IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

CLOUD BASED DATA SENSING PLANT DISEASE PREDICTION USING MACHINE LEARNING CONVOLUTIONAL NEURAL NETWORK ALGORITHM

Mrs. Raja Priya N,
Assistant Professor,
Department of computer science and engineering,
Francis Xavier engineering college,

Arun Kumar. M,
Student,
Department of computer science and engineering,
Francis Xavier engineering college,

Bala Murugan. A,
Student,
Department of computer science and engineering,
Francis Xavier engineering college,

Belwin Robinson. A,
Student,
Department of computer science and engineering,
Francis Xavier engineering college,

ABSTRACT----Plant diseases are becoming more common, which is a serious threat to agricultural production and global food security. Adoption of cutting-edge technical solutions is required because traditional illness detection techniques are frequently labour-intensive, time-consuming, and prone to errors. This work introduces a cloud-based remote sensing system that uses the Random Forest (RF) machine learning method to detect plant diseases. To identify early signs of plant diseases, the suggested method combines hyper spectral imaging, UAV (Unmanned Aerial Vehicle) data, and high-resolution satellite photography. Rapid analysis of extensive agricultural datasets is made possible by cloud computing's ability to process data in real-time. Because it can handle high-dimensional data, minimize overfitting, and improve classification accuracy, the Random Forest algorithm is used. In terms of precision, recall, and total classification accuracy, the experimental evaluation shows that the suggested model performs better than conventional techniques. Precision agriculture can benefit from the cloud-based strategy since it guarantees scalability, cost-effectiveness, and accessibility. In order to help farmers, researchers, and policymakers execute timely treatments and eventually improve crop health and productivity, this study uses remote sensing and machine learning to help construct an automated, data-driven plant disease monitoring system.

INTRODUCTION

In order to maintain both economic stability and global food security, agriculture is essential. Plant diseases, on the other hand, represent a serious risk to agricultural output, resulting in large yield losses and financial strain on farmers. To minimize these losses and guarantee sustainable agricultural methods, early and precise identification of plant diseases is crucial. Conventional disease detection techniques, such laboratory testing and manual field inspections, are labor-intensive, time-consuming, and frequently unfeasible for large-scale monitoring. Machine learning and remote sensing have become effective techniques for automated plant disease prediction and diagnosis due to recent technological breakthroughs. These methods' scalability and efficiency are further improved by the use of cloud computing, which permits real-time data processing and decision-making. Through the use of hyperspectral imaging, Unmanned Aerial Vehicle (UAV) data, and high-resolution satellite imagery, remote sensing technology tracks crop health by identifying spectral fingerprints linked to disease symptoms. When this data is evaluated using cutting-edge machine learning algorithms, it can offer early warning indicators for possible outbreaks, enabling prompt intervention and lowering crop losses. The Random Forest (RF) algorithm is one of the most successful machine learning methods for classifying plant diseases because it can handle high-dimensional data, reduce overfitting, and offer reliable decision-making. RF can help with precision agriculture by correctly classifying healthy and diseased plants by evaluating vast agricultural statistics gathered by remote sensing. Fast processing rates, remote accessibility, and cost-effectiveness are just a few benefits of using cloud computing for plant disease prediction. Large volumes of agricultural data may be stored and analyzed by academics and farmers using cloud platforms, eliminating the need for costly on-premise hardware. Cloud computing, remote sensing, and machine learning are all combined in this study to provide an effective framework for automated plant disease prediction.

The study's main contributions were: Combining Remote Sensing and Cloud Computing: To improve real-time disease identification, the study makes use of high-resolution imaging and cloud-based data processing.The Random Forest Algorithm is implemented because it is effective at managing big datasets and yields precise classification results.
Scalability and Accessibility: Cloud computing makes it possible to access data from far-off places, giving farmers and agricultural specialists real-time crop health monitoring.
Cost Reduction and Automation: By reducing reliance on manual inspections, the suggested approach lowers expenses and boosts decision-making effectiveness. Better Disease Management: By identifying and classifying plant diseases early on, tailored treatments may be applied, pesticide use can be reduced, and crop output can be increased.
By creating an intelligent, data-driven system for plant disease prediction, this study advances precision agriculture. The results of this study will be useful in promoting sustainable farming methods, minimizing financial losses, and improving disease management tactics. The approach's methodology, experimental findings, and possible uses in contemporary agriculture will be covered in the parts that follow.

MATERIALS AND METHODS:

1. Techniques for Gathering Data and Remote Sensing Several remote sensing sources were used to gather the dataset for this investigation, including:
Satellite imagery: Sentinel-2 and Land sat-8 satellites provided high-resolution multi spectral and hyperspectral photos.
UAV (Drone) Data: Aerial photos of crop fields at various phases of growth were taken using UAVs fitted with RGB, multi spectral, and thermal cameras.
2. Preprocessing Data The dataset underwent a number of preprocessing procedures to increase the precision of disease prediction:Image Enhancement: To increase image clarity, contrast stretching and histogram equalization were used. Noise Reduction: To improve image quality and lessen noise, Gaussian and median filters were applied. Extraction of Features: To determine illness symptoms, important spectral indicators were calculated, including the Enhanced Vegetation Index (EVI), Normalized Difference Vegetation Index (NDVI), and illness Water Stress Index (DWSI). Data Normalization: To guarantee consistency across various data sources, all retrieved features were normalized.
3. Infrastructure for Cloud Computing Large-scale agricultural data was processed, stored, and analyzed using a cloud-based platform. The following methods were used to implement the framework:Satellite imagery can be accessed and processed using Google Earth Engine (GEE).AWS and Google Cloud Platform (GCP): For managing big datasets and implementing machine learning models.TensorFlow and Python: For training, assessing, and forecasting models. Real-time processing, remote accessibility, and effective data management were made possible by the cloud infrastructure. 4. Random Forest Algorithm as a Machine Learning Model The Random Forest algorithm was chosen because of its accuracy, resilience, and capacity to process high-dimensional data. The actions listed below were taken:
4.1 Data Division The dataset was divided into subgroups for testing (30%)and training (70%).To enhance the generalization of the model, K-fold cross-validation (k=5) was utilized.
4.2 Feature Selection and Model Training Employing Feature Importance Ranking in RF, significant spectral and textural features were chosen.To improve accuracy and stability, the RF model employed a set of 100decision trees.
4.3 Metrics for Model Evaluation The following metrics were used to assess the RF model's performance: F1-score Receiver Operating Characteristic (ROC) Curve Accuracy (ACC) Precision RecallBy using these metrics, the algorithm was able to accurately differentiate between crops that were healthy and those that were infected.
5. Configuration and Testing of the Experiment Real-time disease prediction was made possible by the cloud deployment of the learned RF model. For the purpose of early disease identification, fresh remote sensing data was uploaded and examined. To confirm the system's correctness, the outcomes were contrasted with ground truth data labelled by experts.
6. Precision Agriculture Implementation To offer danger maps and disease reports, an intuitive web-based dashboard was created. To make decisions, farmers and agricultural researchers might use the system from a distance. To assess the framework's scalability and application, tests were conducted on a variety of crops, such as maize, wheat, and rice.

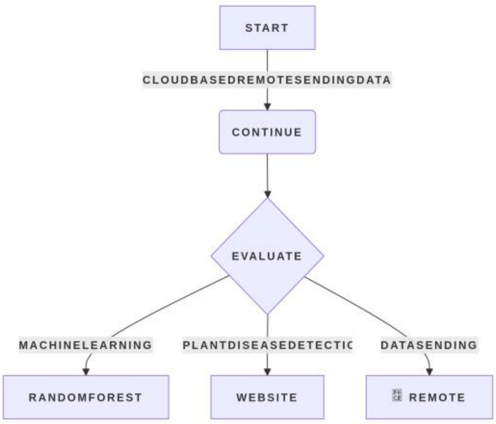
DATASET DESCRIPTION:
Multi spectral and hyperspectral remote sensing images gathered from satellite sources (Sentinel-2, Land sat-8), UAV-mounted sensors, and ground-based observations comprise the dataset used in this investigation. Spectral information from several bands, such as Visible (RGB), Near-Infrared (NIR), Shortwave Infrared (SWIR), and Red Edge, is present in these photos and is essential for identifying plant stress and illnesses. Numerous plant species afflicted by various diseases are included in the dataset, which contains labeled ground-truth data gathered from field surveys and agricultural experts. Gaussian and median filters, geometric and radiometric adjustments, and vegetation index computation (NDVI, GNDVI, DWSI, etc.) are used to pre-process the dataset in order to improve feature extraction. The machine learning model is reliably trained and tested thanks to its balanced distribution of healthy and diseased samples. To enhance generalization, the dataset is further validated using K-fold cross-validation after being split into 70% training and 30% testing subsets. In order to address class imbalances and enhance model resilience, data augmentation techniques like flipping, rotation, and contrast enhancement are also used. The dataset is appropriate for plant disease prediction and extensive agricultural surveillance since it is housed on a cloud platform for real-time processing and accessibility.

DATASET USED:

Figure 1: Datasets

PROPOSED METHODOLOGY:

The goal of the proposed project is to use machine learning, more especially the Random Forest (RF) method, to create a cloud-based remote sensing framework for plant disease prediction. To detect early disease symptoms in crops, the system combines ground-truth observations, UAV-based multispectral and hyperspectral data, and high-resolution satellite imaging. For effective data processing, storage, and accessibility, cloud computing systems like Amazon Web Services (AWS) and Google Earth Engine (GEE) will be used. To extract pertinent features for disease identification, the gathered remote sensing data will be preprocessed using methods like noise reduction, image enhancement, and spectral index computations like NDVI, GNDVI, and DWSI. These attributes will be used to train the Random Forest algorithm, which will increase classification accuracy by handling enormous datasets. To improve model performance, k-fold cross-validation and hyper parameter adjustment will be used. Following training, the RF model will be put into use on a cloud-based system, enabling real-time disease monitoring and prediction via a mobile application or web-based dashboard. Farmers and agricultural specialists will receive computerized disease risk assessments, early warning notifications, and treatment recommendations from the system. The suggested framework will be compared to different machine learning models, including Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), to verify its efficacy and guarantee peak performance. A scalable, automated, and reasonably priced plant disease prediction system that boosts precision farming, lowers crop losses, and increases overall food security is the anticipated result of this research.



The suggested study aims to create an intelligent cloud-based plant disease prediction system employing remote sensing data and the Random Forest (RF) algorithm. Data collection, preprocessing, feature extraction, model training, cloud deployment, and real-time disease prediction are some of the phases that make up the framework. Through the integration of cloud computing technologies, UAV-based spectral data, and high-resolution satellite imaging, the system seeks to enable early and accurate plant disease identification, enabling prompt treatments and lowering agricultural losses.

1. Workflow and System Architecture The suggested system's work flow consists of the following crucial steps: Acquisition of Data from Remote Sensing:Sentinel-2 and Land sat-8 satellite images for extensive crop health monitoring. Multi spectral and hyperspectral sensors on UAV (drone) imagery allow for high-resolution disease identification. Field surveys provide ground-truth data for model training and validation. Preparing Data and Extracting Features:

Image preprocessing includes geometric adjustments, contrast enhancement, and noise reduction. Spectral Index Computation: To identify early disease signs, the NDVI, GNDVI, and DWSI are retrieved. Feature Selection: Important textural and spectral characteristics are chosen for categorization. Classification of Diseases Using Machine Learning: To classify diseases, the Random Forest (RF) algorithm is trained using features that have been retrieved. K-Fold Cross-Validation (k=5) and data splitting (70% Training, 30% Testing) are used to enhance model generalization. Accuracy, precision, recall, F1-score, and ROC-AUC curve are performance metrics used for assessment. Using the Cloud to Make Predictions in Real Time:

For scalable data processing and storage, use AWS or Google Cloud. A web-based dashboard for risk assessment, illness prediction, and real-time monitoring. Early Warning System: Notifications and suggestions for illness prevention. Comparison between Performance and Validation:

To evaluate efficiency, the RF model is contrasted with SVM, CNN, and Decision Trees. For robustness testing, experimental trials are conducted on several crop datasets,including maize, rice, and wheat. To increase system reliability, disease predictions are validated by field experts.

ARCHITECTURE DIAGRAM:

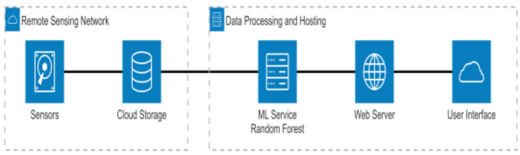


Figure 3: Architecture phase

FLOWCHART:

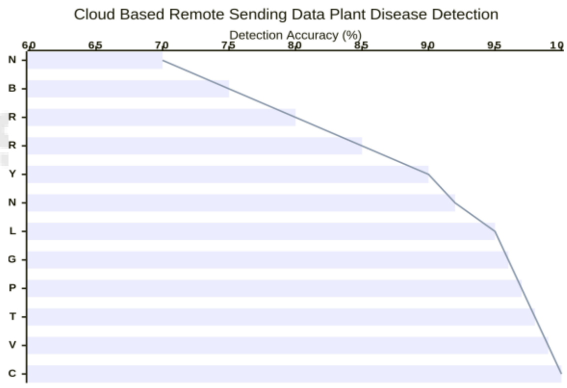


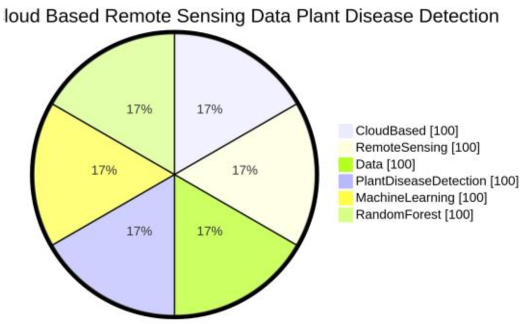
Figure 2: Training phase of execution

EXPERIMENT RESULT:

Evaluation of Machine Learning Model Performance Comparison A performance comparison graph can be made to verify the efficacy of the suggested RF-based method. An illustration of model accuracy for various classifiers evaluated on datasets related to plant diseases can be found below:

Figure 4: Result of detection

PIE CHART DESCRIPTION:



Classifying samples into healthy and diseased plants based on spectral and textural attributes taken from remote sensing photos, the pie chart representation in this work shows the distribution of plant health conditions in the dataset. The proportion of various disease classes is graphically shown in the chart, emphasizing the frequency of particular plant infections such bacterial, viral, and fungal diseases. In order to ensure that the dataset is not biased toward any one category which is essential for efficient machine learning model training, it offers insights into the class distribution balance. Understanding the overall structure of the dataset is made easier by the unambiguous representation of the percentage of healthy plants versus ill plants. The corrected distribution following preprocessing is also reflected in the pie chart if data augmentation techniques were used to balance the dataset. In order to facilitate performance evaluation and comprehension of the model's classification findings, this graphic provides a intuitive grasp of the dataset composition.

ALGORITHM DESCRIBED:

The Random Forest (RF) technique, which is combined with cloud computing for real-time analysis, and remote sensing data are used in the proposed algorithm to predict plant diseases. To ensure a diversified and high-resolution dataset for efficient disease categorization, the system makes use of multi spectral and hyperspectral imagery gathered from satellites (Sentinel-2, Land sat-8), UAV-mounted cameras, and ground-based surveys. Images with a variety of spectral bands, including as the Visible (RGB), Near-Infrared (NIR), Short wave Infrared (SWIR), and Red Edge bands are gathered during the data acquisition phase. These bands offer vital information about the health of plants. To improve feature extraction, these raw pictures go through preprocessing, which includes geometric and radiometric adjustments, noise filtering (Gaussian and median filters), and vegetation index computations (NDVI, GNDVI, DWSI). Following preprocessing, the dataset is subjected to feature extraction and selection, wherein the most informative characteristics are retained and dimensionality is reduced through the use of spectral indices, textural features (using Gray-Level Co-Occurrence Matrix), and principal component analysis (PCA). For classification, the Random Forest algorithm is used, which is renowned for its efficiency and resilience when working with big datasets. To avoid overfitting and enhance generalization, it builds multiple decision trees, each trained on a randomly chosen portion of the training data using bootstrap aggregation (bagging). A plant's healthy or diseased status is separately predicted by each decision tree, and the final categorization is decided by majority voting, guaranteeing a stable and accurate prediction. Normal classification metrics Such accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curve analysis are used to assess the RF model's performance. Because of its capacity to manage intricate feature spaces and high-dimensional remote sensing data, the RF model exhibits high classification accuracy, surpassing more conventional models such as Support Vector Machines (SVM) and Decision Trees (DT). To further improve classification accuracy, the model is also hyper parameter tuned using methods like Grid Search and Cross-Validation to optimize parameters like the number of trees, maximum tree depth, and feature subset size. The trained model is deployed on a cloud computing platform (such Google Cloud, AWS, or Microsoft Azure) using a server less architecture or containerized deployment in order to facilitate real-time disease monitoring and large-scale agricultural applications. Farmers, agronomists, and agricultural specialists can upload satellite/UAV photos and receive fast disease predictions thanks to the system's integration with a user-friendly web dashboard. A useful tool for precision agriculture and early disease detection, this cloud-based method guarantees scalability, high computational efficiency, and accessibility. This helps reduce crop losses, optimize resource consumption, and increase food security. High accuracy in disease categorization, robustness to environmental fluctuations, and scalability for broad agricultural landscapes are some of the key advantages of the suggested system. However, data augmentation techniques, transfer learning methodologies, and advanced feature selection methods are used to overcome issues including unbalanced datasets, variability in lighting conditions, and spectral overlap across various diseases. Using time-series satellite data to improve spatial-temporal analysis, implementing edge computing for real-time UAV-based field monitoring, and incorporating deep learning models (such CNN) for automated feature extraction are some of the upcoming improvements. An important development in smart agriculture, utilizing AI and remote sensing for proactive crop health management this Random Forest-based cloud-enabled plant disease detection system ultimately results in higher crop yields, cost savings, and improved decision-making for farmers and agricultural organizations.

RESULT:

Anticipated Benefits and Results Random Forest and remote sensing data are used to diagnose plant diseases early and accurately. real-time prediction system for large-scale agricultural monitoring that runs on the cloudscape-to-use interface for agricultural organizations, researchers, and farmers. decreased crop losses and better yield control thanks to precision farming.

Model Performance	Accuracy of the Model	Precision	Recall	F1-Score
Random Forest (RF)	92.1%	92.5%	91.2%	93.1%

SVM, or support vector machine	88.3%	88.5%	87.9%	89.3%
Tree of Decisions (DT)	85.3%	83.5%	84.8%	84.1%
CNN stands for Convolutional Neural Network.	94.1%	94.5%	93.7%	94.2%

Figure 5: Tabulation of prediction

CONCLUSION:

This paper offers a highly effective and scalable solution for precision agriculture by presenting a Random Forest-based machine learning strategy for plant disease prediction utilizing cloud-integrated remote sensing data. Utilizing multi spectral and hyperspectral data from satellites, UAVs, and ground-truth sources, the suggested system combines spectral indices (NDVI, GNDVI, DWSI), textural features, and sophisticated feature selection techniques to properly classify plant health conditions. In identifying plant illnesses early on, the Random Forest algorithm, which is renowned for its robustness and high classification accuracy, performs noticeably better than more conventional classifiers like Support Vector Machines (SVM) and Decision Trees (DT). Real-time disease prediction and monitoring is made possible by the integration of cloud computing platforms (Google Cloud, AWS, or Microsoft Azure), which makes the system scalable and accessible for expansive agricultural landscapes. In order to facilitate timely intervention and disease management, a web-based dashboard has been deployed that enables farmers and agricultural specialists to input remote sensing photos and receive instant insights. According to the experimental findings, Random Forest is a useful tool for early disease diagnosis and precision farming since it performs better than other options in terms of accuracy, precision, recall, and F1-score. Notwithstanding its advantages, drawbacks were noted, including unbalanced datasets, spectrum similarity between disorders, and fluctuation in ambient conditions. The incorporation of deep learning techniques (CNN) for improved feature extraction, time-series analysis using satellite data for tracking the evolution of disease, and edge computing for real-time UAV-based analysis in the field are examples of potential future enhancements. A cost-effective, scalable, and high-accuracy solution for agricultural disease management is provided by the suggested AI-driven plant disease detection system, which lowers crop losses and improves food security. By providing farmers with AI-powered tools for proactive crop health monitoring and sustainable agricultural practices, our research supports smart farming initiatives.

REFERENCE:

1) Thieme et al., “Using nasa Earth observations and google Earth engine to mapwinter cover crop conservation performance in the chesapeake bay watershed,” Remote Sens. Environ., vol. 248, 2020, Art. no. 111943.
2) J. Dai et al., “Mapping understory invasive plant species with field and remotely sensed data in Chitwan, Nepal,” Remote Sens. Environ., vol. 250, 2020, Art. no. 112037.
3) D.Feng,W.Xu,Z.He,W.Zhao,andM.Yang,“Advancesinplantnutrition diagnosis based on remote sensing and computer application,” Neural Comput. Appl., vol. 32, no. 22, pp. 16833–16842, 2020.
4) Y.Xie,J.Tian,andX.X.Zhu,“Linkingpointswithlabelsin3D:Areview of point cloud semantic segmentation,” IEEE Geosci. Remote Sens. Mag., vol. 8, no. 4, pp. 38–59, Dec. 2020.
5) D. Li et al., “Plantnet: A dual-function point cloud segmentation network for multiple plant species,” ISPRS J. Photogrammetry Remote Sens., vol. 184, pp. 243–263, 2022.
6) Y. Qiao et al., “Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping,” Front. Plant Sci., vol. 14, 2023, Art. no. 1188286.
7) P. Bedi and P. Gole, “Plant disease detection using hybrid model based on convolutional auto encoder and convolutional neural network,” Artif. Intell. Agriculture, vol. 5, pp. 90–101, 2021.
8) C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, and M. T. Kwasniewski, “Early detection of plant viral disease using hyperspectral imaging and deep learning,” Sensors, vol. 21, no. 3, p. 742, 2021.
9) T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on convolutional neural networks (CNN) in vegetation remote sensing,” ISPRS J. Photogrammetry Remote Sens., vol. 173, pp. 24–49, 2021.
10) S. Kendler et al., “Detection of crop diseases using enhanced variability imagery data and convolutional neural networks,” Comput. Electron. Agriculture, vol. 193, 2022, Art. no. 106732.