer, M.; Pfeffer, K.; Sliuzas, R. Slums from space—15 years of slum mapping using remote sensing. *Remote Sens.* **2016**, *8*, 455. [[CrossRef](#)]
62. Vatsavai, R.R. High-resolution urban image classification using extended features. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 869–876.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).