Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

ISSN 1989-9572

DOI:10.47750/jett.2021.12.02.026

HARNESSING AUTOMATED SATELLITE IMAGE CLASSIFICATION FOR COMPREHENSIVE ENVIRONMENTAL ASSESSMENT AND EARTH MONITORING

Saresh Kumar Ellamla[1],

Parvatham Sathish[1],

Akula Rajini[1]

Journal for Educators, Teachers and Trainers, Vol.12(2)

https://jett.labosfor.com/

Date of reception: 2 Jan 2021

Date of revision: 1 Feb 2021

Date of acceptance: 30 March 2021

Saresh Kumar Ellamla, Parvatham Sathish, Akula Rajini. (2021). HARNESSING AUTOMATED SATELLITE IMAGE CLASSIFICATION FOR COMPREHENSIVE ENVIRONMENTAL ASSESSMENT AND EARTH MONITORING. Journal for Educators, Teachers and Trainers, Vol.12(2). 202-211.

Journal for Educators, Teachers and Trainers
The PROFESIOLab electronic, peer-reviewed, open-access Magazine

Journal for Educators, Teachers and Trainers, Vol. 12(2)

ISSN 1989 –9572

https://jett.labosfor.com/

HARNESSING AUTOMATED SATELLITE IMAGE CLASSIFICATION FOR COMPREHENSIVE ENVIRONMENTAL ASSESSMENT AND EARTH MONITORING

Saresh Kumar Ellamla[1],Parvatham Sathish[1],Akula Rajini[1]
[1]Department of Electronics and Communication Engineering
[1]Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana.

ABSTRACT
Monitoring and assessing land cover changes has relied on satellite photography. Since their launch in the early 1970s, Landsat satellites have given environmental monitoring and land use planning data. Urban planning, forestry, and environmental conservation require land cover change analysis. Traditional land cover change analysis used manual satellite images interpretation or rudimentary change detection techniques. These methods may work, but they are time-consuming, subjective, and may miss subtle or complex changes. The biggest challenge is creating a Landsat-based land cover change analysis system. This requires processing and interpreting vast amounts of information over time to identify land use and land cover changes and classify them. Sustainable land management and environmental conservation need land cover monitoring. This data is essential for urban development, natural resource management, and habitat preservation decisions. Advanced methods like ensemble learning improve land cover change analysis accuracy and dependability. The research, "Analyzing Land Cover Changes with Landsat Satellite Data: An Application to Ensemble Learning," uses advanced ensemble learning to revolutionize land cover change analysis. The collective intelligence of numerous models is used to construct a system that can autonomously and reliably identify and classify land cover changes. Ensemble learning methods handle complex data and improve prediction accuracy, making them ideal for this purpose. This breakthrough could improve environmental monitoring and land management by making Landsat satellite data land cover change analysis more dependable and accurate.
Keywords: Random Forest, Image Processing, Landsat, Ensemble Learning, Urban planning

INTRODUCTION

The utilization of satellite imagery has played a pivotal role in the systematic monitoring and analysis of changes in land cover over time. Since their inception in the early 1970s, Landsat satellites have been instrumental in providing an extensive dataset for environmental monitoring and land use planning. This data is indispensable for various applications, such as urban planning, forestry, and environmental conservation.

Historically, the analysis of land cover changes relied on manual interpretation of satellite imagery or the application of basic change detection algorithms. Although these methods proved effective to a

certain extent, they were marred by drawbacks such as being time-consuming, subjective, and potentially overlooking subtle or intricate changes. Consequently, there arose a pressing need to develop a system capable of accurately analyzing land cover changes using Landsat satellite data.
The primary challenge lay in processing and interpreting large volumes of imagery over time, discerning changes in land use and land cover, and categorizing these changes into meaningful classifications. Monitoring land cover changes emerged as a critical component of sustainable land management and environmental conservation. The insights derived from such analyses became invaluable for making well-informed decisions regarding urban development, natural resource management, and habitat preservation.
To address the limitations of traditional methods, a groundbreaking project titled "Analyzing Land Cover Changes with Landsat Satellite Data: An Application to Ensemble Learning" was initiated. The objective was to revolutionize land cover change analysis by implementing advanced ensemble learning techniques. These techniques, which harness the collective intelligence of multiple models, were envisioned to autonomously and accurately identify and classify land cover changes.
Ensemble learning methods, known for their prowess in handling complex data and enhancing prediction accuracy, were deemed highly suitable for this task. The innovative approach of leveraging ensemble learning held significant promise for advancing environmental monitoring and land management efforts. By providing a more reliable and accurate means of analyzing land cover changes using Landsat satellite data, this project aimed to contribute substantially to the enhancement of environmental conservation and sustainable land management practices.

2. LITERATURE SURVEY

Many works have been done to examine the use of LULC analysis on remotely sensed records. From 1986 to 2001 in Pallisa District, Uganda, Otukei and Blaschke [3] carried out land cover mapping and land cover assessing using DTs, SVMs and MLCs. They explored the use of knowledge mining to find the required classification bands and thresholds for decision. The analysis assessed the efficiency of the classification models, claiming that land cover elements occur at an unpredictable pace. According to desired classes, a few image classification models are available for segmenting a multi-dimensional component space into homogenous regions and labelling segments. Parametric classifiers accept a normally distributed dataset and statistical parameters acquired properly from training data. The most broadly utilized parametric classifier is the maximum-likelihood classifier (MLC), which makes decision surfaces dependent on the mean and covariance of each class. MLC [11] was first applied to IRS LISS-III images between 2001 and 2011 and classified into eight classes. Additionally, the study used a unique methodological framework for post-classification adjustments. It considerably increased total classification accuracy from 67.84% to 82.75% in 2001 and from 71.93% to 87.43% in 2011.
Islam et al. [1] used Landsat TM and Landsat 8 OLI/TIRS images to examine land use changes in Chunati Wildlife Sanctuary (CWS) from 2005 to 2015. ArcGIS and ERDAS imagine were used for land use change assessment. To derive supervised land use categorization, the maximum likelihood classification technique was applied. It was discovered that around 256 ha of the degraded forest area has increased over ten years (2005–2015), with an annual rate of change of 25.56%. Non-parametric classifiers do not accept a particular information appropriation to isolate a multi-dimensional feature space into classes. Most normally utilized non-parametric classifiers incorporate decision trees [4], support vector machines (SVM) [12] and expert systems. ML algorithms have been utilized according to pixel classifiers in remote sensing image analysis [6].
Grippa et al. [13] describes a method for mapping urban land use at the street block level, emphasizing residential usage by utilizing very-high-resolution satellite images and derived land-cover maps as input. For the classification of street blocks, a random forest (RF) classifier is utilized, which achieves accuracies of 84% and 79% for five and six land-use classifications, respectively. RF classifier applied over urban communities Dakar and Ouagadougou, cover more than 1,000 km2 altogether, with a spatial resolution of 0.5 m.

3. PROPOSED SYSTEM

3.1Overview

Here is the overview description of the landcover changes with landsat satelite:

- Uploading Dataset:Users upload their dataset by clicking the "Upload Dataset" button.
- Upon clicking the button, a file dialog window appear, allowing users to navigate to and select the dataset folder containing subfolders for different classes of satellite images.Once the dataset is uploaded, a confirmation message displayed on the GUI.
- Image Preprocessing: After the dataset is uploaded, the "Image preprocessing" button clicked to initiate image processing.The application utilize the VGG16 model to extract features from the satellite images in the dataset.Extracted features be saved along with their corresponding labels.The dataset be split into training and testing sets for model training and evaluation.
- Training and Testing Existing Logistic Regression Model:Upon clicking the "Build & Train Logistic Regression Model" button, the application train a logistic regression model using the preprocessed dataset.The trained model be saved to a file for future use.The model's performance could be evaluated using metrics such as accuracy, precision, recall, and F1-score.The evaluation results may be displayed on the GUI, along with a confusion matrix and classification report.
- Training and Testing Proposed RFC Model:Clicking the "Build & Train Ensemble Learning Model" button may trigger the training of a Random Forest Classifier (RFC) model.The RFC model might be trained using the preprocessed dataset.After training, the model's performance may be evaluated using similar metrics as for the logistic regression model.Evaluation results, including accuracy, precision, recall, F1-score, confusion matrix, and classification report, may be displayed on the GUI.
- Models Evaluation Graphs:Upon clicking the "Performance Evaluation" button, the application generate comparison graphs for evaluating the performance of both models.
- Graphs display metrics such as accuracy, precision, recall, and F1-score for each model.Users visually compare the performance of the logistic regression and RFC models through these graphs.
- Test Image Prediction Using Proposed RFC Model:Users upload a test image by clicking the "Upload test image" button.After selecting an image, the application use the trained RFC model to make predictions on the uploaded image.Predicted class labelsbe displayed on the image or in a separate window, indicating the land cover changes identified by the model.

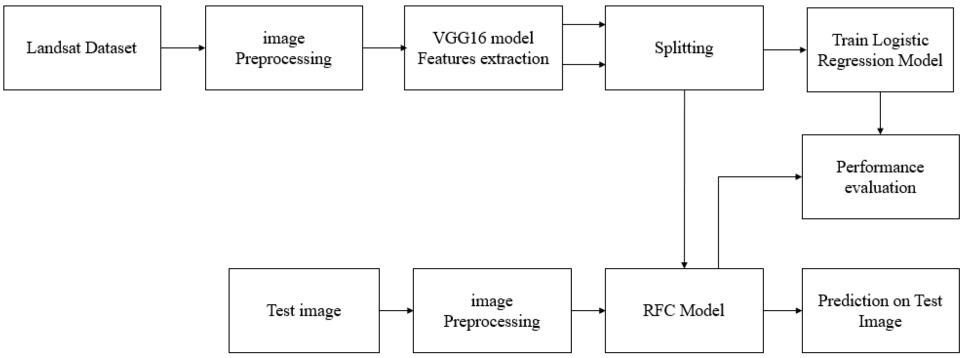


Figure1: Block diagram of Proposed System.

3.2 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

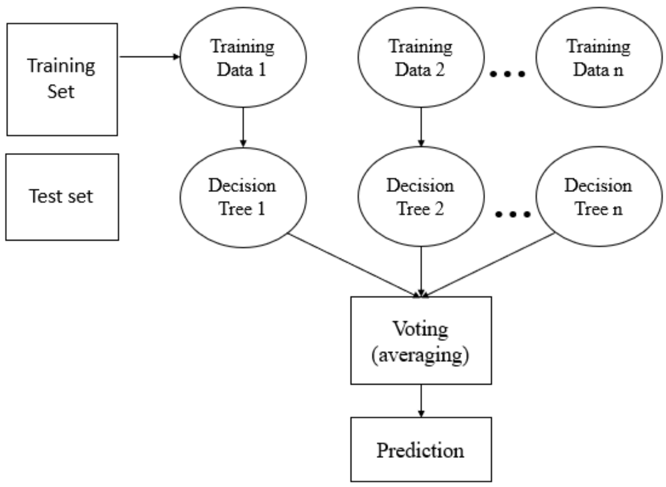


Fig. 2: Random Forest algorithm.

Random Forest algorithm

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Important Features of Random Forest

- **Diversity-** Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- **Immunetothecurseofdimensionality-** Since each tree does not consider all the features, the feature space is reduced.
- **Parallelization-**Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build random forests.
- **Train-Testsplit-** In a random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- **Stability-** Stability arises because the result is based on majority voting/ averaging.

Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.

Below are some points that explain why we should use the Random Forest algorithm

- It takes less training time as compared to other algorithms.
- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

Types of Ensembles

Before understanding the working of the random forest, we must look into the ensemble technique. Ensemble simply means combining multiple models. Thus, a collection of models is used to make predictions rather than an individual model. Ensemble uses two types of methods:

Bagging– It creates a different training subset from sample training data with replacement & the final output is based on majority voting. For example, Random Forest. Bagging, also known as Bootstrap Aggregation is the ensemble technique used by random forest. Bagging chooses a random sample from the data set. Hence each model is generated from the samples (Bootstrap Samples) provided by the Original Data with replacement known as row sampling. This step of row sampling with replacement is called bootstrap. Now each model is trained independently which generates results.

The final output is based on majority voting after combining the results of all models. This step which involves combining all the results and generating output based on majority voting is known as aggregation.

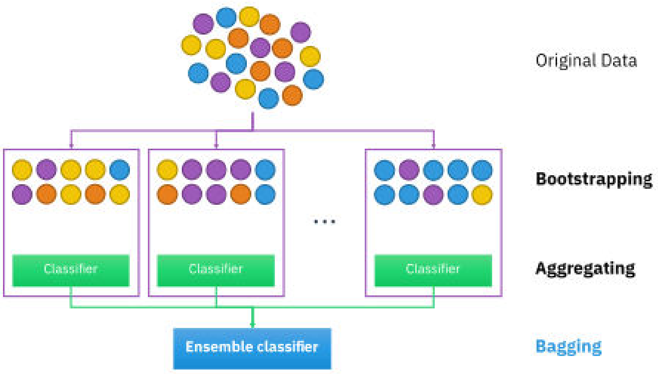


Fig. 3.RF Classifier analysis.

Boosting– It combines weak learners into strong learners by creating sequential models such that the final model has the highest accuracy. For example, ADA BOOST, XG BOOST.

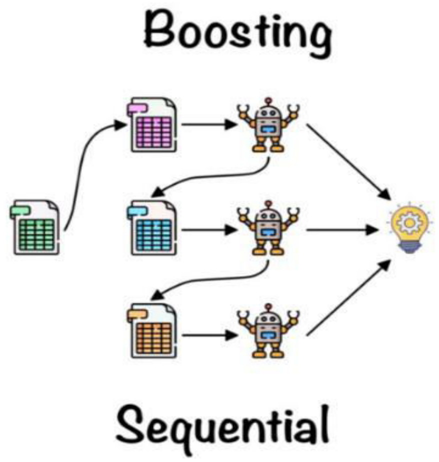


Fig.4: Boosting RF Classifier.

4. RESULTS

Figure 5 a graphical user interface (GUI) designed for analyzing land cover changes using Landsat satellite data. It includes interactive elements for data visualization and analysis. Figure 6 is the dataset uploading process is illustrated, indicating how users can import Landsat satellite data into the GUI for analysis. This step is crucial for accessing the dataset and preparing it for further processing. Figure 7: Displaying the dataset preprocessing and data splitting steps, this figure demonstrates the necessary transformations applied to the Landsat satellite data to enhance its quality and usability. Preprocessing involve normalization, feature scaling, and splitting the data into training and testing sets.

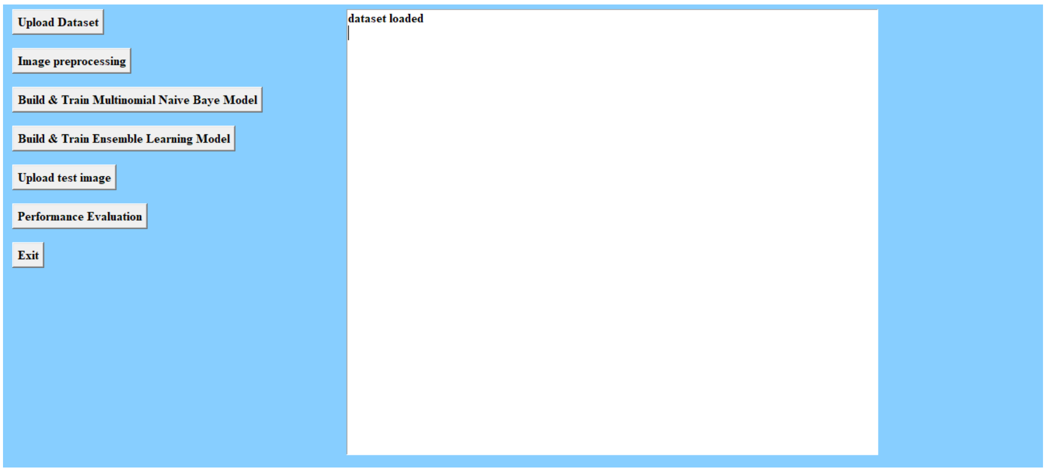


Figure 5: Displays the GUI of Monitoring Environmental assessment.

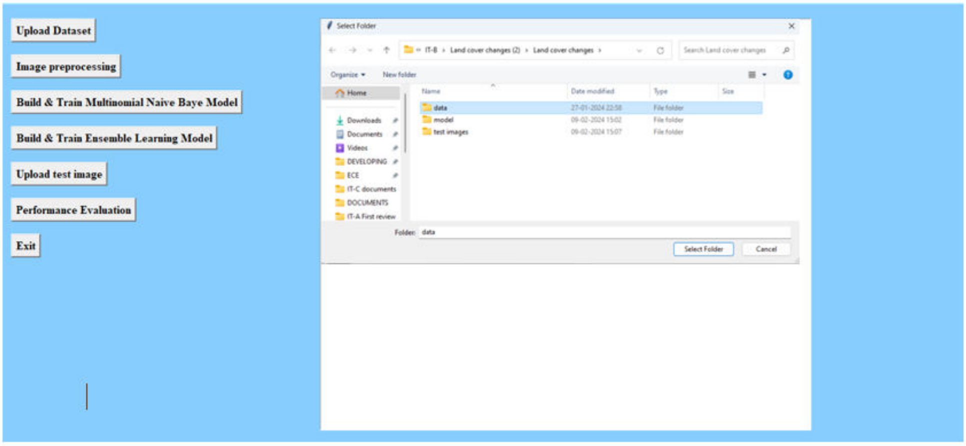


Figure 6: Displays the uploading of dataset.

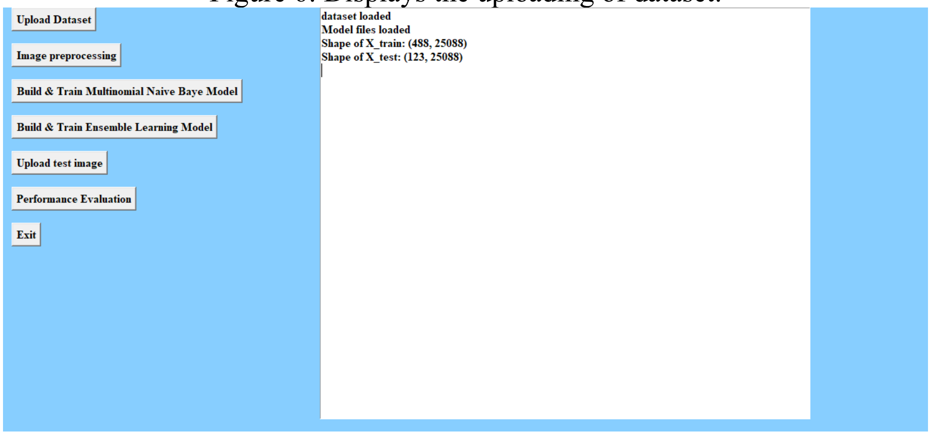


Figure 7: Displays the dataset preprocessing and data splitting.

Figure 8 is a confusion matrix for both the Ensemble model and Logistic Regression model. These matrices provide insights into the performance of each model by showing the counts of true positive, true negative, false positive, and false negative predictions.

Figure 9 is a performance comparison count plot, depicting various evaluation metrics such as accuracy, precision, recall, and F1-score for each model. The plot allows users to visually compare the performance of different models and select the most effective one for their analysis. the proposed Ensemble model's predictions on test images are illustrated. Users can observe the model's classifications of land cover changes based on Landsat satellite data, providing valuable insights into environmental changes over time.

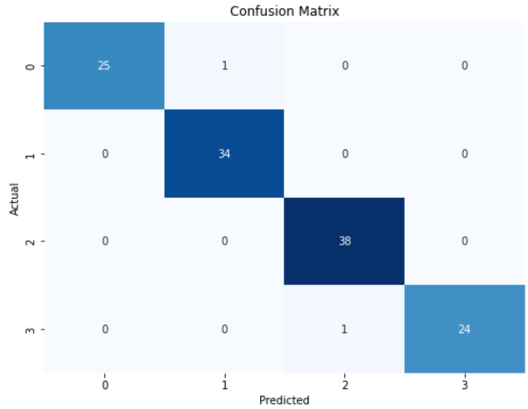


Figure 8: Confusion matrix of Ensemble model.

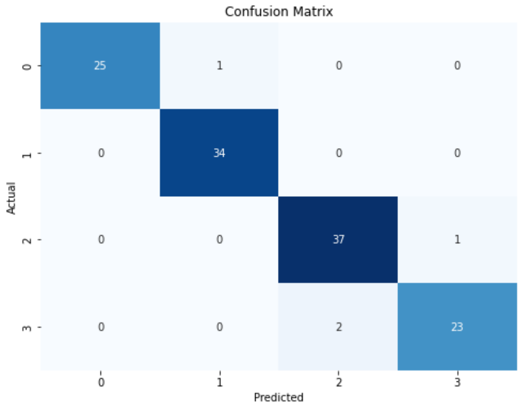


Figure 9: Confusion matrix of Logistic Regression model.

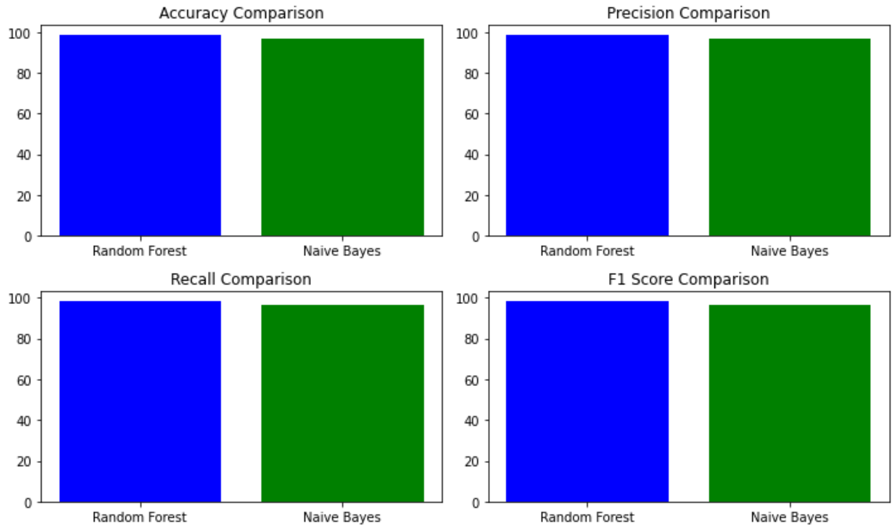
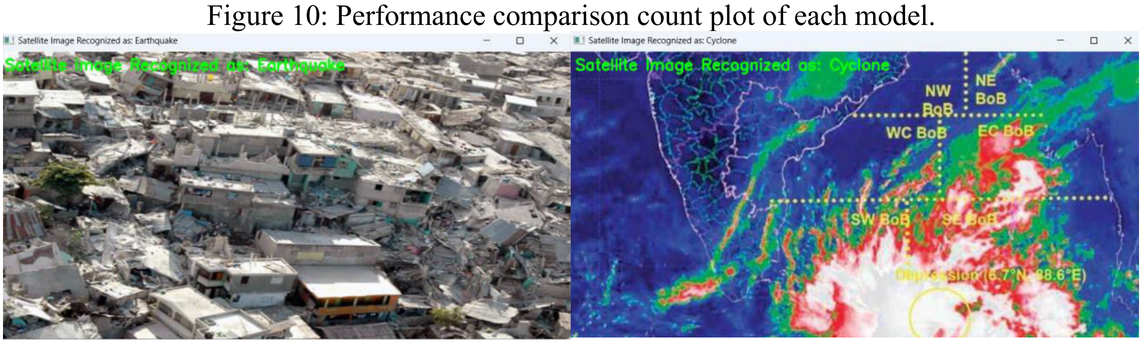


Figure 10: Performance comparison count plot of each model.



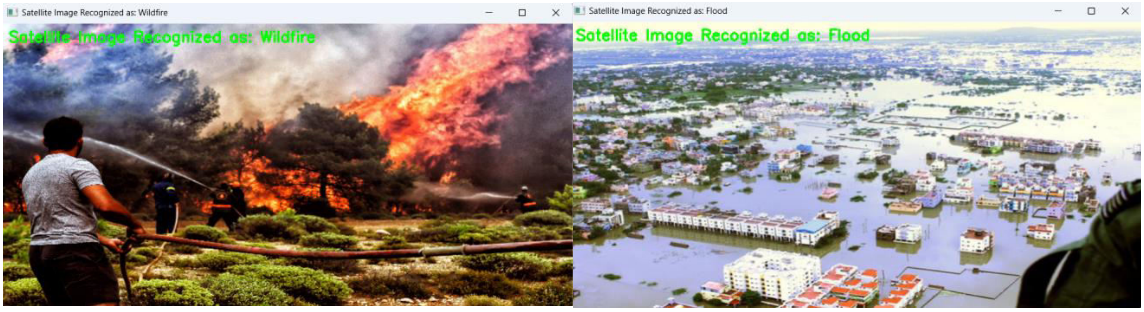


Figure 11: Proposed Ensemble model prediction on test images.

Table 1: Performance comparison of quality metrics by ML models.

Model	Logistic regression	Ensemble Classifier model
Accuracy (%)	96	98
Precision (%)	96	98
Recall (%)	96	98
F1-score (%)	96	98

For the Logistic regression model:

- The Accuracy is 96, indicating the accuracy between the actual and predicted values
- The Precision is 96, suggesting that, on average Precision between the actual and predicted values.
- The Recall is 96, suggesting that, on average Recall between the actual and predicted values.
- The F1-score is 96, representing the average F1-score between the actual and predicted values

For the Ensemble Classifier model:

- The Accuracy is 98, indicating the accuracy between the actual and predicted values.
- The Precision is 98, suggesting that, on average Precision between the actual and predicted values.
- The Recall is 98, suggesting that, on average Recall between the actual and predicted values.
- The F1-score is 98, representing the average F1-score between the actual and predicted values.

5. CONCLUSION

In conclusion, the utilization of Landsat satellite data has been instrumental in monitoring and analyzing changes in land cover over the decades, contributing significantly to environmental monitoring, land use planning, and conservation efforts. The conventional methods of manual interpretation and basic change detection algorithms, while effective to some degree, have inherent limitations such as being time-consuming, subjective, and potentially overlooking subtle or intricate changes.

Recognizing the pivotal role of accurate land cover change analysis in critical domains like urban planning, forestry, and environmental conservation, there is a pressing need to advance existing methodologies. The primary challenge lies in the development of a sophisticated system capable of autonomously processing and interpreting large volumes of Landsat satellite imagery, identifying nuanced changes in land use and cover, and classifying them into meaningful categories.

The proposed reserch represents a pioneering effort to revolutionize land cover change analysis. By leveraging advanced ensemble learning techniques, which harness the collective intelligence of multiple models, the research aims to enhance the accuracy and reliability of identifying and classifying land cover changes. Ensemble learning, known for its proficiency in handling complex data and improving prediction accuracy, emerges as a promising solution to address the challenges inherent in traditional methods.

The outcomes of this research hold great promise for the field, as the application of ensemble learning is poised to provide a more reliable and accurate means of analyzing land cover changes using Landsat satellite data. This advancement is not only crucial for sustainable land management but also for making well-informed decisions regarding urban development, natural resource management, and habitat preservation. In essence, the project signifies a significant step towards advancing environmental monitoring and land management efforts through cutting-edge technologies and methodologies.

REFERENCE

[1] K. Islam, M. Jashimuddin, B. Nath, and T. K. Nath, "Land use classification and change detection by using multi-temporal remotely sensed imagery: The case of Chunati wildlife sanctuary, Bangladesh," The Egyptian Journal of Remote Sensing and Space Science, vol. 21, no. 1, pp. 37–47, Apr. 2018, doi: 10.1016/j.ejrs.2016.12.005.

[2] C. Kok Yang, F. Pei Shan, and T. Lea Tien, "Climate change detection in Penang Island using deterministic interpolation methods," Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), vol. 19, no. 1, pp. 412–419, Jul. 2020, doi: 10.11591/ijeecs.v19.i1.pp412-419.

[3] J. R. Otukei and T. Blaschke, "Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms," International Journal of Applied Earth Observation and Geoinformation, vol. 12, pp. S27– S31, Feb. 2010, doi: 10.1016/j.jag.2009.11.002.

[4] L. Hua, X. Zhang, X. Chen, K. Yin, and L. Tang, "A feature-based approach of decision tree classification to map time series urban land use and land cover with Landsat 5 TM and Landsat 8 OLI in a Coastal City, China," ISPRS International Journal of Geo-Information, vol. 6, no. 11, pp. 1–18, Oct. 2017, doi: 10.3390/ijgi6110331.

[5] I. A. Md Zin, Z. Ibrahim, D. Isa, S. Aliman, N. Sabri, and N. N. A. Mangshor, "Herbal plant recognition using deep convolutional neural network," Bulletin of Electrical Engineering and Informatics (BEEI), vol. 9, no. 5, pp. 2198–2205, Oct. 2020, doi: 10.11591/eei.v9i5.2250.

[6] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," Remote Sensing of Environment, vol. 117, pp. 34–49, Feb. 2012, doi: 10.1016/j.rse.2011.02.030.

[7] A. Jamali, "Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran," SN Applied Sciences, vol. 1, no. 11, Nov. 2019, Art. no. 1448, doi: 10.1007/s42452- 019-1527-8.

[8] D. K. Behera, M. Das, S. Swetanisha, and P. K. Sethy, "Hybrid model for movie recommendation system using content K-nearest neighbors and restricted Boltzmann machine," Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), vol. 23, no. 1, pp. 445–452, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp445-452.

[9] A. A. W. Mohammed and H. T. Khamees, "Categorizing and measurement satellite image processing of fire in the forest greece using remote sensing," Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), vol. 21, no. 2, pp. 846–853, Feb. 2021, doi: 10.11591/ijeecs.v21.i2.pp846-853.

[10] Z. Deng, X. Zhu, Q. He, and L. Tang, "Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area," Advances in Space Research, vol. 63, no. 7, pp. 2144–2154, Apr. 2019, doi: 10.1016/j.asr.2018.12.005.

[11] A. K. Thakkar, V. R. Desai, A. Patel, and M. B. Potdar, "Post-classification corrections in improving the classification of land use/land cover of arid region using RS and GIS: The case of Arjuni watershed, Gujarat, India," The Egyptian Journal of Remote Sensing and Space Science, vol. 20, no. 1, pp. 79–89, Jun. 2017, doi: 10.1016/j.ejrs.2016.11.006.

[12] A. V. Rajeswari, S. Saritha, and G. Santhosh Kumar, "Classification based land use/land cover change detection through Landsat images," in 2014 International Conference on Data Science & Engineering (ICDSE), Aug. 2014, pp. 232–237, doi: 10.1109/ICDSE.2014.6974644.

[13] T. Grippa et al., "Mapping urban land use at street block level using OpenStreetMap, remote sensing data, and spatial metrics," ISPRS International Journal of Geo-Information, vol. 7, no. 7, Jun. 2018, Art. no. 246, doi: 10.3390/ijgi7070246.