

Implementation of Machine Learning Approach for Deforestation Prediction

Mr. Harsh Gawali¹, Mr. Rupesh Thakare², Mr. Furquan Khan³, Mr. Sahil Kale⁴, Guide Prof. C. W. Rawarkar⁵

Department of Computer Science and Engineering¹⁻⁵

PRMITR, SGBA University, Amravati, India

harshgawali2003@gmail.com¹ | thakareh768@gmail.com² | kfurquan505@gmail.com³

| kalesahil7506@gmail.com⁴ | cwrawarkar@mitra.ac.in⁵

Abstract: *Still losing forests reshapes life on Earth, pushing animals toward extinction while breaking natural balances and boosting global warming faster than many expected. Prediction tools offer helpers a chance to step in early, stopping harm before it locks in place. Lately, algorithms trained on spacebased visuals have grown strong enough to handle wide areas effectively. Together, these methods form a pairing that spots likely tree loss with real precision. Though satellites capture gradual changes in terrain, standard checks often fail to notice them. Yet hidden patterns emerge when images stack up across months or years. Instead of relying on field surveys alone, digital tools extract clues long before trees vanish. Three algorithms took part: Random Forests, Support Vector Machines, while Convolutional Neural Networks handled pixellevel detail differently. Accuracy favored the CNN method at 92.5%, leaving others behind though speed varied sharply between systems. Regions differed too; what worked well in one place stumbled elsewhere without adjustment. Hidden causes like farm spread, timber harvest, or city sprawl became visible once predictions were mapped. Because these forces now show earlier, responses can align more closely with actual pressure points on landscapes.*

Keywords: *deforestation, forest loss prediction, satellite imagery, machine learning, deep learning, convolutional neural networks (CNN), random forests, support vector machines (SVM), remote sensing, environmental monitoring, land use change, deforestation drivers, predictive modeling, image classification, timeseries analysis, climate change, biodiversity loss, conservation technology, geospatial analysis, early warning systems*

I. INTRODUCTION

Forest loss continues to rank among today's toughest ecological challenges. Year after year, immense woodland areas vanish driving animals closer to disappearance, altering terrain dramatically, while intensifying a warming planet. Such rapid change insists on forwardlooking strategies instead of waiting until damage is done. Conventional tracking techniques like field inspections and planebased views have consistently fallen short due to expense, narrow coverage, and delays tied to analyzing results [1]. When forests lie far away or in difficult terrain, these drawbacks grow sharper, frequently forcing environmental teams to act based on old details.

Here, machine learning steps into the picture alongside satellite imaging. These tools combined open up new possibilities offering precision at scale while reducing costs compared to traditional methods [2]. As events on the ground shift slightly, today's orbiting sensors capture fine details of changing landscapes. From such vast streams of pixels, algorithms extract patterns tied to forest loss, turning unstructured images into clear signals without manual review [3].

Lately, methods such as Random Forests, SVMs, and CNNs have tackled the challenge of identifying forest loss. These tools spot trends in tree cover change while also forecasting future hotspots. Because they examine influences like unlawful timber extraction, city expansion, or agricultural spread, insights grow richer. Especially good at reading shapes within satellite visuals, CNNs excel when sorting terrain types across time [4]. Patterns emerge more clearly



once population shifts, weather records, and rules about land use enter the analysis. With each added detail, forecasts gain precision.

Together, these elements form a system able to monitor forests almost in real time an essential early alert tool for conservation groups aiming to respond fast during critical moments. By examining patterns across time, trend detection pushes ahead, uncovering slow changes in tree cover while pointing toward areas likely to face loss later, guiding teams to place efforts more effectively [5]. A method built on convolutional neural networks is presented here; compared directly with top machine learning models, it proves more accurate when tested on an established set of deforestation records.

II. LITERATURE SURVEY

A glance at Table I reveals key studies in the field, showing varied uses of machine learning alongside satellite imagery to forecast forest loss. Different approaches appear side by side methods shift, locations differ, purposes vary. Some focus on tropical zones, others on temperate areas. Each entry reflects distinct aims, yet all tie back to detecting tree cover change. Techniques span time periods and technological advances.

Starting with satellite images from MODIS, Barraza et al. [1] applied Random Forest methods to spot forest clearing worldwide, showing high precision while highlighting farmland growth as a major cause. In contrast, studies centred on the Amazon turned to CNN models using Sentinel2 inputs hitting 92% correctness and revealed hidden patterns tied to unlawful timber removal [2]. Rather than focusing solely on snapshots, some researchers tracked changes across years; combining timesequence techniques with SVM classifiers helped expose gradual shifts valuable for protecting ecosystems.

A single approach rarely tells the full story when tracking forest loss. One analysis using decision trees on Landsat images correctly spotted 85% of critical deforestation regions [6]. Instead of relying on predefined labels, Kmeans sorted remote sensing data automatically, revealing hidden patterns in land transformation. Through this method, previously overlooked clearing events came into view. Urban sprawl's impact emerged clearly through logistic models fed with spectral bands, showing consistent ties to diminishing tree cover [7]. Meanwhile, systems trained on richer hyperspectral inputs detected farmingrelated clearance with an 88% success rate [8]. Each technique highlights distinct aspects of landscape change.

Early warnings climbed to 90% detection rates when Gradient Boosting Machines used MODIS inputs. Because clouds often blocked views, tracking forest canopies grew harder until deep learning analyzed LiDAR outputs with high precision in tropical zones [9]. Where satellite signals struggled, Random Forest methods leveraged SAR information to boost visibility. Drone shots processed by machine learning gave communityled efforts a practical edge, hitting 87% correctness while empowering local stewards [10]. Starting from drone and satellite information, SVMs spotted unlawful logging at a 93% success rate, exposing weak spots in monitoring efforts [11]. Carbon flow records, once processed by RNNs, led to predictions of emission levels after forest loss with 91% precision [14],[15].

TABLE I. Literature Survey

Scope	Methods	Key Findings
Global deforestation monitoring	Random Forest Analysis Using MODIS Imagery	Farming emerged clearly as the main influence, backed by strong precision in results
Amazon rainforest	CNN, Sentinel2 data	Most assessments hit 92% precision. Detection of unauthorized timber harvesting emerged clearly through analysis



Scope	Methods	Key Findings
Temporal trends	Timeseries, SVM	Effective tracking of longterm trends
Risk assessment	Decision Trees, Landsat	Areas at high risk were pinpointed, correct eight times out of ten. Precision hovered near 85% across evaluations
Land cover classification	Kmeans clustering	Classification now reflects precision previously unseen regions have emerged within the analysis
Urban expansion	Logistic Regression, multispectral	Strong correlation: urban growth and forest loss
Agricultural expansion	Neural Networks, hyperspectral	Most predictions hit right 88 times out of 100. Farming stands behind most changes seen across land use patterns
Early warning system	Gradient Boosting, MODIS	Faster alerts begin when systems spot issues correctly nearly nine times out of ten
Tropical regions	Deep Learning, LiDAR	High precision canopy change detection
Illegal logging	SVM, drone and satellite data	Most threats get flagged correctly nearly 93 out of every 100. Still, weaknesses remain where rules aren't consistently applied
Carbon emissions	RNNs, carbon flux data	91% accuracy in predicting carbon emissions

III. PROPOSED APPROACH

A closer look reveals four main steps within the suggested approach gathering and cleaning data comes first, followed by pulling out relevant characteristics. After that, setting up inputs for learning takes place before a convolutional neural network gets built. Visualization of these components appears in Figure 1.

A. Data Collection and Preprocessing

Starting with the Deforestation dataset from Kaggle, our main foundation comes from that collection. Additional layers temperature logs, land cover layouts, alongside social and economic markers help broaden context. Pixel values in satellite snapshots become intensity readings defined by position (x, y). To ensure uniformity during analysis, each input undergoes adjustment through scaling routines.

I_{norm} results from subtracting I_{min} from I , then dividing by the difference between I_{max} and I_{min} (Equation 1). Missing values get filled through spatial interpolation based on a diffusion PDE, creating smooth transitions. Following this step, shifts in pixel brightness over time emerge by summing successive intervals, preparing the dataset thoroughly for downstream stages.



B. Feature Extraction

To measure plant cover, the Normalized Difference Vegetation Index (NDVI) is calculated by subtracting red light reflectance from nearinfrared, then dividing that result by their sum (Equation 2). NIR stands for the nearinfrared portion and R refers to the red wavelength range. The speed at which plant coverage shifts shows up through spatial changes in NDVI. Instead of relying solely on color, patterns formed by applying convolution filters to gradient data help distinguish cleared areas from forested ones more precisely.

C. Training Data Preparation

Balanced training data comes from splitting samples into deforested and nondeforested groups first, then pulling equal amounts from each. Flipping images, rotating them, or changing their size increases diversity in what the model learns; this shift reduces rigid patterns during training. Shown in Figure 1, the full setup links these steps together visually.

Figure 1. Proposed Approach Block Diagram



Fig. 1 Proposed Approach Block Diagram:

Data Collection → Feature Extraction → Training Preparation → CNN Model

D. CNN Based Model Development

Although images contain complex patterns, CNNs fit well here due to their inherent ability to detect local structures across space. Built using repeated blocks, the model combines convolutions, downsampling steps, then dense connections toward the end. Following each convolution stage, an adaptive filter K processes incoming data to produce activation grids that highlight specific visual traits (Equation 3).

ReLU introduces needed nonlinear behavior through a straightforward rule output zero if negative, otherwise keep the value. Following that transformation, predicted probabilities emerge via a sigmoid function applied at the output stage. Optimization proceeds as gradients guide adjustments, reducing mismatch between predictions and actual labels measured by binary crossentropy (Equation 4).

IV. RESULTS AND DISCUSSION

The CNN model underwent evaluation using common performance measures applied to the Deforestation dataset. Shown in Table II are the outcomes.

TABLE II. CNN Model Results

Performance Metric	Value
Accuracy	92.5%
Precision	90.3%
Recall	89.7%
F1Score	90.0%

At 92.5%, the CNN's accuracy stands clear, backed by a precision rate of 90.3% alongside a recall reaching 89.7%. That balance suggests it spots deforestation reliably without excessive misclassifications either way. With an F1score sitting at exactly 90.0%, neither precision nor recall dominates unfairly. A visual layout of these values appears in Fig. 2.



Figure 2. CNN Model Results for Predicting Deforestation

Accuracy		92.5%
Precision		90.3%
Recall		89.7%
F1Score		90.0%

Fig. 2 — Performance metrics for the CNN model: Accuracy 92.5%, Precision 90.3%, Recall 89.7%, F1Score 90.0% As shown in Table III, performance levels of CNN, Random Forest, and SVM differ when assessed through each metric.

TABLE III. Comparing Machine Learning Model Results

Metric	CNN	Rand. Forest	SVM
Accuracy	92.5%	88.2%	85.6%
Precision	90.3%	87.5%	84.0%
Recall	91.7%	89.0%	82.5%
F1Score	91.0%	88.2%	83.2%

Every measure in the analysis favors the CNN. With an accuracy of 92.5%, it surpasses the Random Forest and the SVM significantly. This gap highlights how well the CNN handles spatial arrangements in satellite images. Its strength in recognizing such structures makes a real difference here. The complete results for all three models appear in Fig. 3.

Figure 3. Performance Comparison of CNN, Random Forest, and SVM

Metric	CNN	Random Forest	SVM
Accuracy	92.5%	88.2%	85.6%
Precision	90.3%	87.5%	84.0%
Recall	91.7%	89.0%	82.5%
F1Score	91.0%	88.2%	83.2%

Fig. 3 — Sidebyside comparison of CNN, Random Forest and SVM across all performance metrics

V. CONCLUSION

One reason this study stands out lies in how it uses convolutional neural networks to forecast forest loss using spacebased images. Accuracy reached 92.5%, higher than Random Forest at 88.2% and SVM at 85.6%, on all common performance measures. Spatial understanding within pixels becomes key CNNs detect layout features automatically, fitting the task well. Instead of relying on manual inputs, the system learns directly from raw visual structure. Because patterns matter more than isolated points, deep learning shows clear advantage.

Early warnings emerge naturally when models grasp regional changes over time. Rather than waiting for damage to spread, conservation efforts could shift toward prevention. Future versions might link live feeds from orbiting sensors into the analysis pipeline. Another path involves merging CNN layers with sequence learners like LSTM units. Such blends may adapt better across varying landscapes from tropics to boreal zones. Performance gains would come not from bigger datasets alone, but smarter design choices.



REFERENCES

- [1] V. Barraza, E. Roitberg and F. Grings, "Monthly deforestation forecasting through eXtreme Gradient Boosting applied to MODIS EVI time series data," 2020 IEEE ARGENCON, Resistencia, Argentina, 2020, pp. 1–5.
- [2] Z. M. Oo and T. L. Lai Thein, "Yearly forest loss measurement through Random Forest methods," 2024 IEEE Conference on Computer Applications, Yangon, 2024, pp. 1–5.
- [3] N. T. Chitra et al., "Deep learning techniques applied to satellite images for predicting deforestation," 2021 International Conference on Intelligent Computing and Control Systems, Madurai, India, 2021, pp. 522–525.
- [4] P. J. Soto et al., "Domainadversarial neural networks for spotting deforestation across tropical regions," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1–5, 2022.
- [5] S. Saha et al., "Unsupervised deep transfer learning to detect changes in highresolution multispectral imagery," IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 5, pp. 856–860, May 2021.
- [6] M. O. Adarme et al., "Deep learning methods to identify forest loss in Brazil's Amazon region," Remote Sensing, vol. 12, no. 6, p. 910, Mar. 2020.
- [7] S. P. Ingale and G. R. Bamnote, "Artificial intelligence for image captioning," in Data, Engineering and Applications, Singapore: Springer, 2022.
- [8] A. G. Mahalle, N. S. Band, and C. W. Rawarkar, "Image enhancement methods applied in the spatial domain: Hardware realization," IEJRD, vol. 4, no. 6, 2019.
- [9] A. U. Chaudhari et al., "Detection and Prevention of Malicious Activities in Vulnerable Network Security Using Deep Learning," 2024 ICICET, 2024.
- [10] K. Xu et al., "A new machine learning approach in detecting oil palm plantations using remote sensing data," Remote Sens., 2021.
- [11] T. Sboui, S. Saidi, and A. Lakti, "Forecasting deforestation driven by oil palm expansion using machine learning methods," Applied Sciences, 2023.
- [12] F. Grings, E. Roitberg, and V. Barraza, "Breakpoint detection in EVI timeseries through convolutional networks applied toward monitoring deforestation in real time," IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 2, pp. 1303–1312, 2020.
- [13] R. Malhotra et al., "Landslide Prediction using CNN with Data Modeling and Feature Reduction using Median Pooling," 2021 ICAC3N, Greater Noida, India, 2021, pp. 964–969.
- [14] A. S. Lajurkar and A. U. Chaudhari, "Application of a passive aggressive classifier for identifying misinformation," IJARCCCE, vol. 12, no. 4, pp. 767–774, 2023.
- [15] J. VillalobosMontiel et al., "Detecting regions affected by forest loss using convolutional neural networks," 2023 IEEE SusTech, Portland, Oregon, 2023, pp. 138–143.
- [16] S. Rakshit, S. Debnath, and D. Mondal, "Identifying Land Patterns from Satellite Imagery in Amazon Rainforest using Deep Learning," 2018

