AI-Powered Earth Vital Sign Monitoring: Enhancing Disaster Prediction with CNN-VGG16 and Random Forest Integration

Y. Vasantha[1], D. Satheesh[2], P. Venkatesh[3]

[1]Department of Electronics and Communication Engineering

[1]Sree Dattha Institute of Engineering and Science, Sheriguda, Hyderabad, Telangana

ABSTRACT

Monitoring earth vital signs is necessary to assess earth health and ensure human safety. The current system uses the Multinomial Naïve Bayes algorithm to detect Earth's vital signals, such as earthquakes, cyclones, floods, and wildfires, through image analysis. The algorithm's premise of feature independence may limit its capacity to capture complex image dataset relationships, despite its simplicity. These systems may suffer with environmental change. Earth monitoring uses the Naive Bayes model to analyze earthquake, cyclone, flood, and wildfire data. However, the Naive Bayes model's assumption of feature independence may limit its ability to capture complicated interactions in Earth's diverse and dynamic vital sign datasets. CNN using VGG16 and Random Forest is proposed. The CNN-based VGG model extracts features from the input image, preprocesses, trains, and tests the data, while the Random Forest method predicts accuracy and labels. Earth vital signs offers real-time monitoring, predictive analytics, and accuracy. Flexible tool for environmental surveillance and early warning systems for environmental risks and natural disasters.

Keywords: Earth Vital Signs Monitoring, Environmental Surveillance, Natural Disaster Prediction, CNN-VGG16 Feature Extraction, Random Forest Classification.

1. INTRODUCTION

1.1 Overview

Monitoring the Earth's vital signs, encompassing seismic activities, cyclones, floods, and wildfires, has become increasingly critical for disaster preparedness and response. Leveraging ensemble learning models in this context provides a comprehensive and sophisticated approach to understanding and predicting these dynamic phenomena. Ensemble learning involves the integration of predictions from multiple models, offering improved accuracy and robustness. In the case of earthquakes, ensemble models can analyze seismic data from various sources to enhance early detection and prediction capabilities. This is crucial for minimizing the impact on communities and infrastructure. For cyclones and hurricanes, ensemble learning excels in analyzing diverse climate and atmospheric data. By combining predictions from models specialized in different aspects of storm behavior, ensemble models provide more accurate forecasts of the storm's path, intensity, and potential impact areas. In flood monitoring, ensemble learning processes information from sources such as rainfall data, river gauges, and topographical information. The diversity of models within the ensemble allows for a more nuanced understanding of complex hydrological processes, aiding in timely flood predictions and risk assessments. Ensemble learning is also valuable in monitoring and predicting wildfires. By analyzing data from satellites, weather patterns, and historical fire incidents, ensemble models can provide more accurate predictions of wildfire occurrence, spread, and potential areas at risk. The adaptability and real-time processing capabilities of ensemble learning models contribute to effective disaster management. These models can quickly adjust to new data, making them suitable for monitoring the evolving nature of natural disasters. Additionally, ensemble learning helps in assessing uncertainties associated with predictions, aiding decision-makers in making informed choices in the face of potential disasters. In summary, applying ensemble learning models to monitor Earth's vital signs related to earthquakes, cyclones, floods, and wildfires enhances prediction accuracy, adaptability, and resilience. This technological approach plays a crucial role in advancing our capability to respond effectively to natural disasters and mitigate their impact on both human and environmental systems.

1.2 Problem Statement

Monitoring the Earth's vital signs, including seismic activities, cyclones, floods, and wildfires, poses various challenges that can be addressed through the application of ensemble learning models. One major problem lies in the complexity of the data associated with these natural phenomena. Traditional methods struggle to capture the intricate patterns and correlations within large and diverse datasets generated by seismic sensors, meteorological instruments, and satellite imagery. Ensemble learning, with its ability to handle multiple models and diverse data sources, addresses this challenge by providing a more comprehensive analysis of the intricate interactions among Earth's vital signs.Another significant problem pertains to the dynamic and evolving nature of these natural disasters. Earthquakes, cyclones, floods, and wildfires exhibit varying patterns over time, and traditional models may struggle to adapt swiftly to changing conditions. Ensemble learning models, designed to be adaptable and responsive, offer a solution by continuously updating predictions based on real-time data. This enhances the accuracy of early warnings and predictions, crucial for effective disaster preparedness and response.Uncertainty in predictions is a persistent challenge in monitoring Earth's vital signs. Ensemble learning models, by providing estimates of uncertainty through the combination of multiple predictions, contribute to better risk assessment. Understanding the confidence levels associated with predictions is vital for decision-makers, allowing them to make informed choices regarding evacuation plans, resource allocation, and disaster response strategies.Data integration and feature selection pose additional challenges in the monitoring of Earth's vital signs. Ensemble learning addresses these issues by efficiently combining predictions from models that specialize in different aspects of the data. This not only aids in identifying crucial features contributing to each type of natural disaster but also ensures a more robust and accurate overall prediction.In summary, the application of ensemble learning models to monitor Earth's vital signs tackles key challenges related to data complexity, dynamic patterns, uncertainty, and efficient feature selection. By doing so, it significantly enhances our ability to understand, predict, and respond to seismic activities, cyclones, floods, and wildfires, ultimately contributing to more effective disaster management and environmental stewardship.

1.3 Research Motivation

The motivation for researching the monitoring of Earth's vital signs, encompassing seismic activities, cyclones, floods, and wildfires, through the lens of ensemble learning models stems from the critical need for more accurate, adaptive, and timely predictions in the face of escalating environmental challenges. Traditional methods often struggle to capture the intricate patterns and interactions within the vast and dynamic datasets associated with these natural phenomena. Ensemble learning, with its ability to integrate diverse models and data sources, offers a promising solution to enhance the precision of predictions and improve our understanding of the complex relationships between different vital signs.

The dynamic and evolving nature of natural disasters poses a significant motivation for exploring ensemble learning. Earthquakes, cyclones, floods, and wildfires exhibit temporal variations and can rapidly change in intensity and behavior. Ensemble learning models, designed for adaptability, can continuously learn from new data, providing real-time insights that are crucial for effective disaster preparedness and response. The motivation here lies in developing systems that can rapidly adjust predictions to changing conditions, ensuring that the monitoring frameworks remain robust and reliable over time.

Furthermore, the motivation extends to addressing the inherent uncertainty associated with predictions in the realm of Earth's vital signs. Ensemble learning models inherently provide estimates of uncertainty by combining predictions from multiple models. This aspect is crucial for decision-makers, enabling them to assess the reliability of predictions and make informed choices regarding resource allocation, evacuation plans, and disaster response strategies.

Data integration and efficient feature selection represent additional motivations for employing ensemble learning models. The ability of these models to handle diverse data sources and select relevant features ensures a more comprehensive understanding of the factors contributing to each type of natural disaster. This not only enhances the accuracy of predictions but also provides valuable insights into the underlying processes driving Earth's vital signs.

In essence, the research motivation lies in the quest for more effective, adaptable, and reliable tools to monitor Earth's vital signs. Ensemble learning models present a promising avenue for advancing our capabilities in understanding, predicting, and responding to seismic activities, cyclones, floods, and wildfires. By addressing the complexities associated with these vital signs, researchers aim to contribute to the development of innovative solutions that can mitigate the impact of natural disasters and foster sustainable environmental 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.

Jamali [7] compared and contrasted eight machine learning methods for image categorization in the northern region of Iran developed in the Waikato environment for knowledge analysis (WEKA) and R programming languages. Machine learning models [14]–[16] such as RF, SVM [17], [18], decision tree, K-nearest-neighbors (KNN) [19], principal component analysis (PCA) [20] are successfully applied in many application areas. We have built up an ensemble model [21], including SVM and XGBoost [22], that gives better precision when contrasted with other individual machine learning models.

3 PROPOSED SYSTEM

3.1 Overview

Here is the overview description of the leveraging vital signs classification for disaster management and environment risk assessment to safeguard ecosystem:

- 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 labels be displayed on the image or in a separate window, indicating the land cover changes identified by the model.

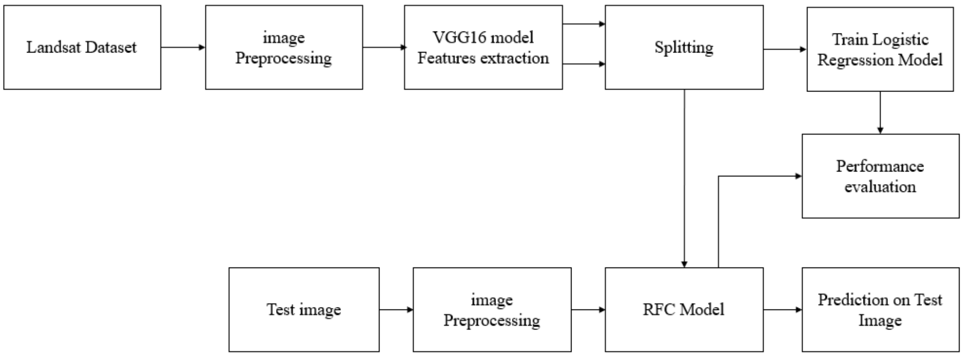

Figure 3.1: Block diagram of Proposed System.

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.

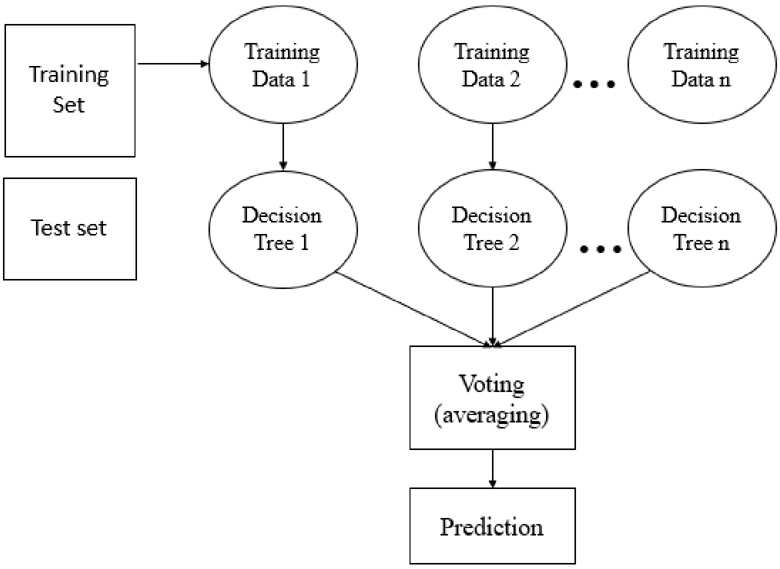

Fig. 3.2: Random Forest algorithm.

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.
- **Immune to the curse of dimensionality-** 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-Test split-** 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.

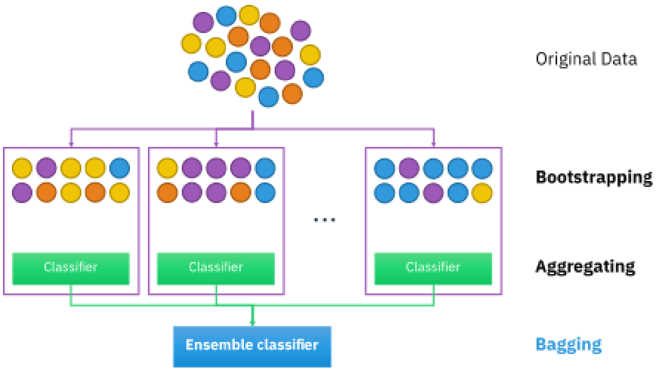

Fig. 3.3: RF Classifier analysis.

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.

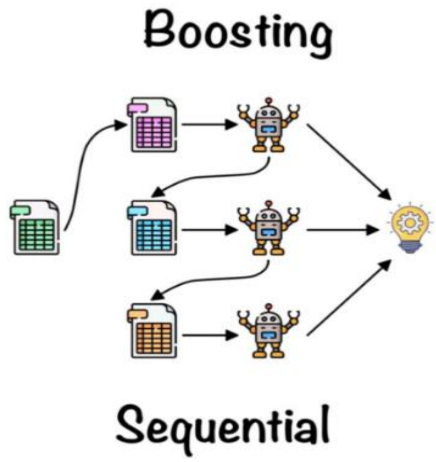

Fig. 3.4: Boosting RF Classifier.

4. RESULTS AND DISCUSSION

Figure 1: This figure showcases the graphical user interface (GUI) designed for analyzing land cover changes using Landsat satellite data. It includes interactive elements for data visualization and analysis. Figure 2: Here, 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 3: 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.

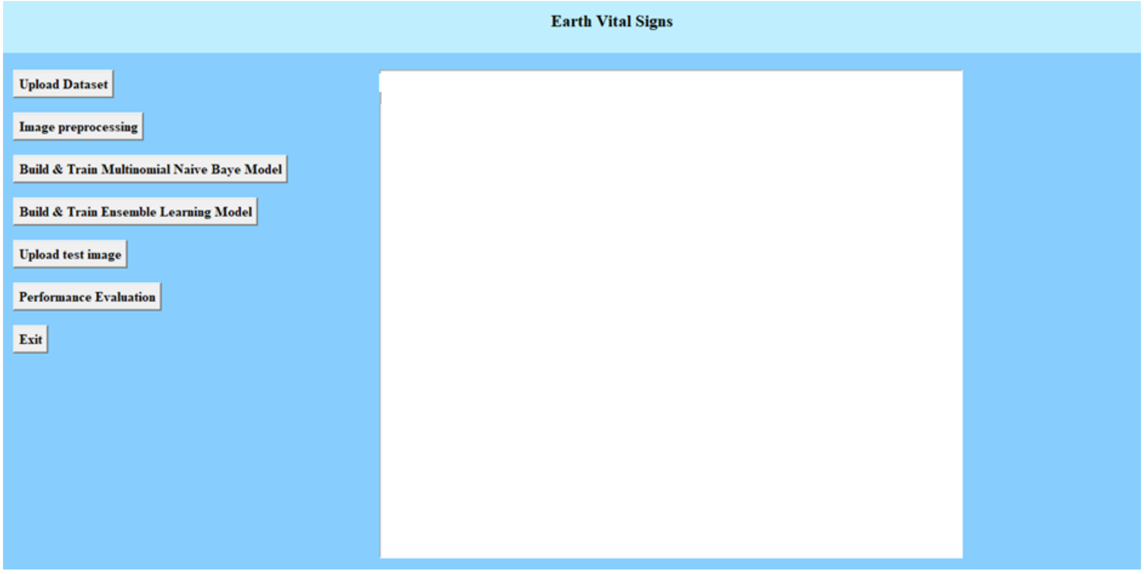

Figure 1: Displays the GUI of land cover changes with landsat satellite.

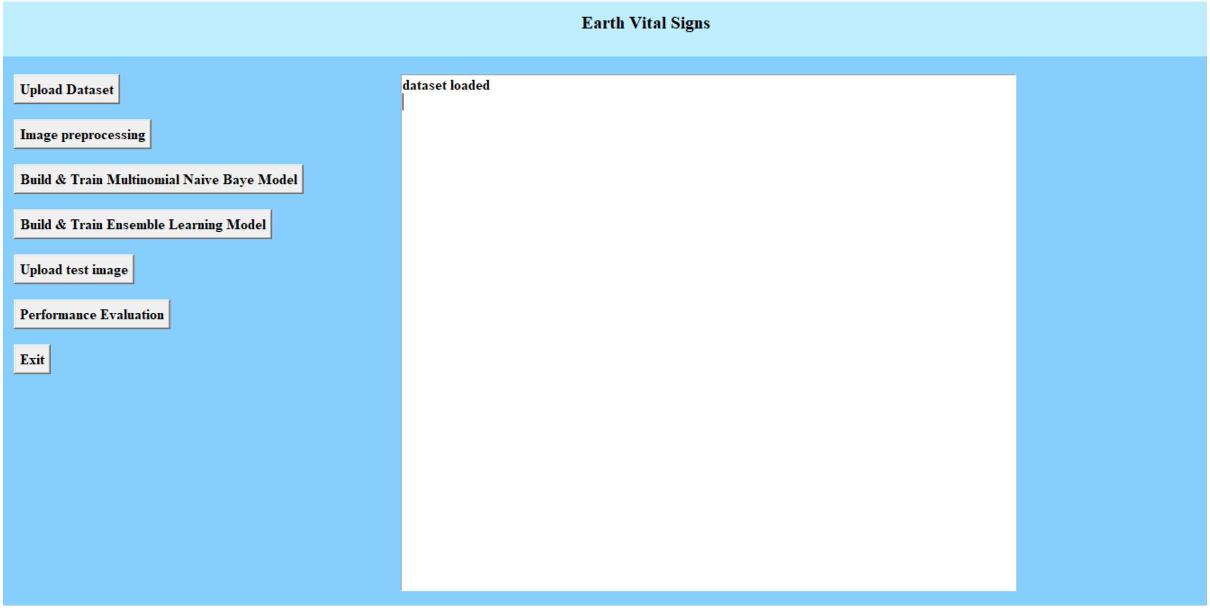

Figure 2: Displays the uploading of dataset.

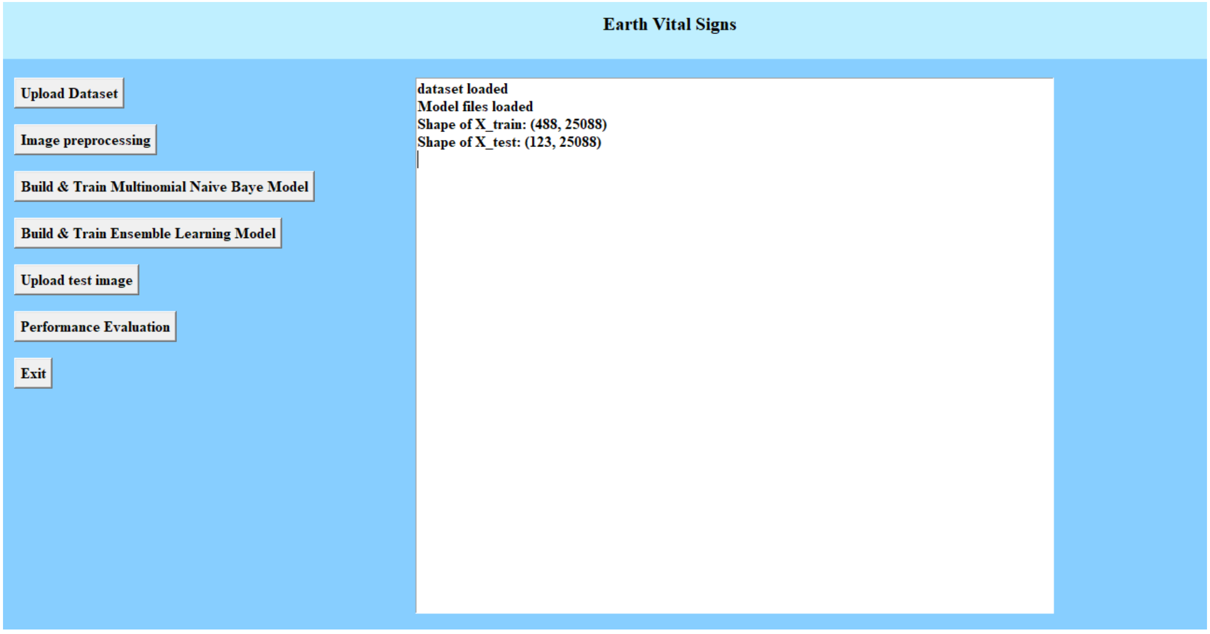

Figure 3: Displays the dataset preprocessing and data spliting.

Figure 4: Presented here are the confusion matrices 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 5: This figure presents 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. Figure 6: Here, 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.

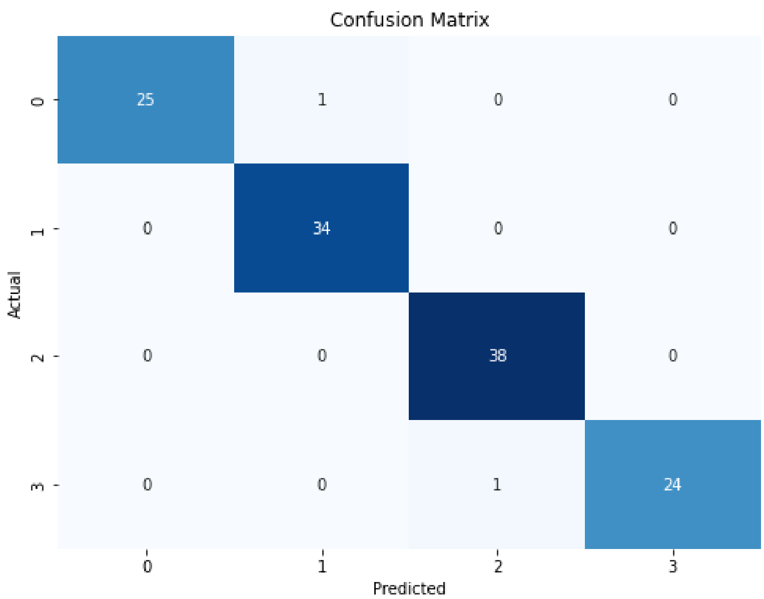

Figure 4: Confusion matrix of Ensemble model.

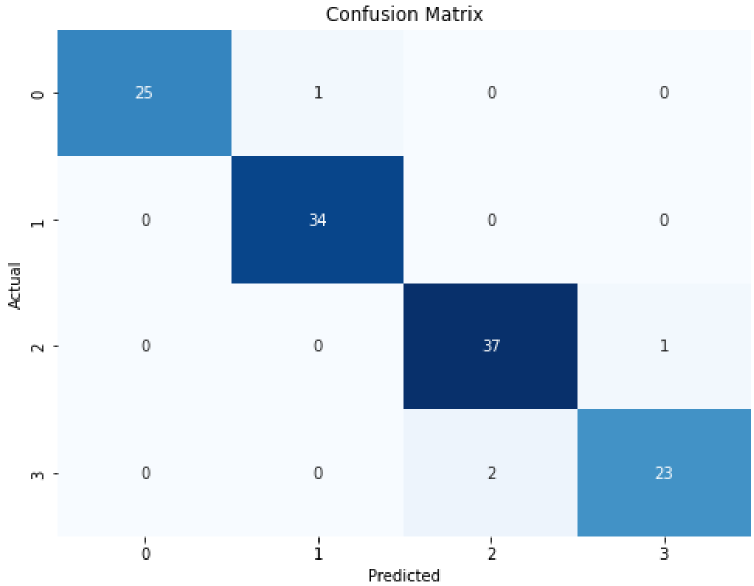

Figure 4: Confusion matrix of Logistic Regression model.

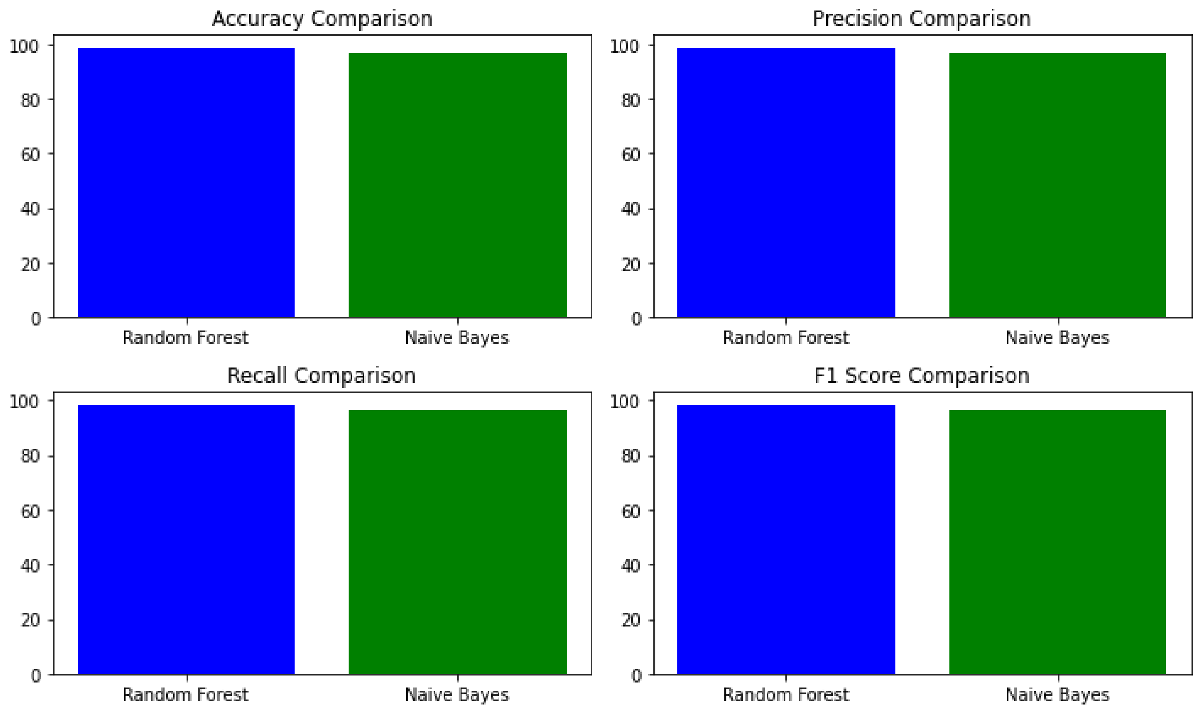

Figure 5: Performance comparison count plot of each model.

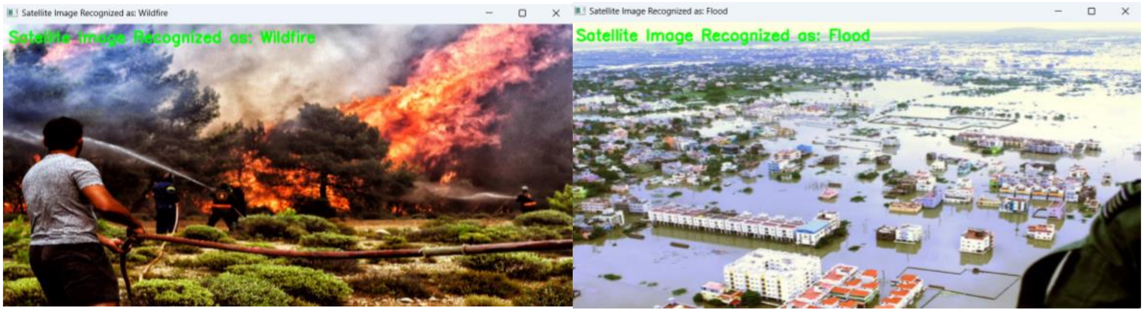

Figure 6: Proposed Ensemble model prediction on test images.

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

5. CONCLUSION

In conclusion, the current state of monitoring Earth's vital signs relies on traditional methods that may lack real-time analysis and predictive capabilities. The use of the Naive Bayes model, while simple and efficient, may face limitations in capturing complex relationships within the diverse and dynamic datasets associated with seismic activities, cyclones, floods, and wildfires. To address these limitations, the proposed system integrates a Convolutional Neural Network (CNN) with the VGG16 model and the Random Forest algorithm as an ensemble learning model. This hybrid approach enhances the system's ability to extract features from input images and make accurate predictions for Earth's vital signs.

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.

[14] W. Hassan, T.-S. Chou, O. Tamer, J. Pickard, P. Appiah-Kubi, and L. Pagliari, "Cloud computing survey on services, enhancements and challenges in the era of machine learning and data science," International Journal of Informatics and Communication Technology (IJ-ICT), vol. 9, no. 2, pp. 117–139, Aug. 2020, doi: 10.11591/ijict.v9i2.pp117-139.

[15] S. M. Praveena, R. Kanmani, and A. K. Kavitha, "A neuro fuzzy image fusion using block based feature level method," International Journal of Informatics and Communication

Technology (IJ-ICT), vol. 9, no. 3, pp. 195–204, Dec. 2020, doi: 10.11591/ijict.v9i3.pp195-204.

[16] T. R. S. Mary and S. Sebastian, "Predicting heart ailment in patients with varying number of features using data mining techniques," International Journal of Informatics and Communication Technology (IJ-ICT), vol. 8, no. 1, pp. 56–62, Apr. 2019, doi: 10.11591/ijict.v8i1.pp56-62.

[17] N. Kulkarni, "Support vector machine based alzheimer's disease diagnosis using synchrony features," International Journal of Informatics and Communication Technology (IJ-ICT), vol. 9, no. 1, pp. 57–62, Apr. 2020, doi: 10.11591/ijict.v9i1.pp57-62.

[18] A. M. Abdu, M. M. Mokji, and U. U. Sheikh, "Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 9, no. 4, pp. 670–683, 2020, doi: 10.11591/ijai.v9.i4.pp670-683.

[19] H. Ohmaid, S. Eddarouich, A. Bourouhou, and M. Timouyas, "Comparison between svm and knn classifiers for iris recognition using a new unsupervised neural approach in segmentation," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 9, no. 3, pp. 429–438, 2020, doi: 10.11591/ijai.v9.i3.pp429-438.

[20] S. M. Shaharudin, N. Ahmad, and S. M. C. M. Nor, "A modified correlation in principal component analysis for torrential rainfall patterns identification," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 9, no. 4, pp. 655–661, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp655-661.