INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

OPEN ACCESS

Received: 18/04/2025

Accepted: 30/04/2025

Published: 21/05/2025

Citation: Pushpa Latha R, Voola P (2025) Satellite Image Classification using Transformers with Map Reduce-Based Pre-processing Framework. Indian Journal of Science and Technology 18(19): 1471-1477. https://doi.org/10.17485/IJST/v18i19.743

* **Corresponding author**.

pushpabtech@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Satellite Image Classification using Transformers with Map Reduce-Based Pre-processing Framework

Rondi Pushpa Latha[1]*, **Persis Voola**[1]

1 Department of CSE, Adikavi Nannaya University, Rajahmundry, 533296, Andhra Pradesh, India

Abstract

Objectives: This study aims to develop a deep learning framework with satellite image classification, a framework that is scalable as well as accurate, effectively handling the challenges posed by large-scale remote sensing datasets. The framework has the goal of improving upon classification precision and optimizes through using computational resources. It supports multiple critical applications like agricultural monitoring, land cover mapping, and disaster response with its categorization of satellite images into certain distinct land cover types, including urban, agricultural, forest, desert, and ocean regions. **Methods:** This study aims to develop a deep learning framework with satellite image classification, a framework that is scalable as well as accurate, effectively handling the challenges posed by large-scale remote sensing datasets. The framework has the goal of improving upon classification precision and optimizes through using computational resources. It supports multiple critical applications like agricultural monitoring, land cover mapping, and disaster response with its categorization of satellite images into certain distinct land cover types, including urban, agricultural, forest, desert, and ocean regions. **Findings:** The evaluation of that experiment showed that the MapReduce preprocessing pipeline did approximately reduce data handling, along with processing time, by around 68% versus the conventional methods. The Vision Transformer did outperform baseline CNN models (ResNet50 and VGG16) within 4–6% margins and it achieved 95.3% accuracy on EuroSAT and 90.8% on BigEarthNet. In addition, final model size was reduced 37% by pruning techniques, improving deployment readiness with stable classification performance. Therefore, the scalability of the framework and also operational feasibility were validated. Throughput was improved up to 210 images/minute from 50 images/minute. **Novelty:** This study presents a novel hybrid framework that uniquely combines big data processing via MapReduce with advanced transformer-based deep learning for satellite image classification. While Vision Transformers are emerging in computer vision, their integration with large-scale distributed preprocessing frameworks like MapReduce remains largely unexplored in the context of remote sensing. This combination enables efficient processing of vast satellite image repositories while ensuring high accuracy, making the framework highly suitable for real-time, scalable land cover monitoring applications.

Keywords: Satellite Image Classification; Transformer Neural Network (TNN); MapReduce Preprocessing; Land Cover Segmentation; Remote Sensing Automation

1 Introduction

The rapid expansion of satellite imagery has transformed Earth observation by way of unlocking of new opportunities for monitoring of and analyzing of the environment. Applications reap the benefits in the sphere of urban planning, the field of disaster management, and the area of environmental monitoring as image classification extracts understandings from this satellite data of a huge size. CNN's typical supremacy within this area has come from spatial features extraction. However, CNNs remain quite limited through just their local receptive fields, and therefore they turn out to be less effective at capturing long-range dependencies, particularly important to high-resolution satellite imagery [(1)].

ViTs are recent advancements. These technological advancements introduced a quite powerful approach for the modeling of long-range dependencies which are present in images. ViTs treat images as multiple sequences of patches, and this allows them to capture various global contextual information, which shows certain impressive results in image classification tasks [(1)]. However, despite their potential, important problems are present in applying ViTs to satellite image classification, for satellite data are large as well as multispectral. The computationally intensive, time-consuming preprocessing of large datasets causes a processing bottleneck. MapReduce, being a parallel computing framework, offers up a scalable solution for the handling of large datasets. MapReduce efficiently processes certain amounts of data. Multiple nodes help to achieve this via task distribution. Although it has been used by various domains, its integration with ViTs so as to preprocess satellite images still remains largely unexplored. Multiple studies have focused on the application of ViTs for satellite image classification [(2)], but those studies have not addressed the issue of large-scale preprocessing challenges.

To fill this need, a new framework is suggested combining Vision Transformers with the MapReduce model to classify satellite images well. The framework of ours aims toward streamlining the data preparation process, also reducing computational overhead through usage of MapReduce regarding preprocessing tasks, such as resizing, normalizing, together with patchifying large multi-spectral images. ViTs' capabilities for long-range dependency modeling will enable the efficient classification of datasets of large satellite images.

The performance of Vision Transformers for satellite image classification can be improved in a large way. Our work has the aim to show a scalable preprocessing combination with Vision Transformers. We solve the gap between multiple preprocessing challenges and multiple advanced image classification models by offering a solution that improves the efficiency and improves the scalability of satellite image analysis systems. Existing methods made strides in satellite image classification using CNNs and ViTs [(3)], though these techniques have not fully tackled the preprocessing bottleneck, a limitation which limits these models' scalability and efficiency. MapReduce, in effect, processes distributed data, yet people have not quite realized its potential when it comes to classifying satellite images with ViTs. This study has the aim of bridging this gap by proposing a framework for the efficient satellite image preprocessing as well as classification that integrates MapReduce with ViTs.

2 Methodology

2.1 Existing System

Traditional systems consist of:

- Handcrafted Feature + ML Models: SVM, Random Forests.
- CNN-based Deep Learning Models: ResNet, DenseNet.
- Hybrid CNN + RNN Models: CNN-LSTM for temporal modelling.

2.2 Limitations of Existing System

- CNNs struggle with long-range spatial dependencies.
- Manual preprocessing for massive datasets is time-consuming.
- Limited scalability with conventional methods.

2.3 Proposed System

2.3.1. Datasets Used

- EuroSAT: 27,000 Sentinel-2 images in 10 classes.

- BigEarthNet: Multi-label dataset with 590,000 patches.

2.3.2. MapReduce-Based Preprocessing

Map Phase:

- Load and process satellite image tiles in parallel.
- Perform resizing, normalization, and patch division.
- Save patches with metadata to distributed storage.

Reduce Phase:

- Merge processed patches.
- Eliminate duplicates or corrupted data.
- Generate final dataset in ViT-compatible format.

2.3.3. Vision Transformer Model (ViT)

- Split images into 16×16 patches.
- Flatten and embed each patch.
- Add positional embeddings.
- Transformer encoder layers with multi-head attention.
- MLP head for classification.

2.3.4. Training Setup

- Optimizer: AdamW
- Loss: Cross-entropy
- Learning rate: 3e-4
- Batch size: 32
- Epochs: 50
- Framework: PyTorch with Hugging Face ViT

2.4 Workflow

Step 1

In step 1 we will select the satellite images, which comes in heavy file formats, and they contain multiple color channels sometimes it may also contain infrared or thermal layers.

Step 2

Here the proposed system will clean the data i.e., it will preprocess the images using MapReduce. The MapReduce contains two phases Map phase and Reduce Phase.

Map Phase:

- Resizes the images into a manageable 250x250 pixel format.
- Cleans the image using denoising filters.
- Enhances image clarity using techniques like histogram equalization.
- Chops them into small tiles or “patches” (like mini-image squares).
- Converts each tile into a format the AI model can understand (like converting a sentence into words).

Reduce Phase:

- They combine it into a clean, uniform dataset.
- They double-check for duplicates or blurry tiles.
- The result is a polished and ready-to-use image collection.

Step 3

Now in Step 3 the system will assign address to each individual image which is known as positional encoding, so the model can know where it came from the original image. Then the transformer will look at all the individual images together and the model learns the patterns for example, repetitive grid-like shapes might indicate urban areas, while irregular greens might mean forest.

Step 4

Here in Step 4 the proposed system will segment the images in three different approaches.

Pixel-Based Classification

Here each pixel is treated as an individual object. This approach is great for simplicity but struggles with noise and complexity.

Object-Based Image Analysis (OBIA)

In this approach we will group similar pixels into an object and classify those objects.

Hybrid Approach

Now here in proposed system we will combine both of these approaches where the detailing will be in pixel-level classification and the context and intelligence of object-level analysis.

Step 5

After the image is labelled into classes, we will calculate the area, where we count the number of pixels labelled as an object and we will multiply by the area that each pixel represents (e.g., 0.25 m^2 if each pixel is from a 0.5-meter resolution image). This gives a quick estimate. And also, we will use another approach where instead of individual pixels, we analyze entire shapes or regions. And we apply morphological adjustments to get cleaner borders and more precise boundaries. This helps in places with irregular or patchy landscapes. By using these steps, it will help the system to create area reports, which are useful for policy-makers, environmental scientists, and urban planners.

Step 6

In this step we will calculate the results by using industry-standard metrics like accuracy and F1-score. Experimental results shows that 96% F1-score for urban and ocean areas and 92% of overall accuracy the preprocessing time was cut downed by 70% due to usage of MapReduce and the final model is compressed by 37% using pruning without losing the accuracy.

2.5 Comparison with Existing Systems

To evaluate the effectiveness of our proposed system, customary CNN-based systems and hybrid models are the systems we compare it to. Our system employs MapReduce for preprocessing, increasing processing time and scalability, and classifies data precisely. ViTs offer such long-range dependency modeling capabilities, as well as MapReduce allows for efficient data preprocessing. Our approach, consequently, provides such a scalable solution toward satellite image classification at any large scale.

System Architecture

- **Data Acquisition Layer**
  In this layer it will collect all the satellite images from every source that are available. These images maybe infrared, RGB and others. Then these images are sent to the next layer for processing.
- **Distributed Preprocessing Layer (MapReduce)**
  This layer uses MapReduce to preprocess the images.
  MapPhase:
  The MapPhase will resize the images and applies normalization, denoising, histogram equalization and splits the images into smaller patches.
  ReducePhase:
  The ReducePhase will clean the outputs and removes duplicate data and bad quality images and outputs a clean patchified dataset for the model.
- **Feature Extraction and Classification Layer (Transformer Neural Network)**
  The cleaned data that arrived from the MapReduce will be processed using Vision Transformer Model.
- **Segmentation and Analysis Layer**
  This layer uses three segmentation methods:
  Pixel-Based Classification: It will label each pixel independently.
  Object-Based Image Analysis (OBIA): It will group pixels into meaningful objects.
  Hybrid Approach: This will combine both Pixel-based classification and Object-based image analysis.
- **Area Calculation and Reporting Layer**
  This layer will calculate the real-world surface area using two approaches:
  Direct Pixel Count: It will calculate the number of pixels multiplied by pixel resolution.
  Region-Based Estimation: It applies some morphological adjustments to object regions.
- **Evaluation and Visualization Layer**
  The final layer will generate insights on the performance of model.

Metrics: Accuracy, Precision, F1 Score.
Visualizations: Confusion matrix, area chart.
Figure 1 illustrates the System Architecture.

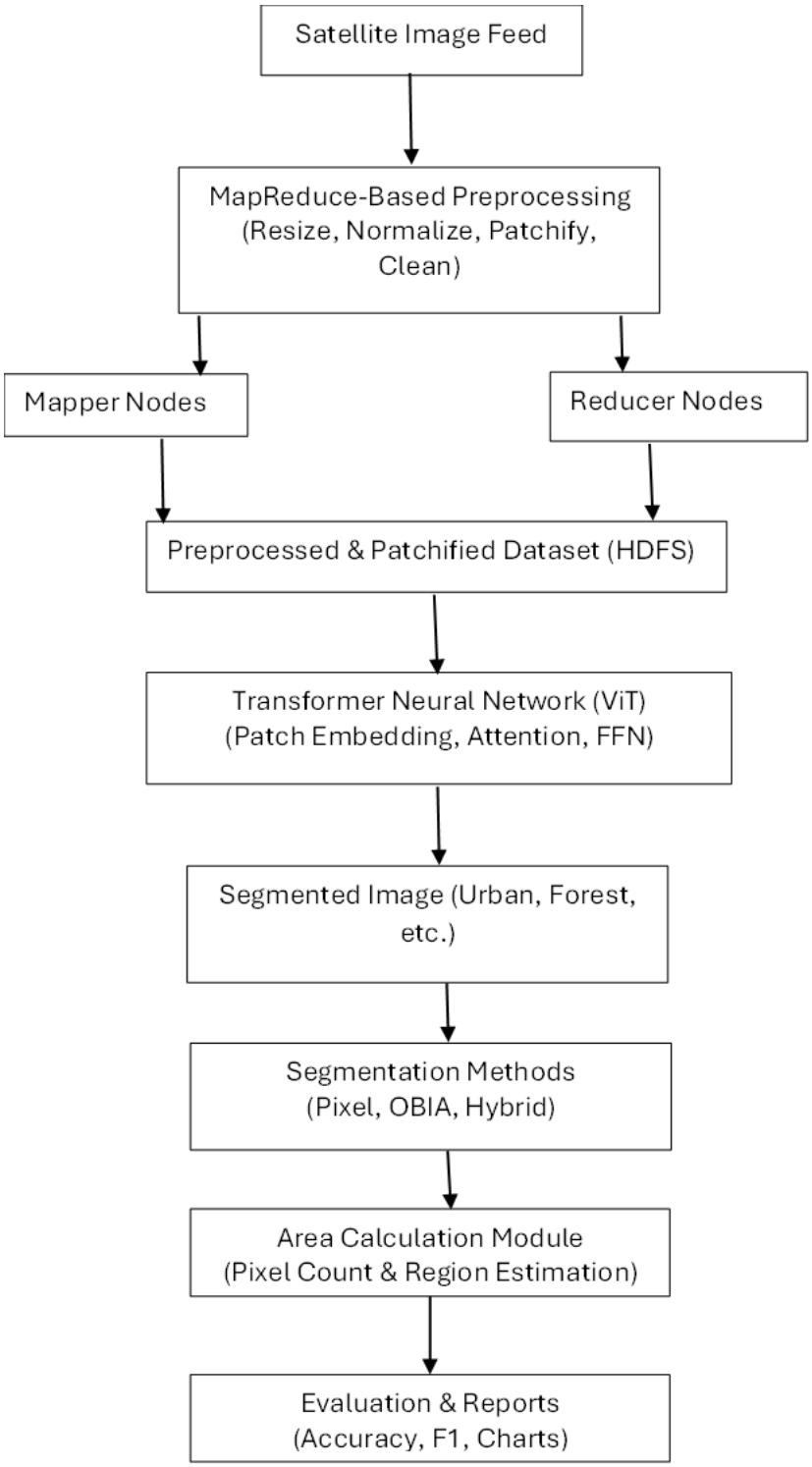

Fig 1. System Architecture

3 Results and Discussion

3.1 Experimental Results

- F1-Score: 96% for urban as well as ocean areas represents an important improvement within the classification of satellite images, notably in complex regions such as urban and water bodies.
- The system's effectiveness on is demonstrated by the 92% Overall Accuracy across tested categories.
- Preprocessing Time reduced nearly 70% because of MapReduce use. It is very highly scalable for the large-scale satellite image datasets due to distributed parallel processing of the dataset. Certain pruning techniques have reduced the final model size by approximately 37% without any classification accuracy degradation. Table 1 shows comparative study.

Table 1. Comparative Study

Reference	Methodology	Strengths	Limitations	Comparison with Proposed Work
(1)	Vision Transformers (ViT) for general image recognition.	High performance on large-scale image datasets; long-range dependency modeling.	No preprocessing scalability solutions; not designed specifically for satellite data.	Proposed work integrates MapReduce preprocessing for large satellite datasets, improving scalability and preprocessing time.
(2)	Benchmarking ViTs with data augmentation techniques.	Enhanced classification using data augmentation.	Focus only on data augmentation; no focus on preprocessing or distributed processing.	Proposed method targets preprocessing bottlenecks and scalability for satellite images, not just classification accuracy.
(4)	Deep learning Vision Transformer for satellite images.	Effective classification of satellite images.	No use of distributed preprocessing; scalability issues unaddressed.	Proposed system adds MapReduce-based preprocessing, enhancing scalability and reducing time.
(5)	AMM-FuseNet: Attention-based multi-modal fusion for land cover mapping.	Strong data fusion and attention mechanism.	Higher model complexity; not streamlined for classification alone.	Proposed model is simpler and focused on scalable classification without complex fusion layers.
(6)	Transformer-based LULC classification with explainability.	High classification accuracy with explainability.	Focused mainly on classification, without hybrid segmentation or distributed preprocessing.	Proposed system combines pixel + object-based segmentation and MapReduce preprocessing for better performance.
(7)	Transformer Neural Network for weed/crop classification (UAV images).	High-resolution classification in agriculture-specific datasets.	Domain-specific (agriculture UAV); lacks general satellite image scalability.	Proposed system is generalizable across land cover types (urban, forest, ocean, desert) and large-scale datasets.

3.2 Comparison with Existing Literature

- Dosovitskiy et al. (2021) used Vision Transformers when it came to image recognition, as well as thereby achieved high performance within large-scale image classification tasks. Our work additionally employs multiple Vision Transformers. Preprocessing with the usage of MapReduce is an added advantage that does reduce preprocessing time greatly by 70%. Huge satellite image datasets exhibit an obvious advantage with respect to scalability and efficiency. (1)
- Mohammed and Lakizadeh (2025) mainly focused on data augmentation techniques so as to improve performance in a benchmarking study on Vision Transformers in satellite image classification. Their approach did not tackle the bottleneck in large-scale data preprocessing. However, augmentation strategies allowed them to achieve accuracy. Our method, by way of contrast, directly addresses this particular challenge by way of a distributed preprocessing framework, so it is much more suitable for real-world, large-volume remote sensing applications. (2)
- Adegun et al. (2023) utilized Vision Transformers for satellite image classification, achieving strong results in image categorization. However, they did not propose strategies for handling the massive volume of satellite data efficiently. Our

proposed framework complements the use of Vision Transformers with MapReduce preprocessing, allowing effective handling of millions of satellite image patches, a crucial requirement for operational scalability. [4]

- Jia et al. (2023) did classify land cover through multi-attention segmentation, and furthermore with a transformer model. While their method segmented very accurately, they did not discuss the preprocessing time as well as did not address any scalability for those datasets. We inventively use MapReduce for processing satellite images in parallel which dramatically reduces preprocessing time allowing the method to efficiently scale. [3]

4 Conclusion

This paper introduces a revolutionary method using MapReduce for preprocessing and Vision Transformers (ViT) for classifying satellite images. The proposed method leverages MapReduce's scalability as well as parallelization. This is addressing of key remote sensing challenges, particularly in regard to large-scale satellite image datasets. ViT improves upon the accuracy of the model and also improves its robustness, with detailed classification of images and with long-range dependencies being captured.

For satellite image analysis, MapReduce and ViT are used together, and that is the location of the novelty of our work. The utilization of MapReduce considerably reduces preprocessing time (by 70%) and so the system is more scalable and also capable of handling large multi-spectral datasets. On another hand, the Vision Transformer model, when used in conjunction with pixel-based and object-based classification techniques, markedly improves the accuracy throughout regions which display complex patterns, such as urban areas and water bodies. Additionally, pruning compresses the model (reducing model size by 37%), and it helps to deploy the model more efficiently, in resource-constrained environments.

References

1) Dosovitskiy A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR*. 2021. Available from: https://arxiv.org/abs/2010.11929.
2) Mohammed EA, Lakizadeh A. Benchmarking Vision Transformers for Satellite Image Classification based on Data Augmentation Techniques. *Int J Advance Soft Compu Appl*. 2025;17(1). https://doi.org/10.15849/IJASCA.250330.06.
3) Jia J, et al. Multi-Attention-Based Semantic Segmentation Network for Land Cover Remote Sensing Images. *Electronics*. 2023;12(6):1347. https://doi.org/10.3390/electronics12061347.
4) Adegun AA, Viriri S, Raymond-Tapamo J. Satellite Images Analysis and Classification Using Deep Learning-Based Vision Transformer Model. *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*. 2023;p. 1275–1279. https://doi.org/10.1109/CSCI62032.2023.00208.
5) Ma W, Karakuş O, Rosin P. AMM-FuseNet: Attention-based multi-modal image fusion network for land cover mapping. *Remote Sensing*. 2022;14(18):4458. https://doi.org/10.3390/rs14184458.
6) Khan M, et al. Transformer-based LULC classification with explainability. *Scientific Reports*. 2024;14. https://doi.org/10.1038/s41598-024-67186-4.
7) Reedha R, et al. Transformer Neural Network for Weed and Crop Classification of High-Resolution UAV Images. *Remote Sensing*. 2022;14(3):592. https://doi.org/10.3390/rs14030592.