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

Machine Learning-Based Deforestation Monitoring and Forecasting Using Sentinel-2 Imagery: A Case Study in Narra, Palawan

Jemar E. Laag[1], Karl Ezra S. Pilario[1]

[1] Artificial Intelligence Program, National Graduate School of Engineering, University of the Philippines, Diliman, Quezon City, 1001, Philippines - (jelaag, kspilario) @up.edu.ph

Keywords: Deforestation Monitoring, Sentinel-2 Imagery, Ensemble Classification, CNN-LSTM, Forest Forecasting, Narra Palawan

Abstract

Manual approaches to deforestation monitoring are time-consuming, inconsistent, and difficult to scale. Most existing systems rely on subjective interpretation or semi-automated tools, limiting repeatability and real-time assessment. While Sentinel-2 offers high spatial and temporal resolution, persistent cloud cover, especially in tropical areas like Palawan, hampers consistent observation. Moreover, most deforestation studies focus only on detecting past changes, lacking spatially aware forecasting capabilities. This study presents an automated, regionally trained pipeline for deforestation monitoring and forecasting using cloud-free Sentinel-2 imagery and machine learning. Imagery from 2017 to 2024 was preprocessed using OmniCloudMask for cloud and shadow masking, followed by VPint2 to fill cloud-covered gaps in the imagery. Manual annotations were used to classify land cover into six classes using Random Forest, XGBoost, and LightGBM. Random Forest achieved the best performance (90.59% accuracy). Forest classification in 2019 was validated against NAMRIA's 2020 Land Cover Map, with an F1-score of 89.5% and IoU of 81% for the forest class, confirming strong agreement with ground truth for forest cover. Forest area declined gradually from 2017 to 2024, especially along edges near expanding croplands. To anticipate future change, a CNN–LSTM model was trained on tree probability maps to forecast forest cover from 2025 to 2029. The model achieved 93.12% accuracy and a forest F1-score of 92.48% when validated on 2024 data. The proposed system provides an objective and repeatable approach for forest monitoring and near-term forecasting.

1. Introduction

Deforestation remains a major global environmental issue, causing biodiversity loss, ecosystem disruption, and accelerating climate change. In the Philippines—one of the most affected countries—Palawan stands out for its rich biodiversity and extensive natural forests. However, the province has experienced significant deforestation, with over 44,000 hectares lost in the past two decades (Palawan News, 2021), driven by agriculture, illegal logging, mining, and infrastructure development (Nolos et al., 2022). Timely and accurate monitoring is essential to guide conservation and sustainable land use planning.

Advances in remote sensing and machine learning have enhanced deforestation analysis by improving detection speed and accuracy. Recent work has explored integrating artificial intelligence and satellite imagery for forest surveillance (Haq et al., 2024). For example, Maheswara and Nurwatik (2024) applied Sentinel-2 data and random forest classification to monitor forest cover and forecast changes in Indonesia (Maheswara and Nurwatik, 2024).

Sentinel-2 is now widely used for land cover monitoring due to its high spatial resolution (10–60 m), frequent revisit intervals, and open-access availability (Phiri et al., 2020). Its multispectral bands support effective analysis of vegetation and surface dynamics. However, persistent cloud cover in tropical areas poses challenges to image usability. While cloud masking tools like Fmask and UnetMobV2 exist, they often struggle with shadow detection and generalization. OmniCloudMask, trained specifically for Sentinel-2, offers improved detection accuracy—91.2% for clouds and 80.5% for shadows (Wright et al., 2025). Complementary to this, VPint2 enables cloud gap reconstruction using spatial-temporal interpolation without requiring extensive training data (Arp et al., 2024).

Global land cover datasets such as Esri's 10-Meter Map and Global Forest Watch have improved access to environmental insights (Esri and Microsoft, 2021, Global Forest Watch, 2024), but often suffer from poor performance in ecologically diverse regions due to limited regional representation (Zhang and Li, 2022). Locally trained models have been shown to outperform these global products in producing more accurate and context-specific maps (Tadesse et al., 2024).

While most platforms focus on classification and historical trends, few offer forecasting capabilities. Existing predictive approaches often rely on basic regressions that lack spatial nuance. Robust forecasting requires models that can learn complex spatial-temporal patterns.

This study proposes a machine learning-based pipeline for deforestation monitoring and forecasting in Narra, Palawan using Sentinel-2 imagery and regionally trained data. The approach addresses current limitations by integrating cloud-resilient preprocessing, local annotation, ensemble classification, and deep learning forecasting to produce reliable and scalable forest cover predictions.

2. Study Area and Dataset

Figure 1 shows the study area of this research, which is Narra, Palawan. It was selected as the study site due to its ecological sensitivity and growing environmental risks linked to land cover change, particularly from mining activities. While large-scale deforestation has not been widely reported, areas like Bato-Bato

have experienced forest loss, flooding, and reduced agricultural productivity following the reopening of open-pit nickel mining in 2011 (Acero, 2023). Soil and vegetation studies in the area show elevated nickel contamination, with potential long-term impacts on ecosystem health and food security. Narra also covers part of the Victoria-Anepahan Mountain Range (VAMR), a Key Biodiversity Area (KBA) known for its rich concentration of endemic and endangered species (Palawan News, 2022).

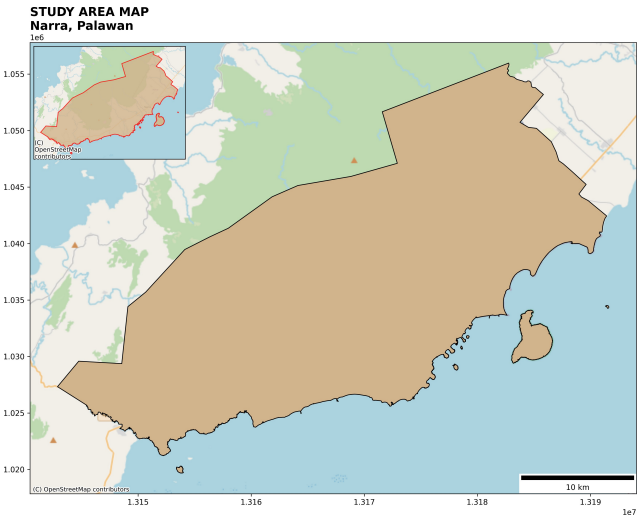

Figure 1. Study Area.

Administrative boundaries for Narra were sourced from the Philippine Statistics Authority (PSA) and NAMRIA via the Humanitarian Data Exchange (HDX) platform (Philippine Statistics Authority (PSA) and National Mapping and Resource Information Authority (NAMRIA), 2023). Sentinel-2 Level-2A imagery was acquired through Sentinel Hub from 2017 to 2024 at two-week intervals, selecting the least cloud-covered scenes during the dry season (February–April) for seasonal consistency. Each image includes thirteen spectral bands, with analysis focused on those available at 10-meter resolution: B02–B04 (visible), B05–B07 (red edge), B08 and B8A (near-infrared), and B11–B12 (shortwave infrared) (Sentinel Hub, 2023), (European Space Agency, 2023). These datasets provided the spectral and temporal foundation for the land cover analysis.

3. Methodology

The flowchart in Figure 2 presents the methodological workflow used in this study for forest and non-forest classification and land cover forecasting using Sentinel-2 Level-2A imagery from 2017 to 2024. The process begins with the acquisition of spatially clipped Sentinel-2 imagery for the Narra, Palawan study area. Cloud-affected pixels were identified using the OmniCloudMask algorithm, which performs multi-class pixel-wise classification to distinguish between clouds, shadows, cirrus, haze, and clear-sky conditions. Cloud-labeled pixels were subsequently masked from the imagery.

To restore missing spectral information in the cloud-masked areas, VPint2 interpolation was applied, generating cloud-free composite images by leveraging temporal consistency across the time series. The reconstructed images were then used for pixel-level annotation, where forest and non-forest classes were manually labeled. These annotations were integrated with spectral reflectance values through multiband reflectance preprocessing, involving per-band scaling, clipping, and stacking into a multispectral data cube. The labeled data points were used to train a supervised land cover classification model.

Following model evaluation, the best-performing model was applied to classify all preprocessed imagery across the study period. The resulting classified satellite images were used for subsequent forest change detection and future land cover forecasting.

3.1 Cloud Masking

This study used the OmniCloudMask (OCM) deep learning algorithm for cloud and shadow masking on Sentinel-2 Level-2A imagery. OCM is a sensor-agnostic segmentation model trained on the global CloudSEN12 dataset, capable of detecting cloud, thin cloud, and shadow using only red, green, and near-infrared bands. It achieves over 91% accuracy for cloud/clear classes and 80% for shadows on Sentinel-2 data (Wright et al., 2025). Key innovations include dynamic Z-score normalization, which adjusts pixel intensities per patch to handle varying illumination, and a mixed-resolution training strategy for consistent performance across sensors and spatial scales.

OCM was selected for its robustness in isolating both thick and thin clouds—crucial for enabling accurate VPint2 interpolation of cloud-contaminated pixels. This ensured the creation of temporally complete, radiometrically consistent image sequences for reliable land cover classification and forecasting. The OCM library is open-source and accessible via PyPI[1] which supports reproducibility and scalability in remote sensing workflows.

3.2 Cloud Reconstruction

Cloud-covered pixels in Sentinel-2 imagery were reconstructed using the VPint2 algorithm (Arp et al., 2024), a training-free spatial interpolation method tailored for optical satellite data. VPint2 fills in missing values by referencing the most recent cloud-free image of the same area and leveraging spatial relationships within the landscape.

After cloud masking via OmniCloudMask, masked pixels were treated as missing data. VPint2 did not copy reflectance values directly; instead, it computed spatial weights based on structural similarity within the cloud-free reference. This approach preserves local textures and features, assuming that spatial structures like roads, field edges, and forest boundaries remain stable over time—even as reflectance may vary due to seasonal or vegetation changes. This ensured visually coherent and radiometrically consistent image sequences for land cover analysis.

To minimize cloud contamination, we restricted image selection to the dry season (February–April), when cloud cover is generally lower and vegetation is captured at a consistent phenological stage. A March acquisition with the least visible clouds was selected as the baseline composite. Residual clouds were addressed by dividing the scene into patches, and for each patch the least cloud-contaminated image was used as a reference before applying VPint2.

Similar to mosaicking, VPint2 relies on cloud-free observations from nearby dates, but instead of directly stitching them into the composite, it adjusts brightness, hue, and texture to match the base image.

[1] https://pypi.org/project/omnicloudmask/

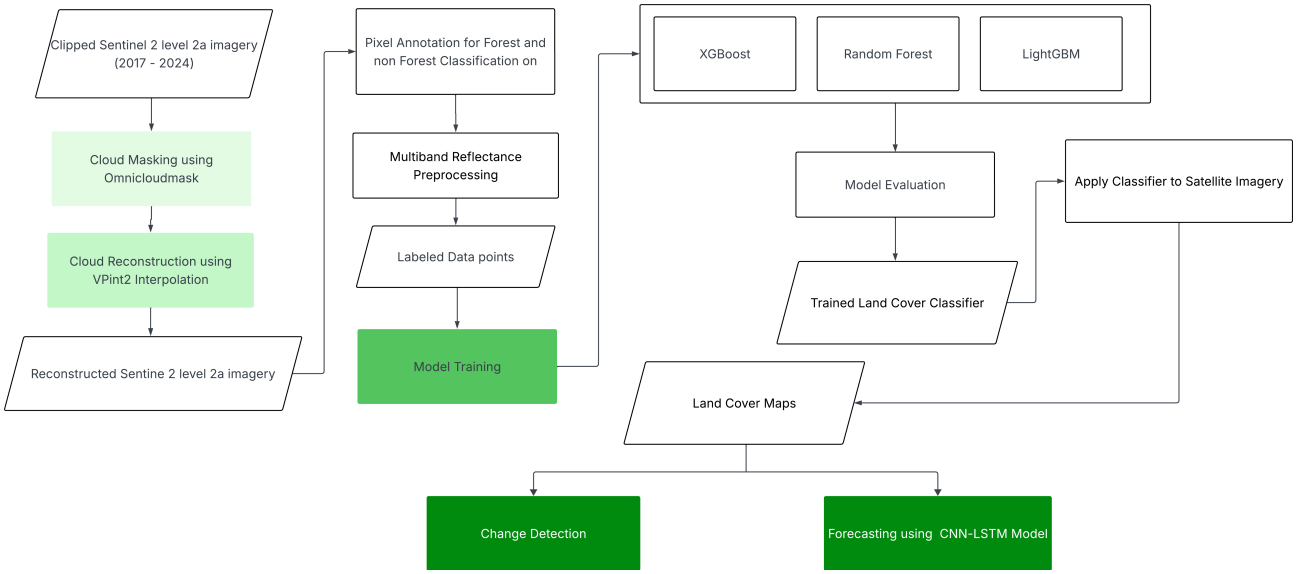

Figure 2. Research Flowchart.

3.3 Annotation and Labeling

Manual annotation was performed in QGIS to generate reference data for supervised land cover classification. Using Sentinel-2 imagery and contextual knowledge, representative points were labeled into six classes based on an adapted version of Esri's Land Cover schema (Esri and Microsoft, 2021): (1) Trees – tall, dense vegetation with continuous canopy; (2) Crops – cultivated vegetation below tree height; (3) Built Area – impervious surfaces like buildings and roads; (4) Bare Ground – exposed soil, sand, or rock with minimal vegetation; (5) Rangeland – open areas dominated by grasses or shrubs; and (6) Water – consistently water-covered areas with little to no vegetation.

To complement these pixel counts, qualitative assessment of classification maps confirmed that forest pixels appeared in contiguous upland blocks, built-up and cropland classes were concentrated along lowlands and road networks, and water pixels were sharply delineated along coasts, consistent with known land use in Narra.

Each point was georeferenced and aligned with the spatial resolution of the VPint2-reconstructed Sentinel-2 stack. Annotations were then rasterized into pixel-wise class masks, forming the labeled dataset used to train and validate the classification model.

Training data were collected from the 2017 imagery through manual annotation of land cover points. Using a single reference year ensured that the model learned stable spectral–textural signatures and reduced risk of overfitting to annual fluctuations. We note that the absence of a formal probability-based sampling design (Stehman and Foody, 2019) may introduce bias and limit representativeness, so accuracy results should be interpreted with caution.

3.4 Machine Learning Models

Three ensemble models—Random Forest (RF), XGBoost, and LightGBM—were used for land cover classification. These methods are accurate, efficient, and robust with limited training data, making them well suited for remote sensing tasks. While deep learning models such as U-Net can achieve strong results, they require large labeled datasets and high computational resources, which were beyond the scope of this study. Since our goal is to develop an operational and deployable workflow, we prioritized lightweight ensemble methods that balance performance, efficiency, and ease of integration.

3.4.1 Random Forest Random Forest is a non-parametric ensemble method that builds multiple decision trees using random subsets of data and features. It reduces overfitting through averaging and performs well on noisy and imbalanced datasets. RF has been widely used in remote sensing for its strong classification performance and ease of interpretation (Belgiu and Dra˘gu¸t, 2016).

3.4.2 XGBoost XGBoost is an efficient gradient boosting algorithm that builds decision trees sequentially, correcting previous errors with each iteration. It includes regularization and pruning mechanisms that improve generalization. XGBoost has shown strong results in land cover and vegetation mapping using satellite data such as Sentinel-2 (Magidi et al., 2025).

3.4.3 LightGBM LightGBM is a fast and memory-efficient boosting algorithm that grows trees leaf-wise with depth limits. Its histogram-based splits and native handling of categorical features make it ideal for large remote sensing datasets. It has been applied successfully to land cover classification across varied spatial and temporal resolutions (Li et al., 2024).

3.5 Model Training, Evaluation, and Change Detection

Three ensemble-based classifiers—Random Forest (RF), XGBoost, and LightGBM—were trained to perform land cover classification using 13 spectral bands from reconstructed Sentinel-2 imagery (2017). Manually annotated points served as reference labels, and min-max normalization was applied. A stratified 70/30 split ensured class balance across training and testing sets (Table 1). Hyperparameters were tuned via grid search with 3-fold cross-validation, and model performance was evaluated using accuracy, precision, recall, F1-score, and class-specific Producer's and User's Accuracy.

Land Cover Class	Train Count	Test Count
Tree	8,480	3,634
Crops	7,913	3,392
Built Area	7,300	3,128
Bare Ground	6,911	2,962
Rangeland	9,882	4,236
Water	2,443	1,047

Table 1. Number of training and testing samples per land cover class.

The best-performing model, based on F1-score, was applied to classify annual images from 2017 to 2024. A 7×7 majority filter was used post-classification to reduce speckle and improve spatial coherence (Danoedoro et al., 2021). To assess classification accuracy and spatial reliability, the 2019 output was compared to NAMRIA's 2020 Land Cover Map, which was derived from Sentinel-2 imagery (Dec 2018–Apr 2019) and validated through field surveys (Dag-uman, 2020). The NAMRIA map was reclassified into six land cover categories to match the model's schema and served as a benchmark for evaluating thematic consistency.

For change detection, we tracked the percentage share of each land cover class annually from 2017 to 2024. Instead of pixel-level differencing, a class-based statistical approach was used, computing per-class pixel counts as a proportion of the total classified area. This method captured general trends in land cover dynamics—highlighting shifts in forest, cropland, built-up, and other classes over time.

3.6 Forecasting Using CNN-LSTM Architecture

To forecast forest cover, we adapted a CNN–ConvLSTM–CNN architecture based on (Varma et al., 2024). The model used 6-year sequences (2017–2022) of tree probability rasters, split into 64×64-pixel patches. Only patches containing some forest pixels were included to ensure informative learning.

The architecture consists of: (1) a CNN encoder to extract spatial features from each annual frame; (2) a ConvLSTM unit to capture spatiotemporal dynamics; and (3) a CNN decoder to generate pixel-wise binary forest predictions. Training used the 2023 land cover map as ground truth and validated on 2024 data, with an 80/20 split. The model was trained for 50 epochs using the Adam optimizer and categorical cross-entropy loss, and saved for future forecasting from 2025 to 2029.

4. Results and Discussion

4.1 Cloud Masking and VPint2 Interpolation

To address cloud cover in Sentinel-2 images, OmniCloudMask and VPint2 were used for masking and reconstruction, respectively. The resulting cloud-free mosaics supported year-to-year comparison from 2017 to 2024. As seen in Figure 5 in the appendix, the resulting mosaics are visually consistent and cloud-free, making them suitable for year-to-year comparison from 2017 to 2024.

4.2 Model Evaluation

We tested three ensemble models—Random Forest, LightGBM, and XGBoost—on a stratified test set covering six land cover classes. As shown in Table 2, Random Forest achieved the best performance (90.6% accuracy) and was selected for final mapping. To gain insight into how well the classes separate in feature space, we ran a t-SNE projection (Figure 3). Tree and Water samples formed clear clusters, while the rest—especially Crops, Range Land, and Built-up areas—showed some overlap, which helps explain the relatively lower performance in those categories.

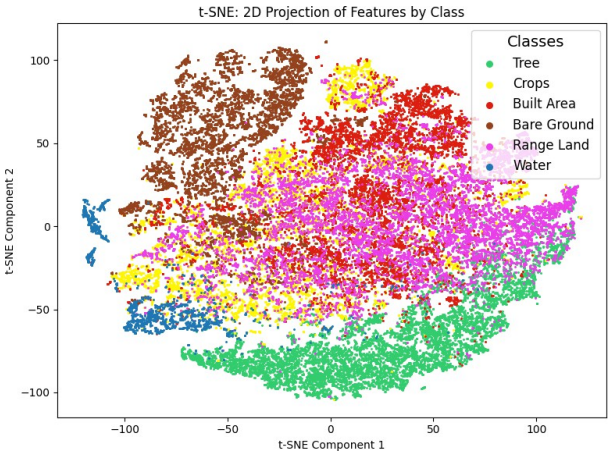

Figure 3. t-SNE projection of labeled land cover samples from 13-band Sentinel-2 data. Tree and Water form distinct clusters, while other classes overlap due to spectral similarity.

Table 3 shows that the model handled Tree, Water, and Bare Ground best, all with F1-scores over 96%. Crops and Range Land had the lowest scores, likely due to their spectral similarity and seasonal variation. Given its strong and consistent results, Random Forest was selected for generating the final land cover maps.

4.3 Land Cover Classification

Maps show stable upland forest and gradual cropland expansion in lowland areas from 2017 to 2024. As shown in Figures 6, and 7 in the appendix, tree cover remained stable in the upland areas, while Croplands gradually expanded in the eastern and southeastern parts. Built-up areas also grew slightly near roads and settlements.

Range Land often appeared in transitional zones between Tree, Crops, and Bare Ground, reflecting its mixed spectral nature. Water bodies stayed mostly consistent, with only minor seasonal or reconstruction-related changes. Overall, the classification maps capture both stable land cover zones and areas of gradual change over time.

One notable limitation was the spectral confusion observed between crops and rangeland, which are spectrally similar in single-season composites. This confusion introduces uncertainty for monitoring, particularly in transitional zones where croplands expand into previously mixed-use areas. Such misclassification may lead to under- or overestimation of cropland-driven forest conversion. Future work could reduce this issue by using additional vegetation indices, seasonal composites, or ancillary data such as topography.

4.4 Comparison with NAMRIA 2020 Land Cover Map

To evaluate the accuracy of the 2019 land cover output, we compared it with the 2020 Land Cover Map developed by NAMRIA. This reference dataset was created through digital interpretation of Sentinel-2 imagery (10-meter resolution) acquired

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.8837	0.8838	0.8837	0.8837
LightGBM	0.9015	0.9012	0.9015	0.9013
Random Forest	**0.9059**	**0.9058**	**0.9059**	**0.9058**

Table 2. Performance comparison of machine learning models for land cover classification.

Class	Producer's Acc.	User's Acc.	Precision	Recall	F1-Score
Tree	0.9684	0.9524	0.9524	0.9684	0.9603
Crops	0.8517	0.8452	0.8452	0.8517	0.8485
Built Area	0.9057	0.8957	0.8957	0.9057	0.9007
Bare Ground	0.9706	0.9843	0.9843	0.9706	0.9774
Range Land	0.8243	0.8420	0.8420	0.8243	0.8331
Water	0.9599	0.9654	0.9654	0.9599	0.9626

Table 3. Per-class accuracy metrics for the Random Forest model.

between December 2018 and April 2019, and was validated with a ground survey (Dag-uman, 2020). For alignment with our classification scheme, NAMRIA's original 12 categories were reclassified into six broader classes as follows: (1) Forest – including Closed Forest, Open Forest, and Mangrove Forest; (2) Cropland – comprising Annual Crop and Perennial Crop; (3) Built-up – consisting only of Built-up areas; (4) Bare Ground – representing Open/Barren land; (5) Rangeland – combining Brush/Shrubs and Grassland; and (6) Water/Others – covering Fishpond, Inland Water, and Marshland/Swamp. The comparison was limited to the Narra municipal boundary, with the NAMRIA map rasterized to 10-meter resolution to match the model output.

As shown in Table 4, the model achieved an overall accuracy of 71.5% , indicating moderate to substantial agreement with the reference. Class-specific metrics (Table 5) show strong performance for the Forest class (F1 = 0.895, IoU = 0.81), which is crucial for deforestation monitoring. Cropland and Rangeland were classified moderately well, while Built-up, Bare Ground, and Water/Others were more difficult to distinguish—likely due to spectral confusion, small object size, or limited training data. Despite these limitations, the results validate the model's ability to map forest cover accurately and support its use in environmental monitoring.

Metric	Value
Overall Accuracy	0.715
Mean Intersection-over-Union (IoU)	0.365

Table 4. Accuracy metrics comparing model output and 2020 NAMRIA land cover.

Class	IoU	Precision	Recall	F1-score
Forest	0.8099	0.841	0.956	0.895
Cropland	0.3942	0.748	0.455	0.565
Built-up	0.1917	0.574	0.224	0.322
Bare Ground	0.2023	0.233	0.607	0.337
Rangeland	0.4210	0.528	0.676	0.592
Water/Others	0.1691	0.424	0.220	0.289

Table 5. Per-class evaluation metrics.

4.5 Change Detection

Deforestation was quantified using post-classification differencing, where a change is defined if a pixel's class label differed between two years. This approach not only identifies where changes occurred but also specifies the type of transition (e.g., forest to cropland, forest to built-up). Such transition-level information provides more insight into land-use pressures than binary change/no-change labels. Although direct change detection from raw time-series imagery is possible, it requires a dataset specifically labeled for change and no-change, which was not available for this study.

From 2017 to 2024, land cover in Narra, Palawan changed gradually. Tree cover remained dominant but declined steadily, likely due to expanding agriculture and rural settlements. Cropland increased notably between 2017 and 2019, while built-up areas grew slowly, reflecting ongoing municipal development (Table 6). Range Land increased slightly in recent years, possibly indicating conversion from former forest or cropland areas into more open, mixed-use landscapes. Bare Ground and Water remained relatively stable, with only minor year-to-year changes that may be tied to seasonal variation or classification differences. Overall, the results suggest low but ongoing land use pressure, with slow forest loss and crop expansion being the most consistent trends.

Year	Tree	Crops	Built Area	Bare Ground	Range Land	Water
2017	41.62	9.60	1.43	0.90	20.85	0.92
2018	39.87	10.82	1.92	1.03	20.84	0.83
2019	37.51	13.40	1.86	1.36	20.50	0.69
2020	36.60	13.25	1.43	1.46	21.80	0.77
2021	37.43	11.83	1.92	1.36	21.77	1.00
2022	37.65	10.29	2.41	1.04	22.67	1.25
2023	37.69	11.20	2.26	1.13	21.75	1.28
2024	36.33	12.00	2.02	1.28	22.86	0.83

Table 6. Annual area (in kilohectares, kha) of land cover classes in Narra, Palawan from 2017 to 2024.

4.6 Forecasting

A CNN–LSTM model was trained on tree probability maps from 2017 to 2022 and used to forecast forest/non-forest classifications from 2025 to 2029. Model performance was first validated by predicting tree cover for 2024 and comparing it with the actual classification, achieving strong results (overall accuracy 93.12% and F1-score 92.48% for forest). This indicates reliable identification of forested areas in the near term. However, forecasts were validated only against 2024 data, since no independent land cover maps beyond this year were available. As a result, predictions for later years, especially 2029, should be interpreted with caution and considered indicative rather than definitive.

The forecasted maps (Figure 9) show a gradual but steady decline in forest cover, with the most noticeable reductions occurring after 2025 along agricultural boundaries and in the western part of Narra. The area trend derived from the forecast maps

(Figure 4) reflects this continued pressure from land conversion. While these results are consistent with observed trends from 2017 to 2024, they also emphasize the need for proactive monitoring and management. To strengthen long-term forecasting, future work will incorporate updated reference maps from NAMRIA or field surveys, explore hindcasting experiments to test robustness, and generate cloud-frequency and uncertainty layers. These additions will help distinguish highly reliable predictions in cloud-free zones from less certain estimates in persistently cloudy areas.

Metric	Value
Overall Accuracy	0.9312
Precision (Tree)	0.9787
Recall (Tree)	0.8766
F1-score (Tree)	0.9248
Precision (Non-tree)	0.8951
Recall (Non-tree)	0.9822

Table 7. Evaluation metrics for tree cover forecasting in 2024.

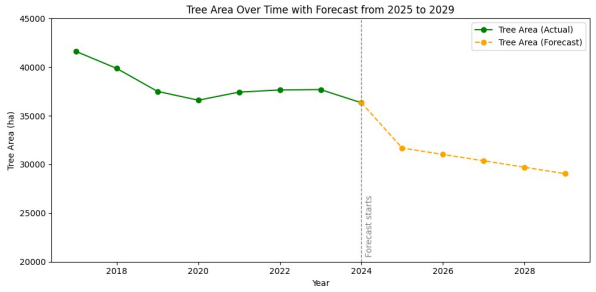

Figure 4. Trend in tree area from 2017 to 2024 (actual) and 2025 to 2029 (forecasted). The dashed line indicates the beginning of the forecast period.

5. Conclusion

This study developed a cloud-aware and locally trained land cover classification and forecasting pipeline for Narra, Palawan using Sentinel-2 imagery. The integration of OmniCloudMask and VPint2 reconstruction allowed consistent, cloud-free composites to be generated for the dry season from 2017 to 2024. Supervised classification using Random Forest outperformed other ensemble models and accurately mapped six major land cover classes.

The analysis revealed that while tree cover remained dominant, its extent gradually declined over the study period, particularly along agricultural frontiers. Change detection captured these class-specific shifts spatially and temporally.

Finally, a CNN-LSTM model was employed to forecast forest cover for 2025–2029, showing a potential continuation of this declining trend. These forecasts provide early warning signals for decision-makers and emphasize the need for proactive land management and conservation strategies.

6. Recommendations

To strengthen land monitoring in Narra, Palawan, several steps are recommended. First, the projected forest loss underscores the need for stronger conservation policies, particularly in areas experiencing agricultural expansion. Second, regular updates to training datasets and the use of recent satellite imagery will improve the responsiveness of the models to ongoing changes. The inclusion of socio-economic and zoning information such as roads, land permits, and population growth can further refine forecasts by linking observed patterns with potential drivers. Third, future versions of the pipeline may adopt advanced architectures such as attention-based networks or spatiotemporal transformers, which are better suited to capturing long-term dynamics. Expanding the training dataset through probability-based sampling designs (Stehman and Foody, 2019) would also reduce bias and provide a stronger foundation for testing deep learning approaches once larger datasets become available. The transferability of the workflow should be assessed by applying it to other regions with different land cover dynamics and socio-economic conditions to ensure broader applicability. Addressing class confusion, particularly between crops and rangeland, is also a priority. Potential solutions include incorporating additional spectral indices such as NDVI and EVI, integrating seasonal composites to capture phenological differences, and using ancillary data such as topography or zoning maps to improve separability. Finally, deploying the workflow as a web or mobile application would empower local governments, civil society, and communities to monitor forest change in near real time, thereby enhancing transparency and enabling proactive responses to emerging deforestation risks.

References

Acero, L. H., 2023. Nickel Content in Plants and Soil: The Case of Mine Tailing Sites in Bato-Bato Narra Palawan, Philippines. *Jurnal Manajemen Hutan Tropika*, 29(3), 200–207.

Arp, L., Hoos, H., van Bodegom, P., Francis, A., Wheeler, J., van Laar, D., Baratchi, M., 2024. Training-Free Thick Cloud Removal for Sentinel-2 Imagery Using Value Propagation Interpolation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 216, 168–184. doi.org/10.1016/j.isprsjprs.2024.07.030.

Belgiu, M., Dra˘gu¸t, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. doi.org/10.1016/j.isprsjprs.2016.01.011.

Dag-uman, D. K., 2020. 2020 land cover map of the province of palawan. National Mapping and Resource Information Authority (NAMRIA). Available from NAMRIA.

Danoedoro, P., Hartono, M. D., Susanti, D., 2021. Evaluating performance of majority rules in per-pixel and obia land-cover classification using worldview multispectral imagery of urban fringe area, indonesia. *Proceedings of the 42nd Asian Conference on Remote Sensing (ACRS)*.

Esri, I., Microsoft, 2021. Esri 10-meter land cover map (2020–2021). livingatlas.arcgis.com/landcover/ (16 May 2025).

European Space Agency, 2023. Copernicus open access hub - sentinel-2. scihub.copernicus.eu/ (16 May 2025).

Global Forest Watch, 2024. Global forest watch dashboard. globalforestwatch.org/dashboards/global/ (16 May 2025).

Haq, B., Ali Jamshed, M., Ali, K., Kasi, B., Arshad, S., Khan Kasi, M., Ali, I., Shabbir, A., Abbasi, Q. H., Ur-Rehman, M., 2024. Tech-Driven Forest Conservation: Combating Deforestation With Internet of Things, Artificial Intelligence, and Remote Sensing. *IEEE Internet of Things Journal*, 11(14), 24551-24568.

Figure 5. **VPint2 cloud-minimization workflow.** (a) Least-cloudy base image (patch) from **March 12, 2017**; (b) Cloud mask from OmniCloudMask: black = clear, gray/white = cloud and shadow.; (c) base image with mask applied, showing residual cloud/opaque areas; (d) **cloud-free reference** from **July 9, 2017** (nearest low-cloud acquisition for this patch); (e) final result after applying *VPint2*. Image selection was restricted to the dry season (Feb–April), and the March scene was chosen as the baseline to keep phenology and illumination consistent. For each patch, a nearby cloud-free observation (here July 9) was used only as a donor. Rather than stitching, *VPint2* transfers brightness, hue, and local texture from the donor to the base so infilled areas match the base image.

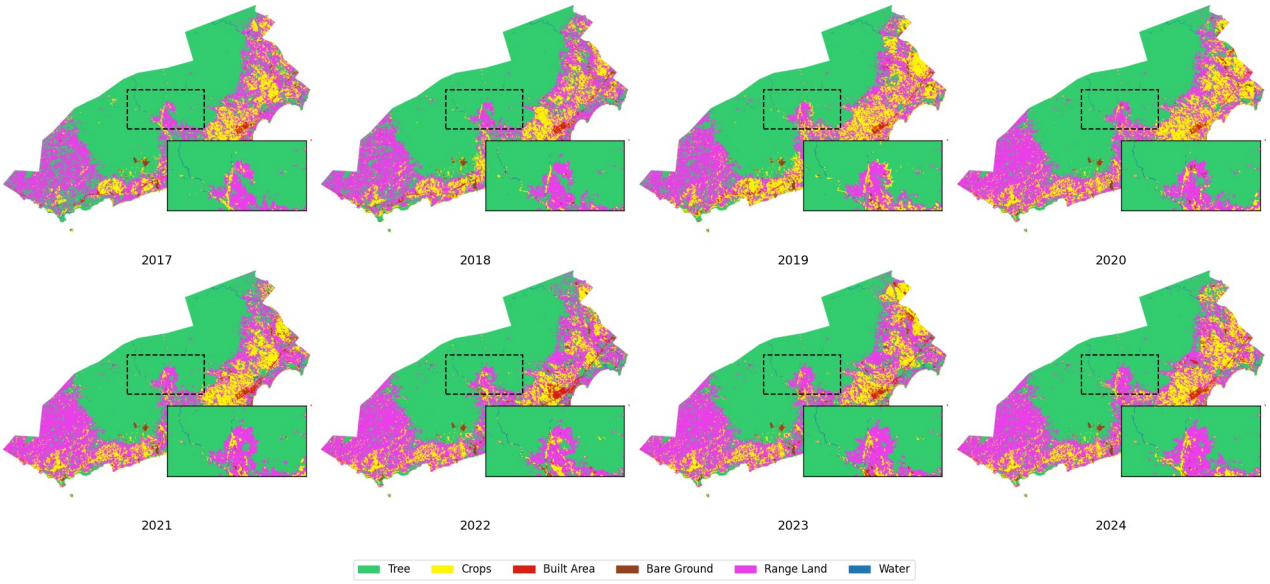

Figure 6. Land Cover maps of Narra, Palawan (2017–2024) from VPint2-reconstructed Sentinel-2 imagery, classified into six classes using the top-performing ML model. Forest cover remains stable in uplands, while croplands and built-up areas expand in lowlands

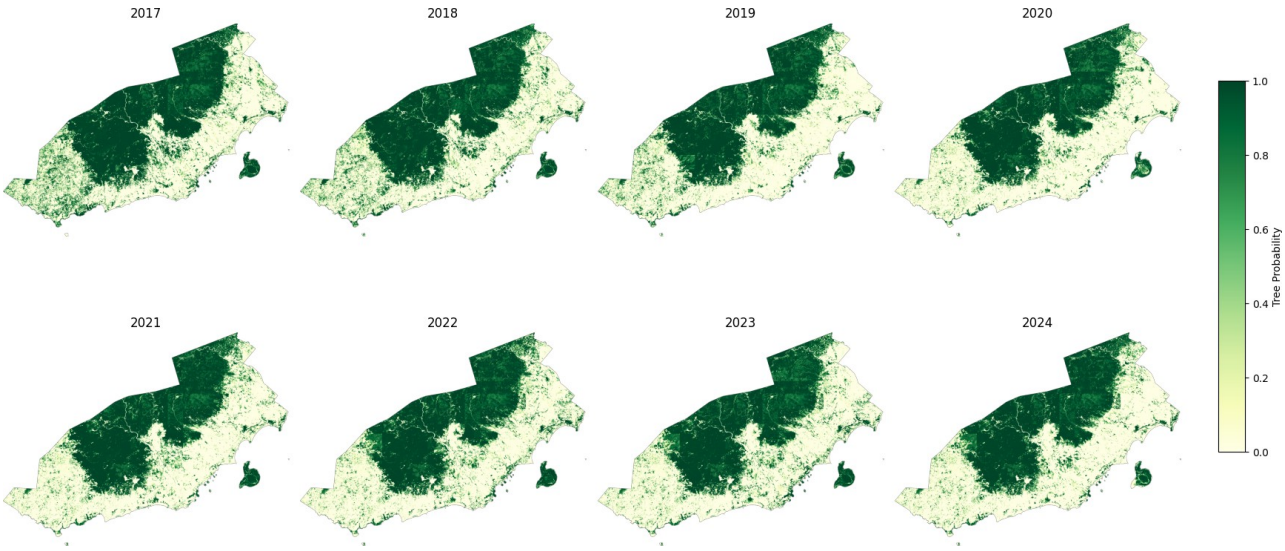

Figure 7. Annual tree probability maps (2017–2024) for Narra, Palawan, derived from the best-performing classification model.

Li, R., Gao, X., Shi, F., 2024. A Framework for Sub- region Ensemble Learning Mapping of Land Use/Land

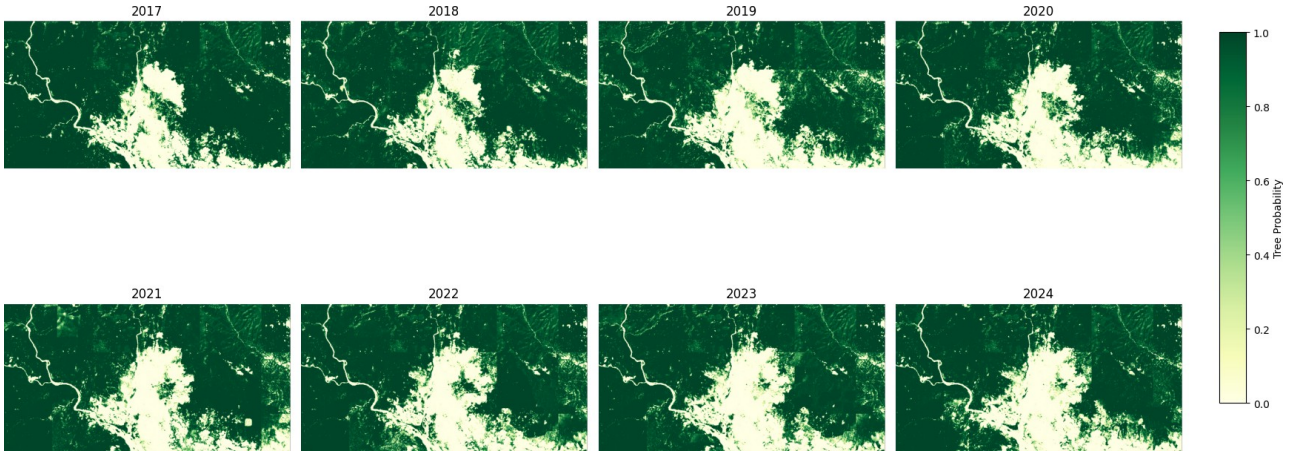

Figure 8. Zoomed-in view of annual tree probability maps (2017–2024) for a representative area in Narra, Palawan. The closer look highlights gradual forest loss and expansion of non-forest classes, particularly along the lowland–upland transition zones.

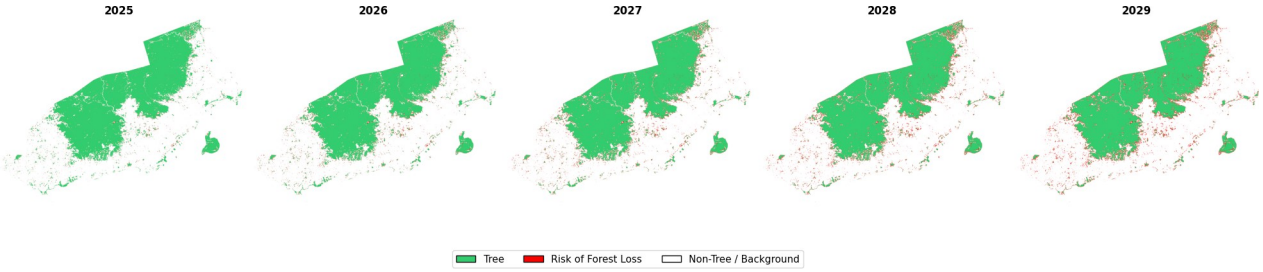

Figure 9. Forecasted binary tree masks from 2025 to 2029 in Narra, Palawan. Areas in green represent predicted forest cover.

Cover at the Watershed Scale. *Remote Sensing*, 16(20). doi.org/10.3390/rs16203855.

Magidi, J., Bangira, T., Kelepile, M., Shoko, M., 2025. Land use and land cover changes in Notwane watershed, Botswana, using extreme gradient boost (XGBoost) machine learning algorithm. *African Geographical Review*, 44(5), 497–517. doi.org/10.1080/19376812.2024.2424378.

Maheswara, S. D. A., Nurwatik, N., 2024. Using Sentinel-2 Imagery and Machine Learning Algorithms to Estimate Deforestation (Case Study: a district in Malang Regency). *IOP Conference Series: Earth and Environmental Science*, 1418(1), 012008. doi.org/10.1088/1755-1315/1418/1/012008.

Nolos, R. C., Zamroni, A., Evina, K. P., 2022. Drivers of deforestation and forest degradation in Palawan, Philippines: An analysis using Social-Ecological Systems (SES) and Institutional Analysis and Development (IAD) approaches. *Geography, Environment, Sustainability*, 15(4), 44–56. doi.org/10.24057/2071-9388-2022-081.

Palawan News, 2021. Palawan lost over 44,000 hectares of natural forest over the past 20 years, says int'l research group. palawan-news.com/palawan-lost-over-44000-kilohectares-of-natural-forest-over-the-past-20-years-says-intl-research-group (16 May 2025).

Palawan News, 2022. Lgus ink agreement for sustainable management of victoria-anepahan mountain range. palawan-news.com/lgus-ink-agreement-for-sustainable-management-of-victoria-anepahan-mountain-range/ (16 May 2025).

Philippine Statistics Authority (PSA), National Mapping and Resource Information Authority (NAMRIA), 2023. Philippines - subnational administrative boundaries (level 0–3). data.humdata.org/dataset/cod-ab-phl (16 May 2025).

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(12), 2291. doi.org/10.3390/rs12142291.

Sentinel Hub, 2023. Sentinel hub api documentation. docs.sentinel-hub.com/api/latest/ (16 May 2025).

Stehman, S. V., Foody, G. M., 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, 231, 111199.

Tadesse, G. A., Robinson, C., Mwangi, C., Maina, E., Nyakundi, J., Marotti, L., Hacheme, G. Q., Alemohammad, H., Dodhia, R., Lavista Ferres, J. M., 2024. Global: Local land-use and land-cover models deliver higher quality maps. Preprint, arXiv:2412.00777v2 (16 May 2025).

Varma, B., Naik, N., Chandrasekaran, K., Venkatesan, M., Rajan, J., 2024. Forecasting Land-Use and Land-Cover Change Using Hybrid CNN–LSTM Model. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5.

Wright, N. et al., 2025. Training sensor-agnostic deep learning models for remote sensing: Achieving state-of-the-art cloud and cloud shadow identification with Omni-CloudMask. *Remote Sensing of Environment*, 322, 114694. doi.org/10.1016/j.rse.2025.114694.

Zhang, C., Li, X., 2022. Land Use and Land Cover Mapping in the Era of Big Data. *Land*, 11(10), 1692. doi.org/10.3390/land11101692.