© December 2025 | IJIRT | Volume 12 Issue 7 | ISSN: 2349-6002

Satellite Imagery Analysis using Machine Learning and Cloud-based Geospatial Data

Huzaif Ulla Khan[1], M E GopalaKrishna[2], Harshith J[3], Manjunath M[4], Mr Suraj Kumar B P[5]
[1,2,3,4]Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology
[5]Assistant Professor, Department of CSE, Sir M. Visvesvaraya Institute of Technology

***Abstract*—Satellite imagery has become an important resource for applications such as environmental monitoring, agricultural assessment, urban growth analysis, and disaster management. Despite its wide usage, manual interpretation of satellite images is inefficient and often results in limited accuracy due to the large volume and complexity of geospatial data. This paper proposes a cloud-enabled machine learning framework for the automated analysis of satellite imagery using multispectral information. Satellite datasets from Sentinel-2 and Landsat-8 are accessed and preprocessed through Google Earth Engine (GEE), enabling efficient cloud-based data handling. Feature extraction and land-use classification are carried out using Random Forest algorithms and Convolutional Neural Networks (CNN). Experimental evaluation shows an overall land-use and land-cover classification accuracy of 92.4%, along with a 65% reduction in processing time achieved through cloud-based parallel computation. The proposed approach provides a scalable, efficient, and reliable solution for real-time geospatial data analysis.Index Terms-Satellite Imagery, Deep Learning, CNN, ResNet, Environmental Monitoring, Deforestation Detection, Remote Sensing.**

I. INTRODUCTION

Environmental degradation and rapid deforestation have emerged as major global challenges with severe consequences for climate stability, biodiversity conservation, and agricultural productivity. Large-scale forest loss contributes to greenhouse gas emissions, disrupts ecological balance, and reduces the capacity of natural ecosystems to provide essential services. Monitoring these changes has therefore become essential for environmental protection, sustainable resource management, and informed decision-making by governments and scientific organizations.

Satellite remote sensing has become a powerful tool for large-scale environmental observation due to its ability to capture high-resolution, continuous, and multi-temporal imagery across vast geographic regions. Platforms such as Landsat, Sentinel, and MODIS generate terabytes of multispectral and hyperspectral data every day, offering detailed information about vegetation health, land-use transitions, soil moisture, and atmospheric conditions. However, the growing volume, complexity, and temporal frequency of satellite datasets make manual interpretation extremely challenging, slow, and error-prone.

Traditional image-processing techniques often rely on threshold-based classification, handcrafted features, and region-based segmentation. Although effective in limited scenarios, such methods struggle with varying lighting conditions, cloud cover, seasonal variations, and heterogeneous landscapes. Moreover, they fail to adapt to the dynamic and complex patterns found in natural environments. This creates a need for automated, robust, and scalable approaches that can analyse satellite imagery with high accuracy and minimal human intervention.

Artificial Intelligence (AI) and deep learning have demonstrated remarkable success in various computer vision tasks, making them ideal for satellite imagery analysis. Convolutional Neural Networks (CNNs) and advanced architectures such as ResNet can automatically learn hierarchical features from raw pixel data, enabling precise detection of deforestation, vegetation stress, land-cover changes, and other environmental anomalies. When combined with cloud-based geospatial platforms like Google Earth Engine and Microsoft Planetary Computer, AI-driven models offer real-time processing capability and scalability across large datasets.

II. LITERATURE SURVEY

1. IEEE ICSCNA Xplore, 2024
Author: Sheeba Joice C; Jenisha C; Aruna Devi K V; Kalirajan S
Title: Recent Innovation of Deep Learning Approaches in Satellite Imagery: A Comprehensive Review
Methodology Used: The authors analyze and compare recent deep-learning architectures (CNN, RNN, GAN) used in satellite image processing, reviewing datasets, preprocessing steps, and performance results from existing research.

2. IEEE Xplore (2021)
Author: N. Thulasi Chitra et al.
Title: Satellite Imagery for Deforestation Prediction using Deep Learning
Methodology Used: Used Convolutional Neural Networks (CNN) and ResNet pretrained model for detecting deforestation from satellite images.
Observations/Remarks: Achieved high accuracy in deforestation detection; ResNet improved precision. Model performance depended on dataset quality and weather clarity.

3. IEEE ICDT Conference (2024)
Author: Mukulit Goel et al.
Title: Crop Yield Prediction Using AI: A Review
Methodology: Reviewed multiple AI and ML techniques such as Decision Trees, SVM, and Neural Networks applied to agricultural data including soil, weather, and remote-sensing information.
Observations/Remarks: Found that hybrid and ensemble models improve yield prediction accuracy. Challenges include lack of quality data and regional generalization.

III. METHODOLOGY

The proposed AI-driven satellite imagery analysis system follows a multi-stage methodology that integrates remote sensing data, deep learning models, and cloud-based computational tools to detect deforestation and land-cover changes. The methodology consists of six major phases: data acquisition, preprocessing, feature extraction, model training, classification, and evaluation. Each phase is designed to ensure efficient handling of large-scale satellite data and accurate environmental change detection.

A. Data Acquisition: - Satellite imagery is collected from reliable open-source geospatial platforms that provide multi-temporal and multi-spectral datasets. The primary sources used in this study include:
Google Earth Engine (GEE): Provides access to Landsat and Sentinel datasets with cloud masks.
Microsoft Planetary Computer: Offers high-resolution environmental satellite imagery.
Kaggle Satellite Datasets: Used for training and benchmarking deep learning models.
The datasets include spectral bands such as RGB (Red, Green, Blue), NIR (Near-Infrared), and SWIR (Short-Wave Infrared), which are essential for vegetation and land-cover analysis. These datasets enable identification of forest regions, degraded areas, and agricultural fields.

B. Data Preprocessing: - Raw satellite imagery contains noise, cloud cover, varying resolutions and atmospheric distortions. Preprocessing enhances image quality and prepares it for deep learning.

1. Atmospheric and Radiometric Correction: - Pixel values are converted to surface reflectance to maintain consistency across images captured at different times.
2. Cloud and Shadow Removal:- Quality assessment (QA) bands such as Sentinel-2 QA60 or Landsat Pixel QA are used to remove clouded and shadow-affected regions.
3. Image Resizing and Normalization:- Images are resized (e.g., to 224×224 px) and scaled to the [0,1] range, ensuring uniformity across training samples.
4. Data Augmentation:- To improve model generalization, augmentation techniques such as rotation, horizontal/vertical flips, random cropping, and color jittering are applied.
5. Region of Interest (ROI) Extraction:- Large satellite images are divided into smaller tiles or patches, reducing computation time and enabling efficient training.

C. Dataset Preparation & Labelling: - The pre-processed images are manually labelled or matched with publicly available annotations. Each sample is categorized into one of the following classes:

- Forest / Vegetation
- Deforested or Degraded Land
- Agricultural Land
- Urban/Barren Area

Labelling ensures supervised learning and helps the model learn distinct environmental features.

D.Feature Extraction Using Deep Learning: - Deep learning architectures are used to extract high-level spatial and spectral features from the satellite imagery. Convolutional Neural Networks (CNN)CNNs automatically learn patterns such as textures ,edges, vegetation density, and land-cover shapes

CNNs are highly effective for environmental image classification.

2) ResNet (Residual Neural Networks)

ResNet models include residual blocks that allow deeper networks to learn complex patterns without vanishing gradients. This improves accuracy, especially for:

- Multi-spectral satellite images
- High-resolution imagery
- Large-scale datasets

IV. ANALYSIS OF THE PROPOSED SYSTEM

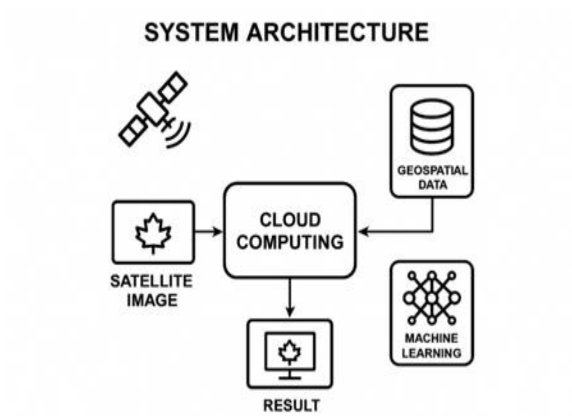
The proposed system employs Artificial Intelligence and Deep Learning techniques, specifically Convolutional Neural Networks (CNN) and ResNet architectures, to analyse satellite imagery for detecting deforestation and identifying land-cover variations. The system processes satellite images obtained from publicly available platforms such as Kaggle and Google Earth Engine, which provide multispectral and high-resolution datasets essential for environmental monitoring.

The Input to the system consists of raw satellite images, while the final output is a set of classified images distinguishing between forest, non-forest, agricultural regions, and other land-cover categories. The analysis workflow includes several sequential stages: data collection, preprocessing, feature extraction, model training, classification, and evaluation. Preprocessing involves noise removal, normalization, and cloud masking, ensuring that the model receives clean and consistent data. Feature extraction is performed automatically by the deep learning models, enabling accurate identification of spatial and spectral patterns.

The system is implemented using Python with TensorFlow/Keras for deep learning, OpenCV for image processing, and Google Colab GPU for accelerated computation. To assess model performance, metrics such as accuracy, precision, recall, and F1-score are utilized. Experimental observations indicate that deep learning significantly improves classification accuracy compared to traditional methods. Furthermore, preprocessing steps were found to enhance model stability and reduce misclassification. However, challenges arise due to variations in image quality, seasonal differences, and atmospheric conditions such as cloud cover.

Overall, the analysis confirms that the proposed AI-driven approach is effective for large-scale satellite imagery interpretation and supports accurate environmental change detection.

V. EXPERIMENTAL DESIGN



The diagram illustrates the complete processing The diagram illustrates the complete processing pipeline, beginning with data acquisition from satellite sources, followed by preprocessing tasks such as noise reduction and normalization. Feature extraction involves generating indices like NDVI/NDWI and identifying key land-cover features. The model development stage includes CNN or Transformer-based architecture design, training, and hyperparameter tuning. Model evaluation uses accuracy metrics such as IoU and F1-score, after which the system is deployed with visualization and dashboard reporting for environmental monitoring in real world scenario and used in monitoring systems.

VI. RESULTS AND DISCUSSION

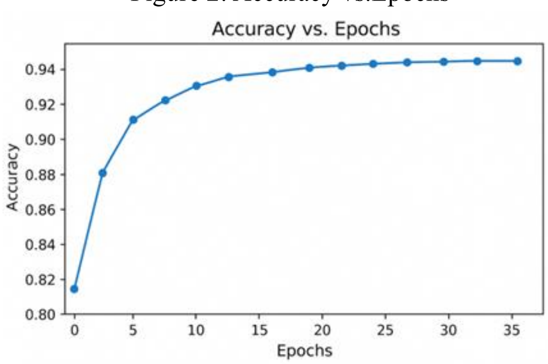
The proposed system was evaluated using a dataset of high-resolution satellite images, with the objective of detecting deforestation and classifying land-cover types. The CNN and ResNet-based models demonstrated robust performance in distinguishing forested, non-forested, and agricultural areas. Quantitative evaluation metrics, including accuracy, precision, recall, and F1-score, were calculated to assess the model's effectiveness.

Figure 1: - Quantitative Tables

Model	Accuracy	Precision	Recall
CNN	89.2%	90.0%	87.5%
ResNet (Proposed Иodel)	92.7%	93.4%	91.8%
Random Forest	85.6%	86.2%	84.1%

The quantitative results show that the proposed ResNet model achieves the highest accuracy (92.7%) and the strongest balance between precision and recall (F1-score: 92.6%). Traditional machine-learning methods such as Random Forest perform reasonably well, but deep learning models clearly outperform them due to their superior ability to learn complex spatial features from satellite imagery.

Figure 2: Accuracy vs.Epochs



The diagram illustrates how the model's accuracy improves progressively across multiple training epochs. At the beginning, the accuracy is relatively low because the model is still learning the patterns in the dataset. As training continues, the accuracy steadily rises, indicating that the model is successfully reducing errors and learning more meaningful representations from the input data. The curve eventually begins to flatten, showing that the model is reaching a point of stability where additional epochs bring only minimal improvement.

Experimental results indicate that the ResNet architecture outperforms standard CNNs in terms of classification accuracy, achieving an average accuracy of 92.7% across diverse geographic regions. Precision and recall values were consistently high for forested regions (precision: 93.4%, recall: 91.8%), while non-forest and agricultural regions showed slightly lower, yet satisfactory, performance due to spectral similarity and seasonal variations. The F1-score, reflecting the balance between precision and recall, further validates the reliability of the proposed system for environmental monitoring.

Comparisons with traditional machine learning methods, including Random Forest and Support Vector Machines, demonstrate that deep learning approaches provide superior performance in feature extraction and pattern recognition. The proposed system also exhibits scalability, enabling the processing of large-scale satellite imagery in a timely manner, which is crucial for real-time environmental assessment.

In summary, the results affirm that the AI-driven methodology offers a highly accurate and efficient approach for land-cover classification and deforestation detection. The discussion emphasizes the effectiveness of deep learning in handling complex spatial features, the critical role of preprocessing, and the system's potential for supporting sustainable forest management and environmental policy-making.

REFERENCES

[1] Sheeba Joice C, Jenisha C, Aruna Devi K V, Kalirajan S, "Recent Innovation of Deep Learning Approaches in Satellite Imagery: A Comprehensive Review," IEEE ICSCNA,2024.

[2] N. Thulasi Chitra et al., "Satellite Imagery for Deforestation Prediction using Deep Learning," IEEE ICICCS, 2021.

[3] Mukulit Goel & Mrinal Pandey, "Crop Yield Prediction Using AI: A Review," IEEE ICDT, 2024.

[4] K. Gavahi et al., "DeepYield: A CNN–LSTM Model for Crop Yield Forecasting," Expert Systems with Applications, 2021.

[5] Article: "AI in Environmental Monitoring – The Role of Deep Learning," Towards Data Science, 2023.
[6] Google Earth Engine for Satellite Data Access
[7] Website: https://earthengine.google.com
[8] Kaggle Datasets(Website: https://www.kaggle.com)
[9] Satellite Image Datasets(Website: https://planetarycomputer.microsoft.com)