Comparison of Supervised Classification Algorithms for Land Use Land Cover in Guayaquil: An Assessment with Landsat and MapBiomas

Divar Castro-Rodas[1,2], André Pico Molineros[3], Rahisha Thottolil[4], Dimitri Bulatov[5], Eder MS de Paula[6], Andrés Velastegui-Montoya[2,3]

[1] Faculty of Mechanical Engineering and Production Sciences, ESPOL Polytechnic University, Campus Gustavo Galindo, Km. 30.5 Vía Perimetral, Guayaquil, 090902, Ecuador – divecast@espol.edu.ec
[2] Laboratory of Geoinformation and Remote Sensing, Faculty of Engineering in Earth Sciences, ESPOL Polytechnic University, Campus Gustavo Galindo, Km. 30.5 Vía Perimetral, Guayaquil, 090902, Ecuador – dvelaste@espol.edu.ec
[3] Faculty of Engineering in Earth Sciences, ESPOL Polytechnic University, Campus Gustavo Galindo, Km. 30.5 Vía Perimetral, Guayaquil, 090902, Ecuador – aipico@espol.edu.ec
[4] Department of Urban and Regional Planning, Faculty of Engineering and the Built Environment, University of Johannesburg, Doornfontein, Johannesburg, South Africa – thottolilr@uj.ac.za
[5] Fraunhofer IOSB Ettlingen, Gutleuthausstr. 1, 76275 Ettlingen, Germany – dimitri.bulatov@iosb.fraunhofer.de
[6] Faculty of Geography, Federal University of Pará, Belém, PA, Brasil – edermileno@ufpa.br

Keywords: LULC, Landsat, Machine Learning, Supervised Classification

Abstract

The analysis of land use and cover maps enables a comprehensive understanding of the territorial transformations associated with urban growth and their impacts on ecosystems. In this study, we employed the Google Earth Engine to pre-process and mosaic the Landsat-9 images from 2023 of the surrounding area of Guayaquil city, Ecuador. Furthermore, we compared the performance of three supervised classification algorithms: Random Forest, Support Vector Machine, and Artificial Neural Network using RStudio. Four land cover classes were defined: forest, crops, no vegetation (buildings, roads and bare soil), and water. Training samples were obtained through visual interpretation and cross-verified with MapBiomas. Validation of the results was conducted using the Kappa coefficient and overall accuracy, compared against reference maps from MapBiomas. The findings indicate that the Support Vector Machine algorithm achieved the highest accuracy (Kappa = 0.91; OA = 93%), slightly surpassing Random Forest (0.89; 92%) and Artificial Neural Network (0.86; 90%). These results confirm the robustness of the SVM algorithm relative to the other methods and highlight its potential for urban monitoring in tropical environments. The methodology employed integrates open satellite data, reproducible tools, and machine learning algorithms, thereby supporting the selection of the most suitable algorithm for the conditions of Ecuador's coastal cities and contributing to sustainable territorial planning aligned with Sustainable Development Goals 11 and 13.

1. Introduction and previous work

Worldwide, the transformation of the environment caused by human activities has significantly altered ecosystems, resulting in biodiversity loss, deforestation fragmentation, and increased greenhouse gas emissions (IPBES, 2019; IPCC, 2019). This phenomenon requires continues development of for tools capable of monitoring land use and cover changes (LULCC), particularly over extensive areas such as capital cities (Song et al., 2018). Analyzing such changes facilitates organized and sustainable territorial management, including the distribution of essential services and the development of urban cadasters (Hersperger et al., 2018; UN-Habitat, 2020). Consequently, LULCC analysis has emerged as a vital evaluative instrument for tracking not only urban sprawl but also environmental degradation in adjacent regions and the erosion of ecosystem services (Yang et al., 2024).

Satellite imagery, in conjunction with remote sensing technologies, has become a source of strategic information for the analysis of LULCC, owing to its capacity for continuous observation and spatial coverage of extensive areas (Bairwa et al., 2025). Notably, the optical sensors of the Landsat series provide complementary multispectral data at a resolution of 30 meters and have maintained an uninterrupted series of data since 1972, thus facilitating the examination of terrestrial landscape evolution over the past fifty years (Wulder et al., 2019). However, the conversion of remotely sensed data into dependable thematic maps is affected by the selection of classification methodology, as various machine learning (ML) techniques may produce differing levels of accuracy or change detection efficacy when applied to identical satellite datasets. (Richardson et al., 2025). In recent years, ML models have enabled diverse classification techniques of satellite images by virtue of their capability to model complex relationships and manage large volumes of data (Southworth et al., 2024).

Among the most widely used ML algorithms are Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) (Talukdar et al., 2020). RF is based on constructing multiple decision trees and has demonstrated high accuracy and robustness against overfitting by applying heuristics (tree pruning or min leaf size), making it one of the most prevalent in the academic literature (Finger et al., 2025; Yimer et al., 2024). SVM, by contrast, maximizes the margins of separation between classes within multidimensional spaces, proving especially advantageous for problems involving nonlinear distributions (Bayable et al., 2025). And ANN, inspired by biological neural networks, facilitates the recognition of intricate nonlinear patterns. (Deng et al., 2024).

Since the beginning of 2010, ML models have been established as an alternative to traditional parametric methods for land use and land cover (LULC) classification (Khatami et al., 2016; Mountrakis et al., 2011). Numerous studies have demonstrated that supervised classification algorithms, such as RF and SVM, achieve high accuracy in generating LULC maps (Khatami et al.,

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-565-2026 |

2016; Talukdar et al., 2020). Nonetheless, their accuracy may be influenced by specific variables, including atmospheric conditions (Lu and Weng, 2007) and landscape heterogeneity within each study area (Heydari and Mountrakis, 2018).
In Europe, a study conducted in the Czech Republic (Svoboda et al., 2022) employed the RF model to classify six LULC categories, attaining an overall accuracy of 89.01% and a Kappa index of 0.84. Nonetheless, the performance of this algorithm is influenced by internal parameters such as the number of trees, the number of variables per split, and the bag fraction. In Mali, Africa, the SVM model was used for the analysis of the LULCC from Landsat satellite imagery (Dembele et al., 2025). The outcomes indicated overall accuracies ranging from 92% to 96%, with Kappa indices varying between 0.85 and 0.94. Furthermore, LULCC analyses revealed that agricultural regions experienced the most significant expansion, with a growth rate of 6.05 km² per year, whereas savannah regions exhibited declines of 3.30 and 3.00 km² annually.

In South America, a study conducted in Rio Grande do Sul, Brazil, employed the RF model in conjunction with the Compound Maximum A Posteriori (CMAP) algorithm to assess enhancements in LULC classification derived from satellite imagery (Quevedo et al., 2024). The findings demonstrated a notable increase in accuracy when utilizing the combined RF-CMAP approach, which facilitated the correct classification of a greater number of areas compared to the standalone RF model. Between the years 2000 and 2008, this integrated RF-CMAP methodology accurately identified ten additional regions relative to those classified solely with RF. In Ecuador, a multi-temporal analysis of LULC in the Southern Andes, was undertaken from 1989 to 2016 using supervised classification. The results indicated overall accuracy levels ranging from 87% to 89% across the evaluated periods, demonstrating that classes with larger territorial extent achieved higher spectral separability, while smaller, heterogeneous classes presented greater classification challenges (López et al., 2020).

Despite advances in technology and scientific methods concerning supervised classification techniques (Castro-Rodas et al., 2026; Freire-Granda et al., 2026; Phiri and Morgenroth, 2017), a notable gap persists in the benchmarking of ML algorithms within tropical environments characterized by limitations such as cloud cover, which impedes the classification of the selected imagery (Whitcraft et al., 2015). In Ecuador, the proliferation of open platforms, such as Google Earth Engine (GEE) (Gorelick et al., 2017), together with collaborative initiatives like MapBiomas, a platform dedicated to mapping and monitoring historical LULCC across the diverse biomes in the Amazon countries using cloud computing (Fundación EcoCiencia, 2024; Souza et al., 2020), has strengthened the capacity to monitor optical satellite variables and generate validated LULC datasets at national and regional scales (Escandón-Panchana et al., 2025). Nevertheless, the challenge of identifying algorithms that deliver superior performance in regions beset by both geographical constraints (e.g., cloudiness, scale, periodicity) (Uday et al., 2025) and operational limitations (e.g., software, processing capacity, licenses) (Dritsas and Trigka, 2025) presents an opportunity for further research, which is the focus of this study.

The city of Guayaquil, Ecuador, responds to the place with accelerated urban dynamics located in a tropical zone with the geographical and operational problems. Unplanned growth has led to the conversion of agricultural areas and mangrove ecosystem into urban and peri-urban areas, modifying the local ecological balance (Velástegui-Montoya et al., 2025). The present study aims to compare the performance of the RF, SVM and ANN algorithms applied to a mosaic of Landsat-9 images from 2023, in order to determine which, one offers greater accuracy for the LULC classification in the Guayaquil area, compared to the information from MapBiomas. The results aim to provide measurable, quantifiable evidence on the effectiveness of ML algorithms in tropical regions, contributing to the strengthening of reproducible methodologies based on open data. Likewise, the maps aim to provide inputs for territorial planning and environmental management in Latin American cities facing similar challenges.

2. Materials and Methods

2.1 Area of Interest

Figure 1 illustrates the geographic location of the area of interest (AOI), which includes Greater Guayaquil. (Hidalgo-Crespo et al., 2023), one of the most populous regions in Ecuador, due to its economic, demographic, and environmental significance. Its central coordinates located at 2°12′00"S and 79°58′00"W and the extent cover approximately 545.9 km², including the municipalities of Guayaquil, Durán, Samborondón, and Daule. According to the National Institute of Statistics and Census, the estimated population exceeds 3.37 million inhabitants (INEC, 2024).

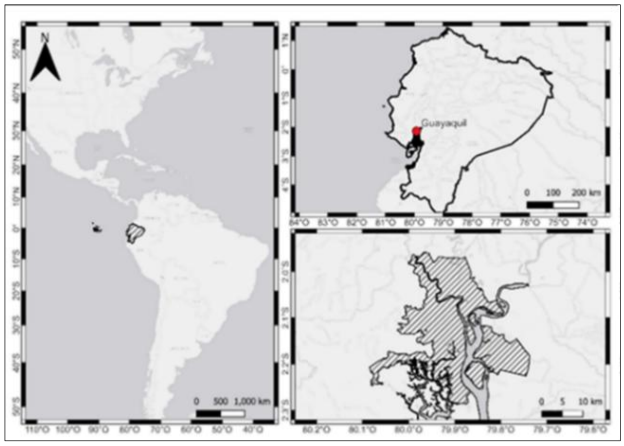

Figure 1. Geographical location map of the Greater Guayaquil metropolitan region in southwestern Ecuador.

The climate of the region corresponds to the tropical savannah (Aw) category as per the Köppen classification (Beck et al., 2023), characterized by an average annual temperature ranging from 25 to 28 °C and receiving annual rainfall between 800 and 2300 mm (INAMHI, 2024). This climate is influenced by the interannual variability of the El Niño-Southern Oscillation (ENSO) phenomenon. The topography is predominantly flat and includes mangrove ecosystems, agricultural zones, urbanized areas, and coastal water bodies (Delgado, 2013). The pronounced landscape heterogeneity renders Greater Guayaquil a representative and methodologically challenging scenario for evaluating the performance of supervised classification algorithms under complex spectral and spatial conditions (Shang et al., 2018).

2.2 Methodological outline

Figure 2 presents the overall framework of the methodological approach employed in this study. The process has been executed in five phases: I) acquisition and preprocessing of Landsat-9 images, including cloud filtering and mosaic generation; II) selection of training and validation samples to ensure each class's representativeness; III) implementation of supervised classification algorithms (RF, SVM, ANN); IV) evaluation of performance utilizing statistical metrics and confusion matrix; and V) external validation with MapBiomas Ecuador and a comparative analysis of accuracy.

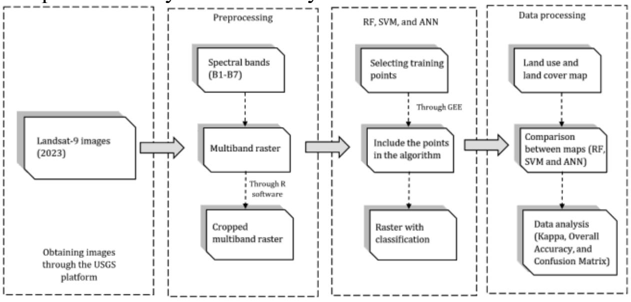

Figure 2. Workflow of LULC classification using Landsat-9 imagery (2023), including image acquisition, preprocessing, machine-learning classification, and accuracy assessment through comparative analysis and validation metrics.

2.3 Procurement and Pre-Processing

Optical images from the Landsat-9 mission, corresponding to the year 2023, were utilized. The mosaic of images was produced via the GEE platform utilizing the data repository "USGS Landsat 9 Level 2, Collection 2, Tier 1". Furthermore, to mitigate cloud noise, pixels with cloud cover exceeding 30% were excluded. The B2–B7 bands (visible, near-infrared, and two short-wave infrared channels) were incorporated and delineated according to the AOI. The optical image was generated using GEE and subsequently exported in GeoTIFF format for analysis in RStudio (R Core Team, 2023).

2.4 Training and validation

Four LULC classes were delineated for classification: Forest, Crops, Impervious (Surface), and Water. The training samples were obtained by visual interpretation of the Landsat-9 mosaic and high-resolution Google Earth photographs within the QGIS basemap. A total of 200 points per class were selected to ensure a homogeneous distribution across the study area, surpassing the minimum sample size of 50–100 samples recommended in literature (Congalton and Green, 2019). The dataset was partitioned into 70% for training and 30% for validation. All samples were exported in .csv format and subsequently processed in RStudio.

2.5 Supervised classification algorithms

2.5.1 Random Forest (RF)

This classifier is constructed upon a framework consisting of multiple decision trees, and each tree contributes a single vote for the most probable class of the input feature vector (Breiman, 2001). A decision tree performs multiple, sequential splits of single features (in our case: image channels) in order to obtain distribution of training data as pure as possible on both sides of the split. In applications involving satellite image classification, RF has been highly preferred due to its robust performance with high-dimensional datasets, its capacity to manage outliers, and its superior accuracy relative to alternative ML models (Belgiu and Drăguţ, 2016).

The model was implemented using the *randomForest()* function within the RStudio environment. The number of trees of 500 was chosen to ensure the stabilization of the classification error, as prior research indicates that accuracy converges and the out-of-bag (OOB) error is minimized around this threshold without incurring excessive computational expenses. (Belgiu and Drăguţ, 2016). Concerning the number of variables randomly sampled as candidates at each split, a value of 2 was employed. This choice adhered to the standard for classification tasks, defined as the square root of the total number of predictors ($\sqrt{p}$, where p=6 spectral bands), yielding an approximate value of 2 (Breiman, 2001).

2.5.2 Support Vector Machine (SVM)

Conversely, the SVM classifier is an ML model that uses kernel functions to fit a hyper-surface separating the training data and is applicable to both classification and regression tasks. It is distinguished by its high generalization capability and discriminative power, which have enabled its application in pattern recognition and binary classification (Cervantes et al., 2020; Dehghananari and Hosseinjanizadeh, 2024). SVM is categorized as a non-parametric supervised learning technique, demonstrating insensitivity to the distribution of the underlying data (Mountrakis et al., 2011). For this study, a radial basis function (RBF) kernel was employed using the *caret* package (Kuhn, 2008).

2.5.3 Artificial Neural Networks (ANN)

The ANN is inspired by the functioning of the human brain, roughly modelling its processes through a massively distributed, parallel processing system composed of simple calculation units (Goodfellow et al., 2016). These models acquire knowledge from their environment through training variables and, owing to their non-linear properties, are capable of processing incomplete data and producing responses even in situations characterized by uncertainty (Atkinson and Tatnall, 1997). Consequently, ANNs have been extensively applied to supervised LULC classification, as they allow for the robust estimation of class distributions from training samples (Mas and Flores, 2008). The fundamental mechanism of these networks lies in the iterative adjustment of connection weights between neurons and the dynamics of their interactions during the learning process (Bishop, 2006).

For training, the *nnet* package was employed (Venables and Ripley, 2010), using a feed-forward architecture with a single hidden layer of 5 neurons. To enhance model generalization and prevent overfitting, a regularization parameter (weight decay) of 0.1 was applied (Krogh and Hertz, 1991). The calibration process incorporated a five-fold cross-validation scheme. The model was trained using 560 samples and six predictors corresponding to the selected spectral bands.

2.6 Evaluation Metrics

2.6.1 Overall Accuracy

The overall accuracy (OA) is the default metric to indicate the overall percentage of correctly classified pixels relative to the total number of observations. Technically, it is calculated by dividing the sum of the diagonal elements of the confusion matrix (correctly classified instances) by the total number of samples (Congalton, 1991).

2.6.2 Kappa coefficient

In studies aimed at generating LULC maps, the Kappa coefficient remains a standard statistical measure of accuracy. It facilitates quantifying the degree of concordance between classified and reference data even in the case of strongly imbalanced data, explicitly correcting for agreements that may occur by random chance, a practice widely endorsed in recent machine learning reviews (Talukdar et al., 2020). This metric is fundamental for benchmarking algorithms, as it enables comparison of inter-rater agreement across models such as RF and SVM within a standardized statistical framework (Tilahun, 2015).

The values of Kappa below 0.40 denote low agreement; values between 0.41 and 0.60 indicate moderate agreement; those between 0.61 and 0.80 signify substantial agreement; and values exceeding 0.80 represent near-perfect agreement (Landis and Koch, 1977).

2.6.3 External validation

A comparative evaluation was conducted between the classification maps produced by each algorithm and the reference data from the MapBiomas Ecuador project (Fundación EcoCiencia, 2024). This comparison facilitated the assessment of the spatial congruence of the classification patterns and the accuracy in delineating the classes. In addition to standard accuracy metrics, a quantitative comparison of the areal estimates was performed. The pixel percentages for each class (Forest, Crops, Non-Vegetated Zones, Water) in the LULC maps generated by each algorithm were compared with the corresponding percentages in the MapBiomas reference map. This metric complemented the spatial validation by providing a detailed analysis of quantity disagreement versus spatial allocation, offering a comprehensive view of land cover distribution (Pontius Jr and Millones, 2011).

3. Results

The performance evaluation for the three algorithms is summarized in Table 1.

Table 1. Performance comparison of supervised classification algorithms, with all values given in percent and best results marked in bold.

Algorithm	Kappa Index	Accuracy [%]	Computation Time [s]
RF	89	92	0.13
SVM	**91**	**93**	**0.53**
ANN	86	90	0.41

The SVM algorithm attained the highest overall performance, evidenced by a kappa coefficient of 91% and an accuracy rate of 93%, surpassing the RF and ANN algorithms. Furthermore, an analysis of the LULC class distribution in terms of pixel percentages was conducted, comparing the results of each algorithm with the MapBiomas reference map (Table 2), alongside the supervised classification maps produced by each algorithm (Figure 3).

Table 2. Comparison of the pixel distribution by LULC class.

LULC	RF [%]	SVM [%]	ANN [%]	MapBiomas [%]
Crop	40.6	41.7	34.9	44.2
Forest	24.6	25.1	26.6	20.8
No Vegetation	23.1	21.6	23.2	20.7
Water	11.8	11.6	15.3	14.2

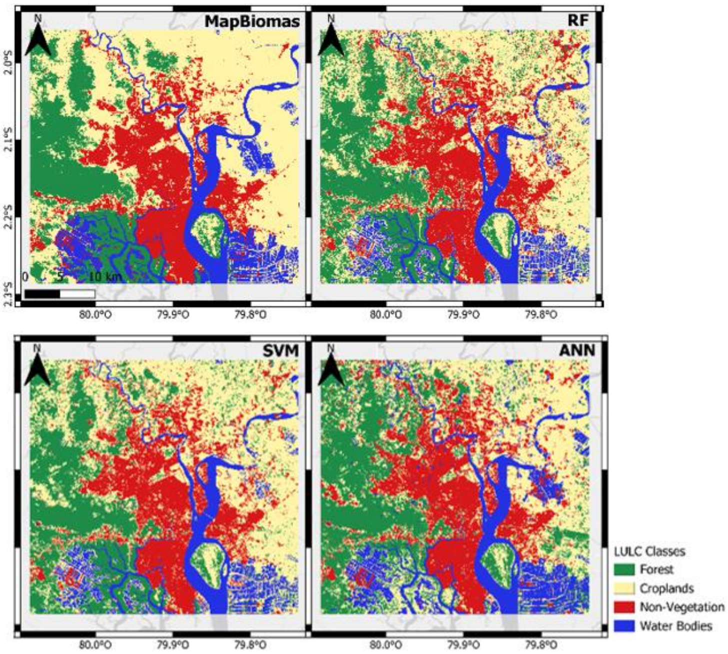

Figure 3. Supervised land-use/land-cover classification maps generated using different machine-learning algorithms, illustrating spatial variations and comparative classification performance.

A comparison of the percentage of pixels across various classes with the reference values provided by MapBiomas revealed that SVM exhibits greater similarity in the crop and no vegetation classes and a minimal correlation with water coverage. RF demonstrated the closest approximate values within the Forest category and ranked second in the crop, no vegetation, and water categories. Conversely, ANN was most prominent in accurately identifying the water class but showed the lowest performance in the remaining categories.

4. Discussion

Overall, SVM demonstrated superior performance, achieving an overall accuracy of 93% and a Kappa coefficient of 0.91. It was followed by RF, which achieved 92% accuracy (k = 0.89), and ANN, which achieved 90% accuracy (k = 0.86). This trend is consistent with comprehensive meta-analyses and similar studies conducted in tropical and subtropical environments, where SVM has frequently attained accuracies exceeding 90%, demonstrating high robustness in delineating complex land cover classes from multispectral data (Khatami et al., 2016; Thanh Noi and Kappas, 2017).

Regarding the RF algorithm, although it is widely acknowledged for its stability and capacity to handle noise and overfitting (Belgiu and Drăguţ, 2016), it exhibited a slightly worse performance. This can be attributed to the high spectral correlation between vegetation and crop classes, which diminished the decision trees' discrimination ability in transition zones or cover mixtures, a documented challenge in heterogeneous landscapes (Maxwell et al., 2018). Furthermore, the configuration of the internal model parameters, such as the number of trees and variables per split, directly affects the predictive classifier performance and its sensitivity to feature redundancy (Rodriguez-Galiano et al., 2012). Nevertheless, RF demonstrated robust, consistent behavior, proving effective at identifying forest regions and water bodies.

Conversely, the ANN algorithm demonstrated commendable performance, albeit inferior to that of the SVM and RF models. This discrepancy can be attributed to the simplistic architecture of the neural network employed, consisting of a single hidden layer with five neurons. Although this configuration mitigates the risk of overfitting, it also constrains the capacity of the model to encode intricate interactions among variables, often leading to underfitting in complex transition zones (Shao and Lunetta, 2012). Nevertheless, the ANN exhibited exemplary performance within the Water class, attaining sensitivities exceeding 95%, thereby illustrating its potential to differentiate coverages characterized by well-defined spectral signatures (Kavzoglu and Mather, 2003).

External validation using the MapBiomas dataset facilitated the evaluation of the spatial consistency of the results, demonstrating a high degree of concordance between the classified maps and the reference database. The SVM classifier demonstrated the highest similarity with MapBiomas, with an average agreement exceeding 90%, followed by RF at 88% and ANN at 86%. The main discrepancies were observed in the urban and agricultural boundaries of the northern and northwestern regions of Guayaquil. These areas are characterized by rapid land use changes and the presence of spectral mixing and shadows, which considerably complicate accurate classification. This aligns with established evidence from large-scale mapping initiatives (Souza et al., 2020) and recent environmental monitoring studies (Yang et al., 2022), which confirm the effectiveness of machine learning algorithms on GEE for producing reproducible thematic maps even within such complex heterogeneous landscapes.

The performance of the models was primarily conditioned by three critical factors. First, the spectral heterogeneity of the landscape, characterized by a complex mosaic of mangroves, agricultural zones, and urban settlements, challenged the discriminatory capacity of the algorithms. In this context, SVM demonstrated a bit superior efficacy in comparison to two other classifiers considered in this study. Second, the size of the training dataset directly influenced model stability. While current protocols suggest 50-100 samples per class to ensure a balanced distribution, empirical evidence indicates that increasing the sample size, particularly for spectrally diverse classes, significantly enhances model generalization and reduces error variance (Heydari and Mountrakis, 2018). Third, the 30-meter spatial resolution of Landsat-9 constrained the ability to resolve fine-scale urban structures and mixed pixels. Consequently, future analyses would benefit from integrating higher-resolution data from sensors such as Sentinel-2 (10 m) or PlanetScope (3 m) to mitigate the mixed-pixel effect in heterogeneous urban fringes (Phiri et al., 2020).

Although these results are promising, the study recognizes inherent limitations stemming from its reliance on optical imagery, which is vulnerable to data loss during periods of dense cloud cover, a perennial issue in equatorial regions. To address this, future research would greatly benefit from the combined utilization of Synthetic Aperture Radar (SAR) data (e.g., Sentinel-1), which provides reliable observation capabilities irrespective of atmospheric conditions. Additionally, concerning the ANN model, transitioning to Deep Learning (DL) architectures, such as Convolutional Neural Networks (CNNs) or hybrid models, has the potential to significantly enhance predictive accuracy. This shift is supported by recent findings in satellite image classification studies (Cecili et al., 2023; Li et al., 2018), which highlight the superior ability of DL techniques to extract complex spatial features within heterogeneous landscapes.

The proposed methodology, utilizing open-access satellite data, cloud computing platforms, and reproducible code, enables the generation of information without the need for extensive computational infrastructure, thereby effectively democratizing environmental monitoring (Tamiminia et al., 2020). Its application in Greater Guayaquil exemplifies the potential of SVM-based approaches to offer actionable insights for decision-making related to urban expansion and environmental impact assessments. On a broader level, this research directly supports the Sustainable Development Goals, particularly SDG 11 (Sustainable Cities and Communities) and SDG 13 (Climate Action) by providing the scientific baseline necessary for advancing urban resilience and sustainable territorial planning, as advocated by the international Earth Observation agenda (Anderson et al., 2017).

5. Conclusions

This study demonstrated that the SVM classifier yielded the highest accuracy within the Greater Guayaquil area, outperforming both RF and ANN. This finding highlights SVM's robust capacity to model non-linear relationships and precisely differentiate spectrally complex classes, an advantage in heterogeneous urban landscapes.

External validation against the MapBiomas Ecuador dataset revealed spatial concordance exceeding 90%, thereby affirming the reliability of the proposed methodological framework, grounded in open data and reproducible techniques. These results confirm that integrating Landsat-9 imagery with supervised classification algorithms provides a cost-effective, scalable solution for environmental monitoring, particularly in regions with limited geospatial data availability.

Beyond the local context, this research establishes a replicable workflow for deploying ML models in further urban centers experiencing rapid urbanization. The generated data offers a validated baseline to support decision-making focused on sustainability, directly contributing to SDGs 11 (Sustainable Cities and Communities) and 13 (Climate Action). Future investigations should prioritize integrating high-resolution imagery and SAR data to overcome atmospheric limitations, alongside exploring DL architectures to enhance temporal change detection and delineate complex urban features.

6. References

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., Friedl, L., 2017. Earth observation in service of the 2030 Agenda for Sustainable Development. Geo-spatial Information Science 20, 77–96. https://doi.org/10.1080/10095020.2017.1333230

Atkinson, P.M., Tatnall, A.R.L., 1997. Introduction Neural networks in remote sensing. International Journal of Remote Sensing 18, 699–709. https://doi.org/10.1080/014311697218700

Bairwa, B., Sharma, R., Kundu, A., Sammen, S.Sh., Alshehri, F., Pande, C.B., Orban, Z., Salem, A., 2025. Predicting changes in land use and land cover using remote sensing and land change modeler. Front. Environ. Sci. 13, 1540140. https://doi.org/10.3389/fenvs.2025.1540140

Bayable, G., Gebrie, G., Melese, T., Melaku, A., 2025. Land use/cover classification using machine learning algorithms and their impacts on land surface temperature and soil moisture in the Alawuha Watershed, Ethiopia. Environmental and Sustainability Indicators 27, 100797. https://doi.org/10.1016/j.indic.2025.100797

Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., Van Dijk, A.I.J.M., Miralles, D.G., 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. Sci Data 10, 724. https://doi.org/10.1038/s41597-023-02549-6

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bishop, C.M., 2006. Pattern recognition and machine learning, Information science and statistics. Springer, New York.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Castro-Rodas, D., Velastegui-Montoya, A., Guimarães, Y.C., Nascimento, E.S., Pereira, M.D.A., Santos, L.B.L., Alvarez, C., Sanclemente Ordoñez, E., 2026. Detecting Urban Change with Open-Source GIS programs: A Case Study in Guayaquil. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-3/W4-2025, 49–54. https://doi.org/10.5194/isprs-annals-X-3-W4-2025-49-2026

Cecili, G., De Fioravante, P., Dichicco, P., Congedo, L., Marchetti, M., Munafò, M., 2023. Land Cover Mapping with Convolutional Neural Networks Using Sentinel-2 Images: Case Study of Rome. Land 12. https://doi.org/10.3390/land12040879

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Congalton, R.G., Green, K., 2019. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third Edition, 3rd ed. CRC Press. https://doi.org/10.1201/9780429052729

Dehghananari, M., Hosseinjanizadeh, M., 2024. Investigation of the Support Vector Machine (SVM) algorithm for land use changes, a case study of Kerman Province, Iran. Journal of Geological Remote Sensing 2, 1–14. https://doi.org/10.48306/jgrs.2024.467600.1009

Delgado, A., 2013. Guayaquil. Cities 31, 515–532. https://doi.org/10.1016/j.cities.2011.11.001

Dembele, B.A., Traore, S.S., Omobowale, M., 2025. Analysis and monitoring of land use and land cover (LULC) changes from 1990 to 2023 in Faragouaran Municipality, Southern Mali. GeoJournal 90, 203. https://doi.org/10.1007/s10708-025-11451-0

Deng, Y., Zhang, Y., Pan, D., Yang, S.X., Gharabaghi, B., 2024. Review of Recent Advances in Remote Sensing and Machine Learning Methods for Lake Water Quality Management. Remote Sensing 16, 4196. https://doi.org/10.3390/rs16224196

Dritsas, E., Trigka, M., 2025. Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. Remote Sensing 17, 550. https://doi.org/10.3390/rs17030550

Escandón-Panchana, P., Velastegui-Montoya, A., Pico-Saltos, R., Martínez-Cuevas, S., 2025. Applications of geomatics in multidisciplinary knowledge fields. A review for decision-makers. Discov Appl Sci 7, 1409. https://doi.org/10.1007/s42452-025-07928-9

Finger, A.P., Ferreira, R.L.C., Lana, M.D., Silva, J.A.A.D., Silva, E.A., Breunig, F.M., Bispo, P.D.C., Liesenberg, V., Nogueira, S.S., 2025. Comparison Between Traditional Forest Inventory and Remote Sensing with Random Forest for Estimating the Periodic Annual Increment in a Dry Tropical Forest. Forests 16, 998. https://doi.org/10.3390/f16060998

Freire-Granda, D., Velastegui-Montoya, A., Guimarães, Y.C., Nascimento, E.S., Pereira, M.D.A., Santos, L.B.L., Alvarez, C.I., Sanclemente, E., 2026. From Constraints to Urban Growth: Satellite-Based Monitoring of Guayaquil’s UrbanExpansion with Sentinel-2. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. X-3/W4-2025, 149–155. https://doi.org/10.5194/isprs-annals-X-3-W4-2025-149-2026

Fundación EcoCiencia, 2024. MapBiomas Ecuador: Colección 2.0 de Mapas Anuales de Cobertura y Uso del Suelo.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hersperger, A.M., Oliveira, E., Pagliarin, S., Palka, G., Verburg, P., Bolliger, J., Grădinaru, S., 2018. Urban land-use change: The role of strategic spatial planning. Global Environmental Change 51, 32–42. https://doi.org/10.1016/j.gloenvcha.2018.05.001

Heydari, S.S., Mountrakis, G., 2018. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26

Landsat sites. Remote Sensing of Environment 204, 648–658. https://doi.org/10.1016/j.rse.2017.09.035

Hidalgo-Crespo, J., Velastegui-Montoya, A., Zwolinski, P., Riel, A., Amaya-Rivas, J.L., 2023. Formalization of recyclable waste transfer stations within the Grand Guayaquil. Procedia CIRP 116, 456–461. https://doi.org/10.1016/j.procir.2023.02.077

INAMHI, 2024. Climatological and Meteorological Data for Ecuador.

INEC, 2024. Estructura poblacional – Censo de Población y Vivienda 2022.

IPBES, 2019. Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES Secretariat, Bonn, Germany.

IPCC, 2019. Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Kavzoglu, T., Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. International Journal of Remote Sensing 24, 4907–4938. https://doi.org/10.1080/0143116031000114851

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote Sensing of Environment 177, 89–100. https://doi.org/10.1016/j.rse.2016.02.028

Krogh, A., Hertz, J., 1991. A Simple Weight Decay Can Improve Generalization, in: Moody, J., Hanson, S., Lippmann, R.P. (Eds.), Advances in Neural Information Processing Systems. Morgan-Kaufmann.

Kuhn, M., 2008. Building Predictive Models in *R* Using the **caret** Package. J. Stat. Soft. 28. https://doi.org/10.18637/jss.v028.i05

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159. https://doi.org/10.2307/2529310

Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q., 2018. Deep learning for remote sensing image classification: A survey. WIREs Data Min & Knowl 8, e1264. https://doi.org/10.1002/widm.1264

López, S., López-Sandoval, M.F., Gerique, A., Salazar, J., 2020. Landscape change in Southern Ecuador: An indicator-based and multi-temporal evaluation of land use and land cover in a mixed-use protected area. Ecological Indicators 115, 106357. https://doi.org/10.1016/j.ecolind.2020.106357

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28, 823–870. https://doi.org/10.1080/01431160600746456

Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. International Journal of Remote Sensing 29, 617–663. https://doi.org/10.1080/01431160701352154

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Phiri, D., Morgenroth, J., 2017. Developments in Landsat Land Cover Classification Methods: A Review. Remote Sensing 9, 967. https://doi.org/10.3390/rs9090967

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M., 2020. Sentinel-2 Data for Land Cover/Use Mapping: A Review. Remote Sensing 12. https://doi.org/10.3390/rs12142291

Pontius Jr, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing 32, 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Quevedo, R.P., Maciel, D.A., Reis, M.S., Rennó, C.D., Dutra, L.V., Andrades-Filho, C.D.O., Velástegui-Montoya, A., Zhang, T., Körting, T.S., Anderson, L.O., 2024. Land use and land cover changes without invalid transitions: A case study in a landslide-affected area. Remote Sensing Applications: Society and Environment 36, 101314. https://doi.org/10.1016/j.rsase.2024.101314

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, G., Knudby, A., Crowley, M.A., Sawada, M., Chen, W., 2025. Machine learning approaches to Landsat change detection analysis. Canadian Journal of Remote Sensing 51, 2448169. https://doi.org/10.1080/07038992.2024.2448169

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67, 93–104. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2011.11.002

Shang, M., Wang, S.-X., Zhou, Y., Du, C., 2018. Effects of Training Samples and Classifiers on Classification of Landsat-8 Imagery. Journal of the Indian Society of Remote Sensing 46, 1333–1340. https://doi.org/10.1007/s12524-018-0777-z

Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing 70, 78–87. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2012.04.001

Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature 560, 639–643. https://doi.org/10.1038/s41586-018-0411-9

Southworth, J., Smith, A.C., Safaei, M., Rahaman, M., Alruzuq, A., Tefera, B.B., Muir, C.S., Herrero, H.V., 2024. Machine learning versus deep learning in land system science: a decision-making framework for effective land classification. Front. Remote Sens. 5, 1374862. https://doi.org/10.3389/frsen.2024.1374862

Souza, C.M., Z. Shimbo, J., Rosa, M.R., Parente, L.L., A. Alencar, A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P.W.M., De Oliveira, S.W., Rocha, W.F., Fonseca, A.V., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C., Siqueira, J.V., Viera, J.L., Neto, L.C.F., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V.V., Azevedo, T., 2020. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. Remote Sensing 12, 2735. https://doi.org/10.3390/rs12172735

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, 1189. https://doi.org/10.3390/rs14051189

Talukdar, S., Singha, P., Mahato, S., Shahfahad, Pal, S., Liou, Y.-A., Rahman, A., 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sensing 12, 1135. https://doi.org/10.3390/rs12071135

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS Journal of Photogrammetry and Remote Sensing 164, 152–170. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2020.04.001

Thanh Noi, P., Kappas, M., 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 18, 18. https://doi.org/10.3390/s18010018

Tilahun, A., 2015. Accuracy Assessment of Land Use Land Cover Classification using Google Earth. AJEP 4, 193. https://doi.org/10.11648/j.ajep.20150404.14

Uday, G., Purse, B.V., Kelley, D.I., Vanak, A., Samrat, A., Chaudhary, A., Rahman, M., Gerard, F.F., 2025. Radar versus optical: The impact of cloud cover when mapping seasonal surface water for health applications in monsoon-affected India. PLoS ONE 20, e0314033. https://doi.org/10.1371/journal.pone.0314033

UN-Habitat, 2020. World Cities Report 2020: The Value of Sustainable Urbanization. United Nations Human Settlements Programme, Nairobi, Kenya.

Velástegui-Montoya, A., Brito-Matamoros, R., Jaya-Montalvo, M., Caiza-Morales, L., Adami, M., Morante-Carballo, F., Carrión-Mero, P., 2025. Land use change and mangrove conservation strategies in the Gulf of Guayaquil Ecuador through spatial and delphi based analysis. Discov Appl Sci 7, 1121. https://doi.org/10.1007/s42452-025-07761-0

Venables, W.N., Ripley, B.D., 2010. Modern Applied Statistics with S, 4. ed., [Nachdr.]. ed, Statistics and computing. Springer, New York.

Whitcraft, A.K., Vermote, E.F., Becker-Reshef, I., Justice, C.O., 2015. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. Remote Sensing of Environment 156, 438–447. https://doi.org/10.1016/j.rse.2014.10.009

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment 225, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Yang, H., Wang, Y., Tu, P., Zhong, Y., Huang, C., Pan, X., Xu, K., Hong, S., 2024. Evaluating the effects of future urban expansion on ecosystem services in the Yangtze River Delta urban agglomeration under the shared socioeconomic pathways. Ecological Indicators 160, 111831. https://doi.org/10.1016/j.ecolind.2024.111831

Yang, R., Chen, H., Chen, S., Ye, Y., 2022. Spatiotemporal evolution and prediction of land use/land cover changes and ecosystem service variation in the Yellow River Basin, China. Ecological Indicators 145, 109579. https://doi.org/https://doi.org/10.1016/j.ecolind.2022.109579

Yimer, S.M., Bouanani, A., Kumar, N., Tischbein, B., Borgemeister, C., 2024. Comparison of different machine-learning algorithms for land use land cover mapping in a heterogenous landscape over the Eastern Nile river basin, Ethiopia. Advances in Space Research 74, 2180–2199. https://doi.org/10.1016/j.asr.2024.06.010