ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-5/W4-2025
Philippine Geomatics Symposium (PhilGEOS) 2025 "Enhancing Human Quality of Life through Geospatial Technologies",
24–25 November 2025, Quezon City, Philippines

Dasymetric Mapping of Population Using Urban Settlement-Focused LULC Classification: A Comparative Evaluation of Random Forest, OBIA, and Hybrid OBIA-RF Methods

Ynah Andrea D. Sunga[1], Althea Marie B. Capucion[1], Gabriel Drew S. Palma[1], Julia Clarice Uy[1], Jennieveive B. Babaan-Mabaquiao[1]

[1]Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City, Metro Manila, Philippines
ydsunga@up.edu.ph, abcapucion@up.edu.ph, gspalma1@up.edu.ph, juy1@up.edu.ph, jbbabaan@up.edu.ph

Keywords: Artificial Neural Network-Multi-Layer Perceptron (ANN-MLP), informal settlement, Land Use/Land Cover (LULC), Object-Based Image Analysis (OBIA), population estimation, Random Forest Classification

Abstract

Urbanization of cities introduces growth of population in urban settlements, formal and informal. However, these informal settlements commonly go unrecorded due to legal threats, thus the scarcity of data. This study aims to bridge the scarcity of urban settlement information by dasymetric mapping of population using census blocks and land use/land cover (LULC) map, focused on formal and informal urban settlements. To achieve this, three classification methods for LULC were compared: random forest (RF), object-based image analysis (OBIA) using a Bayes classifier, and a hybrid approach combining OBIA with RF. Land classification was applied to Sentinel-2 L1C images from 2015 and 2020, with Google Earth Engine utilized for RF and QGIS Orfeo Toolbox for OBIA and OBIA-RF. To evaluate which of LULC classification methods is most accurate, the F-scores of each generated LULC map per method were compared. Furthermore, the reliability of the LULC classification methods for population dasymetric maps were compared to built surfaces of Global Human Settlement Layer through Global Similarity Value. F-score values ranging from 86.37%–94.44% for formal settlement, 65.18%–72.19% for informal settlement, and 0.4978 Global Similarity value from hybrid OBIA-RF show that it has most capability in mapping informal settlements due to incorporation of spectral, spatial and contextual characteristics. The most effective method for LULC allowed dasymetric population mapping projected for 2025, such that LULC prediction was executed using an artificial neural network-multilayer perceptron (ANN-MLP). These outputs are expected to provide valuable insights into population distribution in informal settlements, supporting urban planning and resource management efforts.

1. Introduction

Urbanization drives rapid city growth, particularly in developing countries, often resulting in informal settlements with high density, irregular layouts, and substandard housing. These areas pose unique challenges for urban planning and resource allocation (McFarlane et al., 2024). Reliable population data is essential for sustainable development. It informs infrastructure planning delivery of public services (e.g., education, healthcare), and equitable policy interventions (Asian Development Bank, 2002). It also informs land-use policies to manage urban expansion. In turn, effective planning requires a holistic approach such as integrating housing, disaster resilience, and social services alongside community participation to ensure equitable outcomes (Habitat III, 2016).

A key challenge in this process is the spatial inaccuracy of population data, especially in complex urban areas. Dasymetric mapping offers a valuable solution by refining population distribution estimates by using ancillary data like land use/land cover (LULC) to refine administrative unit data. Unlike choropleth maps that assume uniform distribution, this method creates more realistic population allocations (Kyaw, 2020) that is based on LULC changes. This method is particularly useful for cities like those in the Philippines, where census-block (e.g. barangay-level) data may not accurately capture variations in LULC and population distribution. By using LULC maps as finer zones, it enables analysis of urban growth dynamics and ecological impacts. For this accurate implementation to be achieved, reliable LULC data is necessary.

However, LULC classification presents particular challenges in identifying complex urban features containing both formal and informal settlements. Mudau and Mhangara (2021) identified difficulties in mapping informal settlements due to diverse roof materials and complex morphology. Researchers have explored various remote sensing techniques to address these challenges, including high-resolution satellite imagery, pixel-based classification, object-based image analysis (OBIA), machine learning, texture analysis, and even manual digitization (Alrasheedi et al., 2023; Mudau and Mhangara, 2021).

Among these, Random Forest (RF) and OBIA have emerged as promising methods for urban LULC classification. RF, a machine classification algorithm, effectively handles high-dimensional data and complex feature relationships for land classification (Alrasheedi et al., 2023). Its integration of spectral values and ancillary geospatial variables makes it ideal for urban mapping, as shown in studies combining multispectral imagery, digital surface models (DSM), and local knowledge (Matarira et al., 2022). Assarkhaniki et al. (2021) successfully combined Landsat 8, OSM data, and RF classification for informal settlement detection in Jakarta, highlighting RF's robustness and its ability to avoid overfitting in complex urban classifications. For these reasons, RF has become a widely used approach in urban mapping (Mudau and Mhangara, 2021).

OBIA, on the other hand, is a classification method that considers spatial relationships and contextual information (Mudau and Mhangara, 2021), allowing for a more meaningful representation of urban landscapes. It utilizes object geometry, texture, and neighborhood relationships for classification, unlike pixel-based methods that rely on spectral information alone (Mudau and Mhangara, 2021). That said, it proves to be applicable for landcover mapping, as well as urban settlement classification such that it can delineate buildings using scale, shape and

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W4-2025-401-2026

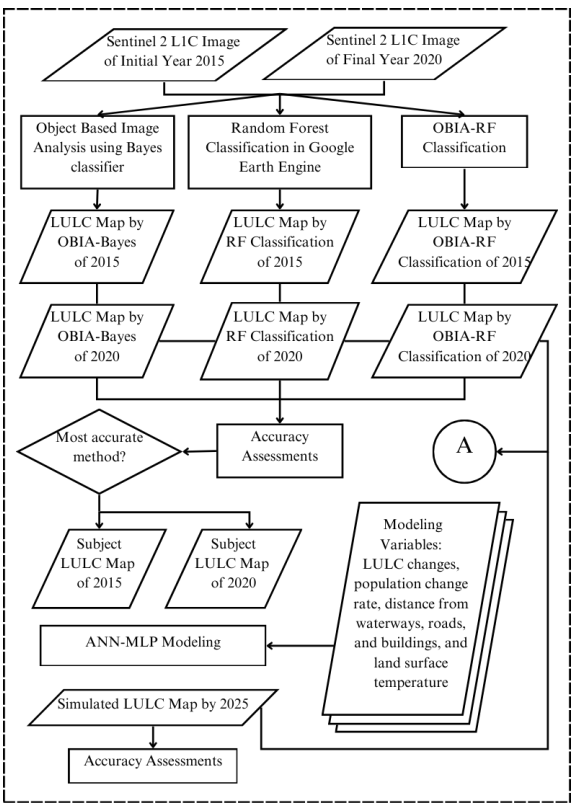
compactness (Alrasheedi et al., 2023). OBIA excels in morphological analysis but requires local typology knowledge (Mudau and Mhangara, 2021). For such unique and often irregular features of informal settlements, OBIA would depend on the understanding and selection of informal settlement local typology and image-based proxies during image classification. With this in mind, RF and OBIA techniques may be explored with respect to their applicability in urban settlement mapping, particularly for informal settlements. Both methods have merits: RF processes complex datasets efficiently, while OBIA better captures settlement morphology.

Moreover, despite the growing use of RF and OBIA in LULC classification, few studies have systematically compared these methods for urban settlement-focused dasymetric mapping, particularly in data-scarce contexts like the Philippines.

This study addressed this gap by developing a comprehensive dasymetric population map of Marikina City, Philippines, emphasizing the inclusion of urban settlements as a distinct land use and land cover category. Marikina City's mix of industrial zones and residential areas creates uneven population densities, highlighting the need for precise mapping to target planning efforts. The study evaluated the performance of three distinct LULC classification approaches: (1) RF classification, (2) OBIA utilizing a Bayes classifier, and (3) a hybrid approach integrating RF and OBIA with a Bayes classifier. By comparing the accuracy and effectiveness of these methods, this research determined the most appropriate approach, for creating a precise and reliable population dasymetric population map of Marikina City.

2. Methodology

1.1 Overview



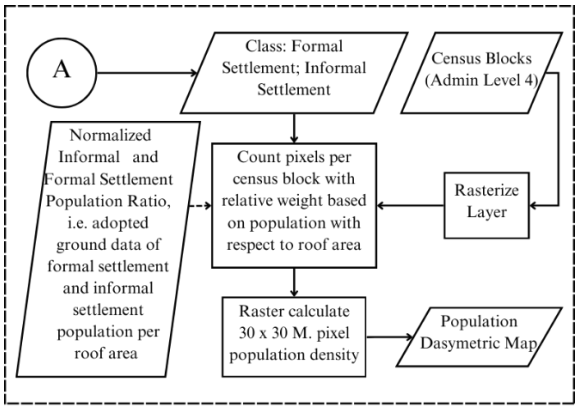


Figure 1. Methodological flowchart of the study

Figure 1 presents the key methods and datasets in achieving the objectives of this study. In summary, three LULC classification methods: OBIA, RF, OBIA-RF, were performed and compared by accuracy assessment. LULC maps with substantial accuracy were used for future LULC simulation. To apply such LULC in connection with demographic characteristics, dasymetric mapping based on the simulated LULC of 2025 was executed for geospatial population data refinement, with particular focus on urban settlement. These are explained further in subsequent sections.

2.1 Data Gathering

To generate accurate LULC classification and ancillary data to be used for dasymetric mapping of Marikina City, this study gathered diverse geospatial datasets such as satellite imagery, population, roads, rivers, and building vectors.

2.1.1 Satellite Imagery and Reference Maps

Sentinel satellite system was utilized as this allows computation of built-up-focused indices discussed in Section 3.3.3. This study also consistently used Sentinel-2 L1C satellite images of Marikina City to consider the year 2015. Moreover, the dates of the satellite images were verified to be consistent with the three LULC classification approaches done.

2.1.2 Informal Settlement Reference Maps

The reference map for determining the location of informal settlements was obtained from a World Bank report in 2015 where in-depth information about informal settlements in Metro Manila was discussed (Singh and Gadgil, 2017). This was then georeferenced and brought to the proper extents of the project. As a supplementary reference map, a high-resolution Google Satellite imagery tile was integrated into the process.

2.1.3 Modeling Variables

Modeling Variable	CRS	Spatial Resolution	Source
Distance from Roads	From WGS84 to UTM Zone 51N	30 M.	OSM
Distance from Buildings			OSM
Distance from Waterways			OSM

Land Surface Temperature			Sentinel-3
Population Change Rate			WorldPop
LULC Change			LULC map from authors

Table 1. Summary of Modeling Variables

Modeling variables employed as ancillary data in this study for simulation of 2025 LULC are distances from roads, rivers, and buildings, Land Surface Temperature (LST), population density, and area changes map of the LULC. Roads, rivers, and buildings were exported from OpenStreetMaps (OSM) and initially in vector shapefile format. LST was obtained from satellite images of the study area from the Sentinel-3 database, and gridded population density data were sourced from WorldPop, having a 100-m resolution. Moreover, LULC change was based on the LULC classification maps done by the authors in 2015 and 2020 using the determined accurate approach.

2.1.4 Population Census Blocks

Values for the population census blocks were obtained from Marikina City's ecological profile. As these were inputs for the dasymetric map in 2020 and 2025, 2020 Philippine Statistics Authority data and 2025 projected Marikeño population was utilized, based on the Geometric Growth Approach from the City Government of Marikina.

2.2 Data Processing

To prepare the gathered datasets for LULC classification and subsequent dasymetric mapping, a series of pre-processing and analytical steps were undertaken.

For the methods involving training data, two independent operators processed the three methods (i.e. one for OBIA and OBIA-RF, and one for RF). Identification of informal settlements was agreed upon to be near Marikina River, as being near waterbodies is one of the characteristics of slums (Singh, 2015). Furthermore, informal settlement topology was agreed to be those that are coarse and irregular.

2.2.1 Preparation of Modeling Variables

Road, river, and building vectors were clipped to Marikina City's boundaries, rasterized, and standardized to WGS 84. QGIS's Raster Proximity tool calculated distance rasters for each feature, then clipped to the study area. Land Surface Temperature (LST) data was provided by Sentinel-3 and was reprojected to WGS 84 in SNAP. Population change rates were derived by processing initial and final year density data through raster calculator operations.

2.2.2 Object-Based Image Analysis using Bayes

OBIA extends pixel-based classification by first segmenting images into meaningful objects. This study utilizes QGIS's Orfeo Toolbox (OTB), which provides segmentation algorithms and machine learning classifiers for multispectral image analysis. Mean-shift segmentation was applied to cluster similar reflectance values in Sentinel-2 imagery, adjusting the range radius parameter to 0.00005 (matching image reflectance units) while maintaining default values for other parameters. The segmented output was then processed with Zonal Statistics to calculate spectral metrics (mean, standard deviation, minimum, maximum) for each band across all segments, generating 13 statistical sets per segment corresponding to Sentinel-2's spectral bands. Classification was performed using the Normal Bayes classifier.

A model file for OTB was created using test samples, with all computed statistics as classification attributes and sample class identifiers as predictors. The Normal Bayes classifier was employed, assuming normality in distributed segment values (other parameters remained in default) although not strictly independent. This classifier models each class's attributes as Gaussian distributions, where values cluster near the mean with fewer outliers. Unlike Naïve Bayes, it accounts for inter-class relationships through covariance matrices (Modica et al., 2021). During classification, the algorithm evaluates how well segments fit each class's distribution and combines this with prior probabilities to predict class membership. The resulting model classified the statistical-segmented vector to produce a LULC map.

2.2.3 Random Forest Classification

The Random Forest (RF) algorithm makes use of several decision trees, where each tree classifies each pixel, based on the chosen input parameters (Svoboda et al., 2022) This algorithm then assigns the final class after taking the majority classification by all trees.

The RF method was implemented in Google Earth Engine due to its known cloud computing capabilities (Matarira et al., 2022). Data refinements done were cloud masking (Bits 10-11 for Sentinel-2 L1C) and SIAC atmospheric correction to convert raw DNs to reflectance values (0-1). For the training data, at least 40 points were included per class and merged into a single dataset. The model incorporated seven established indices namely NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), MNDWI (Modified Normalized Difference Water Index), SAVI (Soil-Adjusted Vegetation Index), NDBI (Normalized Difference Built-up Index), BRBA (Band Ratio for Built-up Area) and UI (Urban Index), which can discriminate between LULC classes (Adepoju et al., 2019). The indices appear in Equations (1) to (7), with Sentinel-2 Level-1C spectral bands B2, B3, B4, B5, B7, B8, and B11.

$$NDVI = \frac{B8-B4}{B8+B4}, \quad (1)$$

where $B8 =$ Near-infrared (NIR) band
$B4 =$ Red band

$$\text{NDWI} = \frac{B3-B8}{B3+B8}, \quad (2)$$

where $B3 =$ Green band
$B8 =$ Near-infrared (NIR) band

$$\text{MNDWI} = \frac{B3-B11}{B3+B11}, \quad (3)$$

where $B3 =$ Green band
$B11 =$ Shortwave-infrared (SWIR) band

$$\text{SAVI} = 1.5 \times \left(\frac{B8-B4}{B8+B4+0.5}\right), \quad (4)$$

where $B8$ = Near-infrared (NIR) band
$B4$ = Red band

$$\text{NDBI} = \frac{B11-B8}{B11+B8}, \quad (5)$$

where $B11$ = Shortwave-infrared (SWIR) band
$B8$ = Near-infrared (NIR) band

$$\text{BRBA} = \frac{B4}{B11}, \quad (6)$$

where $B4$ = Red band
$B11$ = Shortwave-infrared (SWIR) band

$$\text{UI} = \frac{B7-B5}{B7+B5}, \quad (7)$$

where $B7$ = Vegetation red-edge 3 band
$B5$ = Vegetation red-edge 1 band

The execution of the classification process was implemented using the *ee.Classifier.smileRandomForest* package. Using the default parameter of 100 trees and random sampling, 80% of the training samples were created, to which a decision tree was generated for each training sample. The remaining twenty percent (20%) were used to validate and test the classification accuracy of the model (Matarira et al., 2022). The resulting output of the RF model is the classified map of the study area and was exported into the same scale of the original image, i.e. 10 m with the coordinate system of WGS 84 (EPSG: 4326).

2.2.4 OBIA-RF Classification

This method combines the OTB tool for segmentation, classifier modeling and uses RF as the classifier within the same plugin. Default parameters were used; particularly the maximum number of trees (100), maximum depth (5), and maximum number of samples (10) to avoid underfitting and overfitting. The OTB algorithm for RF is OpenCV-based (open-source computer vision library) specializing in remote sensing and raster datasets.

2.2.5 Accuracy Assessment

The validation process utilized QGIS's AcATaMa plugin with stratified random sampling, employing area-proportional sample allocation (standard error target: 0.01) and 0.0001 degree (≈100m) minimum spacing. Reference data combined Google Earth imagery with a georeferenced Metro Manila informal settlements map from Singh (2015)'s OBIA-based study. The assessment generated standard accuracy metrics including overall, producer's, and user's accuracy.

Moreover, the F-score was calculated, which is a statistic that considers both the user's accuracy (UA) and producer's accuracy (PA). F-score indicates the performance of the classifier by calculating the harmonic mean of UA and PA (Zurqani et al., 2019), as shown in Equation (8).

$$F-score = 2 \times \frac{(User's\ Accuracy \times Producer's\ Accuracy)}{(User's\ Accuracy + Producer's\ Accuracy)}, \quad (8)$$

The classified map's informal settlement extents were cross-validated against manually digitized areas from VHR Google Earth imagery (2020/2025), using 2024 imagery for the 2025 map. This approach follows established validation methods for informal settlement mapping (Matarira et al., 2023), particularly valuable given limited ground truth data in Philippine slums. Accuracy was quantified through feature-similarity in Equation (9) and patch-based mean absolute percentage error (MAPE) in Equation (10).

$$Feature\ Similarity = \frac{min.area_{selected\ feature}}{max.area_{selected\ feature}}, \quad (9)$$

where "selected feature" refers to corresponding features from the classified and reference maps used for validation

$$MAPE = \left|\frac{actual\ area\ patch - forecast\ area\ patch}{actual\ area\ patch}\right|, \quad (10)$$

To determine the standard for acceptable value, 0.60 to 0.79 Cohen's Kappa was agreed to be a substantial percentage of data reliability (Parraga-Alava et al., 2021). For MAPE, Lewis (1982), as cited by Montaño (2013), standardizes that a value of less than 0.50 means reliable data forecasting. These are applied upon verification of LULC mapping and simulating.

2.2.6 LULC Map Simulation by ANN-MLP Modeling

The MOLUSCE (Modules for Land Use Change Evaluation) plugin in QGIS was employed to predict future LULC changes and maps. The plugin requires inputs of the initial state map (2015) and the final step map (2020) to analyze the succeeding changes of future maps. Modeling variables were integrated, as these factors affect the LULC changes over time. Pearson's correlation coefficient must be close to zero, denoting little to no correlation between modeling variables. This is to avoid redundancy of input layers, which may cause overfitting upon simulation. General statistics and transition matrix were also displayed in the Area Changes Tab, to quantify the change in areas of the land use and land cover types in the two selected years. This also produced the LULC change map as one of the modeling variables.

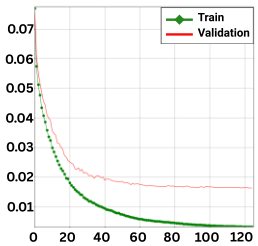


Figure 2. Transition Potential Modeling

For the transition potential modeling, Artificial Neural Network Multi-Layer Perceptron (ANN-MLP) was the chosen method to train the data, as it has been proven to produce good results with high accuracy (Souza et al., 2022) for modeling LULC changes. Parameters such as the maximum number of iterations and number of hidden layers were modified until both the learning curve and error curve appear smooth, and the Current Validation Kappa is about 0.8 or higher. For instance, DEM and slope of the study area were initially included as modeling variables. However, exclusion of DEM and slope resulted in a higher validation kappa. This may be due to the relatively even elevation of the city.

The outcome of the neural network method is the simulated LULC map for 2025. Figure 2 shows the transition potential modeling with 2015 as the initial LULC map and 2020 as the final LULC map with the aforementioned modeling variables as

input layers. The graph shows smooth, non-undulating, downwards behavior, implying diminishing marginal error and non-over-or-underfitting.

Validation of the predicted map was then compared to the closest real-world reference map of 2025, utilizing the latest 2024 Google Earth imagery of Marikina. Patch-based metrics such as MAPE and feature similarity were obtained to evaluate geometric consistency with reference data.

2.3 LULC Application

From the subject LULC maps based on the method that produced most accuracy, dasymetric mapping of population using urban settlement-focused LULC classification through RF algorithm and OBIA was done to validate effectiveness in estimating population distribution within informal settlements of the classification methods, providing evidence-based resource by spatially linking population density to settlement types. This approach does not only test the classification's real-world utility for slum density mapping but also generates actionable data to address urban planning gaps in underrepresented communities.

2.3.1 Population Estimation by Dasymetric Mapping

Dasymetric mapping was done with the 2020 LULC maps using OBIA, OBIA-RF, and RF classification, producing three comparable thematic population maps in 2020 of Marikina City. To further support which LULC classification method is most accurate, the dasymetric maps of 2020 were compared to built-surface Global Human Settlement Layer (GHSL) in 2018 with a 10-m resolution. Note that it is an open-sourced data from Copernicus that is closest to assessing the population estimates by human settlement thematic map. Hot Spot Analysis (Getis-Ord Gi*) for the 2018 GHSL and 2020 population estimation maps based on LULC using OBIA, OBIA-RF, and RF classification methods were generated. These layers were used for Hot Spot Comparison Analysis, such that the reference map is the 2018 GHSL. Thereon, Global Similarity Values were determined to distinguish which classification method used to derive a population map is most similar to the 2018 GHSL, as an accuracy assessment aside from the F-scores.

As an application of the most reliable classification method for urban settlement-focused LULC mapping, the 2025 LULC simulation and population projections were integrated to create a population estimate, dasymetric map using LULC as ancillary data. Census blocks were rasterized, while the LULC layer was reclassified by settlement type (formal/informal) based on roof-area population densities (Galeon, 2008). Using spatial analyst tools, population densities were redistributed across barangays according to LULC classifications.

3. Results and Discussion

3.1 Classified LULC Maps

Generated classified LULC maps from the OBIA, RF and OBIA-RF methods are presented in Figure 3. For OBIA-Bayes generated maps, some areas of informal settlement correspond with the 2015 Metro Manila informal settlement type distribution map. This agreement is attributed to the OBIA's leverage in identifying physical characteristics of slums such as irregular shape, texture and relations to neighboring objects. which was also used as the basis for classifying informal settlements. In this study, however, it was visually observed that not all areas of informal settlement were mapped when compared to the reference maps. This aligns with Matarira et al. (2023) where the OBIA approach failed to identify some discrete informal settlement patches due to textural complexity. This implies that varying roof materials indeed affect the identification of urban settlements, particularly informal areas.

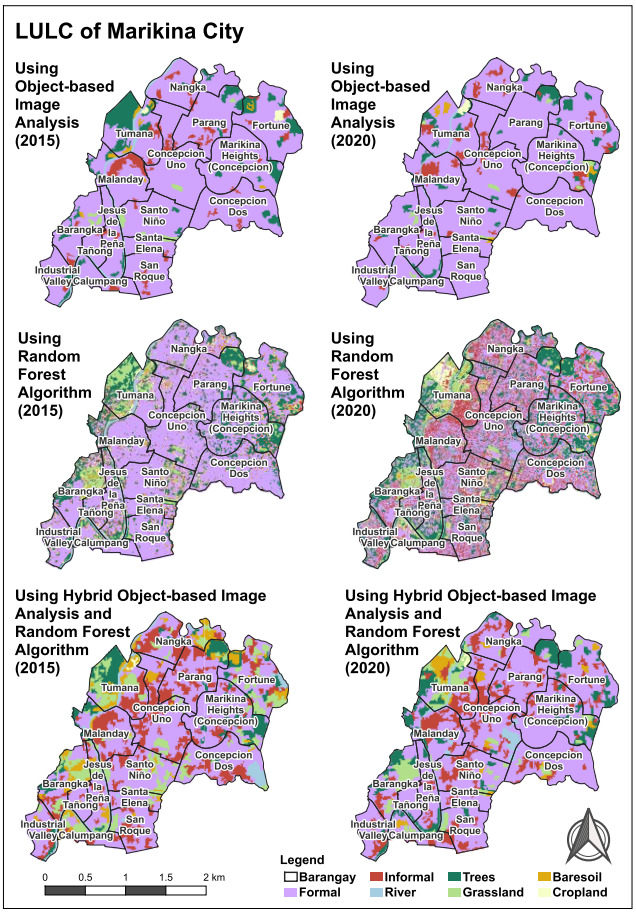


Figure 3: Classified land cover maps of Marikina City for 2015 and 2020 generated using three methods: OBIA-Bayes, RF, and OBIA-RF

For instance, the RF-classified maps show scattered informal settlements, particularly distributed in urban centers, deviating from reference data. This discrepancy likely stems from training data limited to galvanized steel roofs near Marikina River, causing misclassification of similar roofing materials common in formal structures. As a pixel-based method, RF fails to incorporate contextual neighborhood features critical for informal settlement identification. While RF proves less suitable than OBIA for Philippine informal settlements due to prevalent metal roofing, it achieves reasonable accuracy for non-settlement land use and land covers, reflecting machine learning's strength in heterogeneous classification (Alrasheedi et al., 2024).

Meanwhile, OBIA-RF maps showed high agreement with the informal settlement areas, as opposed to solely OBIA and RF. Although this hybrid approach may appear to have overclassified blocks of slums, this was attributed to interurban variability of

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-5-W4-2025-401-2026 | © Author(s) 2026. CC BY 4.0 License.

morphological slum features (Stark et al., 2020) and similarities in spectral reflectivity (Qu et al., 2021) between formal and informal areas. These overestimated classifications have its effects on dasymetric mapping thereafter, but this method was still able to map out informal settlements with higher precision than only OBIA. Findings from the accuracy assessment have shown that the OBIA-RF method consistently obtained the highest overall accuracy (OA), UA, and PA among the three methods in terms of formal and informal settlements. Accordingly, the highest F-scores for urban settlements are attained by OBIA-RF as shown in Figure 4, implying that this method is the best classifier for settlement classes among the three approaches.

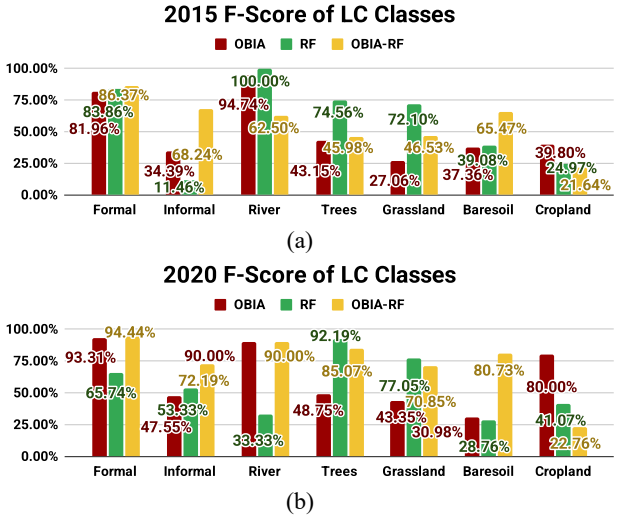


Figure 4. Comparison of F-Scores for All Classes in (a) 2015 and (b) 2020 LULC maps

These results were also observed in the accuracy assessment considering only the two settlement classes in Figure 5. OBIA and RF alternated in which method performed better for the 2015 and 2020 accuracy assessments due to the LULC changes that suited one approach over the other. Meanwhile, the hybrid OBIA-RF algorithm remained to be the best classifier. The findings align with prior research (Alrasheedi et al., 2024) demonstrating that combining OBIA and RF improves informal settlement classification accuracy. This hybrid approach leverages RF's capacity to process diverse geospatial data efficiently and OBIA's strength in object-level analysis. The neighborhood physical characteristics of slums are precisely mapped due to OBIA's high image segmentation capabilities, and its corresponding classification limitations are addressed by RF through its robustness in handling complex environments. Incorporating both methods resolves the challenges that the separate methods encounter which better captures informal settlements' morphological characteristics. Further improvements could incorporate additional indicators like textural variables and local expertise (Alrasheedi et al., 2024).

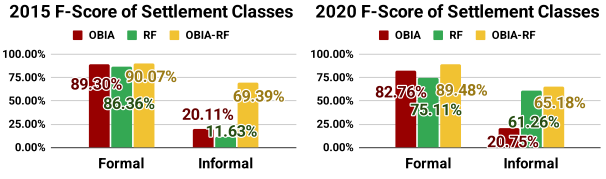


Figure 5. Comparison of F-Scores for Settlement Classes

Overall, OBIA-RF attained F-score ratings of 86.37%–94.44% for formal settlement and 65.18%–72.19% for informal settlement. Differences in the range of values were observed when compared to above 80% accuracies attained by OBIA-RF in previous studies on informal settlement mapping (Matarira et al., 2022; Qu et al., 2021). This variation may be attributed to dissimilarities in the study sites' environment, slums' physical characteristics, and data resolution. For instance, the validation for land cover classes in this study relied on publicly available reference data, which includes the informal settlement map and Google Earth imagery. The average quality of historical images, however, presents challenges in precisely identifying and delineating features of informal areas. To address this in future works, street-level imagery (Veeravalli et al., 2025) or drone images may be obtained for higher quality of validation data.

Nevertheless, the results of F-scores for the hybrid approach combining OBIA and RF LULC classification methods are further supported by patch-based MAPE and feature similarity indices of 30.32%–40.77% and 59.22%–78.03% respectively, reinforcing the robustness of the F-scores. This provides a strong justification in the selection of OBIA-RF, validating its spatial reliability. Given the study's focus on population mapping, which is closely tied to settlements, the OBIA-RF method was selected for its classification superiority. The 2015 and 2020 OBIA-RF classified maps consequently served as inputs for simulating the projected LULC map in 2025.

3.2 Dasymetric Maps of Estimated Population

Using the classification methods, thematic map of population estimates for 2020 of Marikina City were processed and compared to the open-sourced, 10-m fine resolution of urban settlement layer from Copernicus. Figure 6 shows the maps that reflect urban settlement and the population of Marikina City.

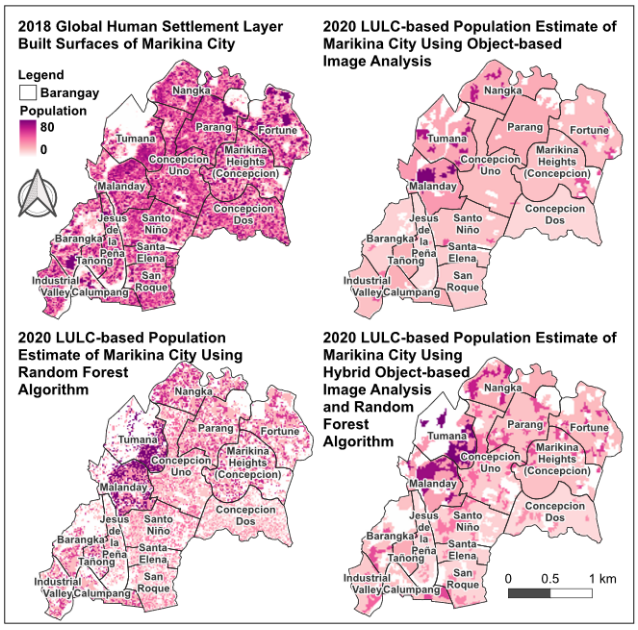


Figure 6: 2018 GHSL of Built Surfaces; 2020 LULC-based Population Estimate Maps Using OBIA, RF, and OBIA-RF classification methods

The Global Similarity Values, derived from the comparison of LULC-based population maps with the 2018 GHSL (as detailed in Section 2.4.1), were analyzed to determine the reliability of

each classification method. OBIA-RF achieved the highest Global Similarity Value of 0.4978, highlighting its suitability for urban settlement mapping that is consistent with the F-score results. On the other hand, the similarity values are 0.4497 for OBIA and 0.4691 for RF classification method. All values fall within the moderate similarity range, indicating that some areas of the LULC maps correspond to GHSL, yet some also significantly differ. These observed variations in the population estimate maps are found to be influenced by LULC misclassifications. Complexities in defining the term informal settlement induce uncertainties in the evaluator's classification (Matarira et al., 2022), considering that housing types in the Philippines could be further categorized other than formal and informal settlements (Crawford and Stephan, 2015; Galeon, 2008). Nevertheless, certain regions of the maps were found to closely align with the GHSL. This implies that while the values are moderate, they confirm that the method captures meaningful settlement patterns and produces results suitable for dasymetric population mapping in complex urban settings.

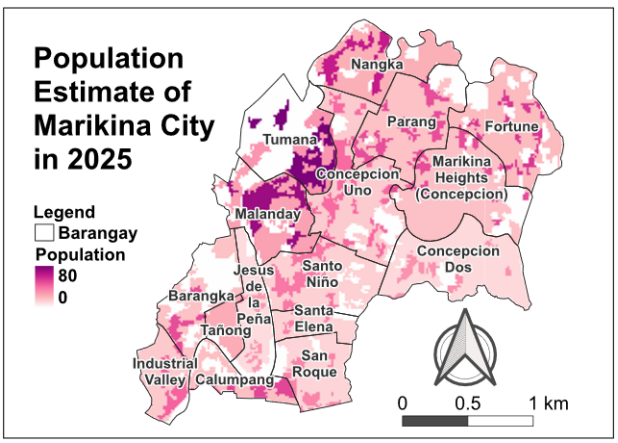


Figure 7. Dasymetric Mapping of Population Estimate of Marikina City in 2025 Using OBIA-RF LULC Map

Figure 7 displays a 30m-resolution population density map derived through dasymetric mapping, incorporating simulated LULC (based on 2015–2020 classifications), LST, area changes, and proximity to infrastructure. High-density areas correlate with informal settlements near Marikina River, which are consistent with resettlement studies (Delos Reyes and Francisco, 2014). The classified map cross-validated to 2024 VHR Google Earth and resulted in an F-score of 70.21%, 75.65% feature similarity based on area, and 33.54% MAPE. Hence, this approach indeed improves coarse barangay/block census data by distributing population according to LULC characteristics. Given challenges in surveying informal settlements due to residents' reluctance (Gupta, 2024), dasymetric mapping provides a valuable alternative for estimating populations in data-scarce areas.

4. Conclusion and Recommendations

Dasymetric mapping is a method used to refine population distribution in connecting to ancillary data such as LULC geospatial information. RF algorithm and OBIA are effective methods in generating LULC data based on spectral and morphology data. With the challenge of getting slum dweller information in line with LULC mapping of an urbanized city, this study focused on comparing three different methods for LULC classification: (1) RF classification, (2) OBIA using a Bayes classifier, and (3) a hybrid technique combining RF and OBIA with a Bayes classifier; the determined approach that is most accurate and reflective of actuality is the LULC classification used to proceed with dasymetric mapping of estimated population. The OBIA-RF algorithm effectively maps irregular informal settlements by incorporating spectral, geometric, and textural features during segmentation, proving particularly adept at slum differentiation.

Determination of an effective LULC classification method allows generation of accurate LULC maps, which may be synthesized with the area's demographic characteristic. For instance, dasymetric map of estimated population was done to show population trends alongside LULC changes, with focus on informal settlement of the urbanized city. In line with the LULC classification, the population dasymetric map reflects the geospatial distribution of the city's population, with notably high concentration in slum dwellings. This provides information regarding commonly unrecorded demographic characteristics of the city about informal settlers. The dasymetric maps produced in this study offer direct utility for Marikina City's governance challenges, particularly in disaster resilience and equitable resource allocation. For instance, the high-resolution population surfaces reveal dense informal settlements along the Marikina River, a known flood hazard zone, which enables targeted interventions, such as early warning system deployment or prioritized infrastructure upgrades. Local government units (LGUs) could integrate these maps with existing hazard maps to refine evacuation plans or zoning regulations. At the same time, NGOs might leverage them to allocate health and education services to informal communities. By aligning with the Philippines' Community-Based Disaster Risk Reduction framework (Delos Reyes and Francisco, 2014), these outputs bridge technical modeling with community and targeted planning, addressing a gap in data-driven decision-making for informal settlements.

For future work, improving Normal Bayes accuracy requires balanced training samples per class, as this directly factors into the algorithm's calculations. Enhanced a priori knowledge of settlement characteristics (topology, morphology) can better accommodate informal settlements' irregular textures. Well-defined training samples enable more reliable method comparisons. Additional approaches could explore GEE clustering algorithms combined with OBIA. High-resolution imagery remains recommended for small areas due to superior spatial, spectral, and radiometric resolution.

References

Adepoju, K., Adelabu, S., Fashae, O., 2019: Vegetation response to recent trends in climate and land use dynamics in a typical humid and dry tropical region under global change. *Advances in Meteorology*, 2019(1), 4946127.

Alrasheedi, K.G., Dewan, A., El-Mowafy, A., 2023: Mapping informal settlements using machine learning techniques, object-based image analysis and local knowledge. IGARSS 2023 – *IEEE International Geoscience and Remote Sensing Symposium*, 7249–7252. IEEE.

Alrasheedi, K.G., Dewan, A., El-Mowafy, A., 2024: Combining local knowledge with object-based machine learning techniques for extracting informal settlements from very high-resolution satellite data. *Earth Systems and Environment*, 1–16.

Asian Development Bank, 2002: Metro Manila Urban Services for the Poor Project. adb.org/sites/default/files/project-documents/31658-phi-tacr.pdf (2 December 2024).

Assarkhaniki, Z., Sabri, S., Rajabifard, A., 2021: Using open data to detect the structure and pattern of informal settlements: an outset to support inclusive SDGs' achievement. *Big Earth Data*, 5(4), 497-526. doi.org/10.1080/20964471.2021.1948178

Crawford, R.H., Stephan, A., 2015: Living and learning: research for a better built environment. 49th International Conference of the Architectural Science Association, 287–297.

Delos Reyes, M., Francisco, A., 2015: Building Sustainable and Disaster Resilient Informal Settlement Communities. *Public Policy*, 173.

Galeon, F., 2008: Estimation of population in informal settlement communities using high resolution satellite image. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XXXVII-B4, 1377–1381.

Gupta, H., 2024: Informal settlements and slum upgrading in the Philippines. The Borgen Project. borgenproject.org/informal-settlements/ (15 December 2024).

Habitat III, 2016: The Philippine National Report: A New Urban Agenda. The United Nations Conference on Housing and Sustainable Urban Development.

Kyaw, N.P., 2020: Integration of remote sensing and GIS for population estimation, Magway district, Magway region, Myanmar. sutir.sut.ac.th:8080/jspui/handle/123456789/9173 (2 December 2024).

Matarira, D., Mutanga, O., Naidu, M., 2022: Google Earth Engine for informal settlement mapping: A random forest classification using spectral and textural information. *Remote Sensing*, 14(20), 5130. doi.org/10.3390/rs14205130

Matarira, D., Mutanga, O., Naidu, M., Vizzari, M., 2022: Object-based informal settlement mapping in Google Earth Engine using the integration of Sentinel-1, Sentinel-2, and PlanetScope satellite data. *Land*, 12(1), 99. doi.org/10.3390/land12010099

McFarlane, C., Saguin, K., Cunanan, K., 2025: Density textures: the crowd, everyday life, and urban poverty in Manila. *Urban Geography*, 46(5), 1222-1241.

Modica, G., De Luca, G., Messina, G., Praticò, S., 2021: Comparison and assessment of different object-based classifications using machine learning algorithms and UAVs multispectral imagery: a case study in a citrus orchard and an onion crop. *European Journal of Remote Sensing*, 54(1), 431–460. doi.org/10.1080/22797254.2021.1951623

Moreno, J. J. M., Pol, A. P., Abad, A. S., Blasco, B. C., 2013: Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25(4), 500-506.

Mudau, N., Mhangara, P., 2021: Investigation of Informal Settlement Indicators in a Densely Populated Area Using Very High Spatial Resolution Satellite Imagery. *Sustainability*, 13(9), 4735. doi.org/10.3390/su13094735

Parraga-Alava, J., Alcivar-Cevallos, R., Vaca-Cardenas, L., Meza, J., 2021: UrbangEnCy: An emergency events dataset based on citizen sensors for monitoring urban scenarios in Ecuador. *Data in Brief*, 34, 106693.

Qu, L.A., Chen, Z., Li, M., Zhi, J., Wang, H., 2021: Accuracy improvements to pixel-based and object-based LULC classification with auxiliary datasets from Google Earth Engine. *Remote Sensing*, 13(3), 453. doi.org/10.3390/rs13030453

Reyes, M.L. (Ed.), 2004: Risk-sensitive land use planning: Towards reduced seismic disaster vulnerability; the case of Marikina City, Metro Manila, Philippines. Kassel Univ. Press.

Singh, G. (2015). From Satellites to Settlements: Identifying Slums from Outer Space within Metro Manila's Complex Urban Landscape. EOWorld2, World Bank. https://thedocs.worldbank.org/en/doc/564861506978931790-0070022017/original/NavigatingnformalityMetroManila72617.web.pdf (8 December 2024)

Singh, G., Gadgil, G., 2017: Navigating informality: Perils and prospects in Metro Manila's slums. International Bank for Reconstruction and Development / The World Bank, Washington, DC, USA.

Souza, C., Padilha, M. D. S., Arcoverde, S. N., Rafull, L. Z., 2022: Artificial neural networks to predict efficiencies in semi-mechanized bean (Phaseolus vulgaris L.) harvest. *Engenharia Agrícola*, 42, e20210097. doi.org/10.1590/1809-4430-eng.agric.v42nepe20210097/2022

Stark, T., Wurm, M., Zhu, X. X., Taubenböck, H., 2020: Satellite-based mapping of urban poverty with transfer-learned slum morphologies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5251-5263.

Svoboda, J., Štych, P., Laštovička, J., Paluba, D., Kobliuk, N., 2022: Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia. *Remote Sensing*, 14(5), 1189.

Veeravalli, S. G., Haas, J., Friesen, J., Georganos, S., 2025: Understanding Informal Settlement Transformation through Google's 2.5 D Dataset and Street View based Validation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLVIII, 245-251.

Zurqani, H. A., Post, C. J., Mikhailova, E. A., Allen, J. S., 2019: Mapping urbanization trends in a forested landscape using Google Earth Engine. *Remote Sensing in Earth Systems Sciences*, 2, 173-182. doi.org/10.1007/s41976-019-00020-y