ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

Machine Learning Approaches for Crop Suitability Assessment and Disease Prediction

V Raviteja Kanakala[1]

Dept of CSE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, Pin:522501,

raviteja.kanakala@gmail.com[1]

K.Jagan Mohan[2]

Dept of Information Technology, Annamalai University, Tamilnadu, Pin:508002,

aucsejagan@gmail.com[2]

V.Krishna Reddy[3]

Dept of CSE, Gandhi Institute of Technology and Management, Andhra Pradesh, Pin:530045,

kvuyyuru@gitam.edu[3]

Y Jnapika[4]

Dept of CSE, Rajalakshmi Institute of Technology, Tamilnadu, Pin: 600124,

anjanadevi.aby06@gmail.com[4]

Abstract—

A sizable portion of farmers, especially in India, lack the expertise needed to choose crops and apply fertilizer in an intelligent manner. While machine learning algorithms have achieved significant strides in automating the identification of diseases, several problems have prevented Deep Learning from reaching its full potential. These include the need for high-quality training data, processing power limitations, and the models' restricted generalizability, all of which make application of the models difficult. An accessible and open-source web application has been created to tackle these problems and maybe improve agricultural yield. Plant disease prediction, fertilizer prescription, and crop appropriateness are the three main areas of emphasis for this application. Notably, attempts have been made to use interoperability techniques to explain the predictions produced by the crop disease detection system. Two different machine learning models were used in the study to identify plant diseases in leaves: a K-nearest Neighbors (KNN) model and a Convolutional Neural Networks (CNN) model. Furthermore, five machine learning techniques were applied for crop compatibility and fertilizer recommendation: Random Forest, Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Naïve Bayes. To determine which

machine learning model performed the best, a variety of measures were used to analyze the models.

Key words: Deep Learning, Machine Learning, Crop Prediction, Fertilizer recommendation, Disease Prediction

Introduction

By 2050, there will likely be nine billion people on the planet, which means that more food will need to be produced to keep up with demand. However, if we don't boost our output per unit area, this could cause issues with food security. We need cutting-edge technologies that can help us boost agricultural productivity per unit area to meet this challenge.

Using simulation models to maximize agricultural yield is one strategy. But in contemporary nations, the application of deep learning—a branch of computer science that makes machines behave like people—has grown in popularity. Deep learning holds great promise for agriculture, especially for crop condition monitoring (water scarcity, plant population, soil moisture content, etc.).

There are several agricultural applications for deep learning. For instance, by anticipating the amount of water required and accounting for weather, it may regulate irrigation water in the field. Because it can distinguish between weeds and crop seedlings, it can also be used for non-chemical weed control. Artificial intelligence-enabled drones may also offer precise mapping of crops in the field and apply fertilizers, insecticides, and pesticides that are tailored to each crop's unique needs. Because they have the potential to drastically lower crop production and jeopardize quality, plant diseases pose a serious danger to agricultural security. Convolutional neural networks (CNNs), one type of deep learning technique, have been utilized to precisely diagnose and identify plant illnesses. Most of the technological difficulties involved in classifying plant diseases can be solved by CNNs since they employ many layers to extract higher-level information from raw input. Still, the lack of comprehensive plant disease classification systems is due to dataset limits in terms of sample variety and quantity. Pattern recognition, classification, clustering, dimensionality reduction, computer vision, natural language processing, regression, and predictive analysis are just a few of the issues that deep learning can be used to. Vulnerable farmers, especially small

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

stakeholders, can use deep learning to take suitable preventive or mitigation actions in the event of crop illnesses, unfavorable weather, or problems with the health of the soil. During the past couple decades, deep learning has completely changed how machine learning is applied in agriculture. In conclusion, we need new and creative technology to boost agricultural production as the world population expands and the demand for food rises. A promising approach to solving agriculture's problems is deep learning, especially when it comes to crop management and plant disease detection. We can guarantee food security, boost crop yields, and lessen the harm caused by plant diseases by utilizing deep learning.

1. LITERATURE REVIEW

Muhammad Faheem et al. (2022) [1] introduced an IoT-based fertilizer recommendation system, evaluated its accuracy in soil fertility mapping, and compared it to conventional methods. Machine learning models, including SVM, LR, GNB, and KNN, were used to suggest tailored fertilizer recommendations based on soil type and macronutrient concentration. While a deep neural network was not suitable due to limited data, the authors suggested that combining a fresh dataset with a deep learning application could be a valuable future contribution.

In Yolo County, California, Zhong L. et al. (2019) [2] developed a deep learning-based framework for categorizing summer crops based on Landsat Enhanced Vegetation Index (EVI) time series. Long Short-Term Memory (LSTM) and one-dimensional convolutional (Conv1D) layers served as the foundation for the creation of two deep learning models. The best non-deep learning classifier was XGBoost, whereas the best F1 score (0.73) and accuracy (85.54%) were attained by the Conv1D-based model.

Mohanty et. al. (2016) [3] in a study 54,306 images of plant leaves with 38 class labels were analysed to predict crop-disease pairs using only the plant leaf image. The downscaled images were used for model optimization and predictions, achieving an overall accuracy ranging from 85.53% to 99.34% across different experimental configurations, demonstrating the potential of deep learning in similar prediction tasks.

Zhang et al. (2018) proposed deep learning models for recognizing maize leaf diseases. By adjusting parameters, changing pooling combinations, adding dropout operations, and reducing classifiers, they obtained two improved models with significantly

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

fewer parameters than VGG and AlexNet. The GoogLeNet model achieved a top-1 average identification accuracy of 98.9% for eight maize leaf diseases, and the Cifar10 model achieved 98.8%. The improved methods enhance accuracy and recognition efficiency.

Wang et al. (2021) developed a "Fertilizer Strength Prediction Model Based on Shape Characteristics" using machine vision and support vector machines. The model predicts fertilizer strength using a combined kernel function and optimized intrinsic parameters with a differential evolution method. The model showed high accuracy and reliability with an error rate below 5%, and the suggested combined kernel function outperformed other functions. The proposed model can be used for fertilizer production and quality inspection.

Nida Rasheed et al. (2021) [6] developed a decision support framework for crop production planning that utilizes spatio-temporal data to address policy gaps and management implications of crop allocation. The model aims to maximize revenue while adhering to production limits and includes historical data, national demand, and export needs. The proposed framework assists in managing overproduction and crop wastage and aids stakeholders in crop allocation, planting, and management.

Yan Qiao et al. (2020) [7] presented a novel approach for image restoration of crop leaf disease photos using Generative Adversarial Networks (GANs), which is a first in the agricultural disease image processing field. Their DATFGAN model, which utilized residual and dense connections and a dual-attention mechanism, outperformed state-of-the-art methods in terms of visual quality and classification performance. DATFGAN can improve classification accuracy while reducing network parameters, making it useful for real-world applications.

Rab Nawaz Bashir et al. (2022) [8] developed a machine learning model using IoT for predicting blister blight disease in tea plants by detecting environmental factors. Regression line models were used to determine the association between disease progression rate and environmental factors such as temperature, humidity, and rainfall. The model achieved high prediction accuracy and improved over time with the use of first-hand observation data for training. This proposed system aims to promote sustainable agriculture through judicious pesticide usage and disease management.

Jinge Xing et al. (2021) [9] presented a tomato leaf disease identification model using a restructured deep residual dense network. By incorporating the ResNet and DenseNet, the

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019**

model achieved 95% accuracy on the Tomato test dataset. Future work includes transferring the model to other plants via model changes to increase generalization capacity and contribute to agricultural intelligence development.

A modified recursive feature elimination (MRFE) technique was presented by G. Mariammal et al. (2021) [10] to pick significant traits from soil and environmental data to forecast crop compatibility. In order to determine which crop would be best suited for cultivation, the classification algorithms KNN, NB, DT, SVM, RF, and bagging were employed. The MRFE method with bagging classifier works better than other classifiers, according to the results. The effectiveness of the MRFE methodology was assessed and contrasted with existing methods, demonstrating that speed improvements are necessary for large datasets.

Ahlad Kumar et al. (2021) [11] Ahlad Kumar et al. (2021) suggests a real-time detection technique for plant diseases using an MLP model and ten characteristics obtained through soil sensors. The model can accurately diagnose the four most frequent illnesses with a subset accuracy of 94.36%. The utilization of sensors reduces the expert system's cost, making it more dependable and cost-effective. Future efforts will concentrate on expanding the sensor network to analyze spatial variances and enhance accuracy.

Shujuan Zhang et al (2021) [12] conducted a detailed assessment of recent research on plant leaf disease identification using deep learning. They discussed the significance of large datasets, data augmentation, transfer learning, and visualization of CNN activation maps to improve classification accuracy. However, most DL frameworks proposed in the literature are not resilient and require increased toughness to adjust to varied illness datasets. To overcome this, a big dataset of plant diseases under real-world situations is needed. Additionally, problems affecting the broad application of hyperspectral imaging (HSI) in early diagnosis of plant diseases must be overcome.

Akhilesh Kumar Sharma et al (2021) [13] proposed a crop recommendation system called "WB-CPI" that utilizes MapReduce and K-means clustering to determine the average yield for a variety of crops in a given location. The model considers factors such as soil type, seed type, ideal temperature, rainfall, and wind speed. The system can be scaled to suggest crops for different states and could be improved further by including additional factors such as soil moisture, irrigation, and cloud cover.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

Anjana Devi et al (2019) [14] proposed a framework for detecting and diagnosing diabetic retinopathy, a condition that can lead to vision loss or blindness. The traditional methods for diagnosis are costly and time-consuming, but deep learning techniques such as CNNs and SVMs have shown promising results with accuracy rates up to 97.93%, and some have been used in clinical settings.

Mirza Muhammad Waqar et al (2013) [15] proposed a method for assessing land suitability for rice cultivation in the Punjab region of Pakistan using existing soil datasets and GIS. They found that 72.2% of the agri-land in the study area was suitable for rice cultivation, with soil texture, water availability, and quality being key factors. GIS was deemed useful for evaluating crop acreage and planning agricultural operations, and addressing these factors could lead to improved rice production and increased yields.

Norasmanizan Binti Abdullah et al (2012) [16] proposed a land suitability mapping approach for precision farming, which involves image processing and integration with GIS. They found a correlation between soil compaction and pixel value in satellite images, and used regression techniques to determine the optimal correlation. The study highlights the importance of accurate image processing in remote sensing activities for identifying optimal sites for agriculture and enhancing productivity.

Sammy V. Militante et al (2019) [17] developed a deep learning-based method to detect and recognize plant diseases using convolutional neural networks (CNNs). The proposed method achieved an accuracy rate of 96.5% in identifying 32 different plant species and their diseases. The approach can be used in real-time by farmers to identify and manage plant diseases, potentially improving crop yield and quality. The authors suggest expanding the dataset and experimenting with different CNN designs to further enhance the model's performance.

2. PROPOSED METHODOLOGY

The proposed system involves a website in which the following applications are implemented - crop recommendation tool, a fertilizer recommendation tool, and a plant disease prediction tool. With the crop recommendation application, users can input their soil data and receive predictions on which crops would be most suitable for their site. Similarly, the fertilizer recommendation tool takes in soil data and crop information to make

ISSN PRINT 2319 1775 Online 2320 7876
Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

recommendations on how to improve soil deficiencies or excesses. Finally, the Plant Disease Predictor application allows users to upload pictures of diseased plant leaves, and the tool predicts the type of disease and provides information on how to cure it. By providing these applications, we hope to assist farmers in making data-driven decisions that can ultimately lead to increased crop yields and profits.

Advantage:

- ML and DL systems are helping to improve the overall harvest quality and accuracy – known as precision agriculture.
- This technology helps in predicting disease in plants, pests, and poor nutrition of farms.
- This technology has Improved resilience and reduced the risk of crop failure.

BLOCK DIAGRAM:

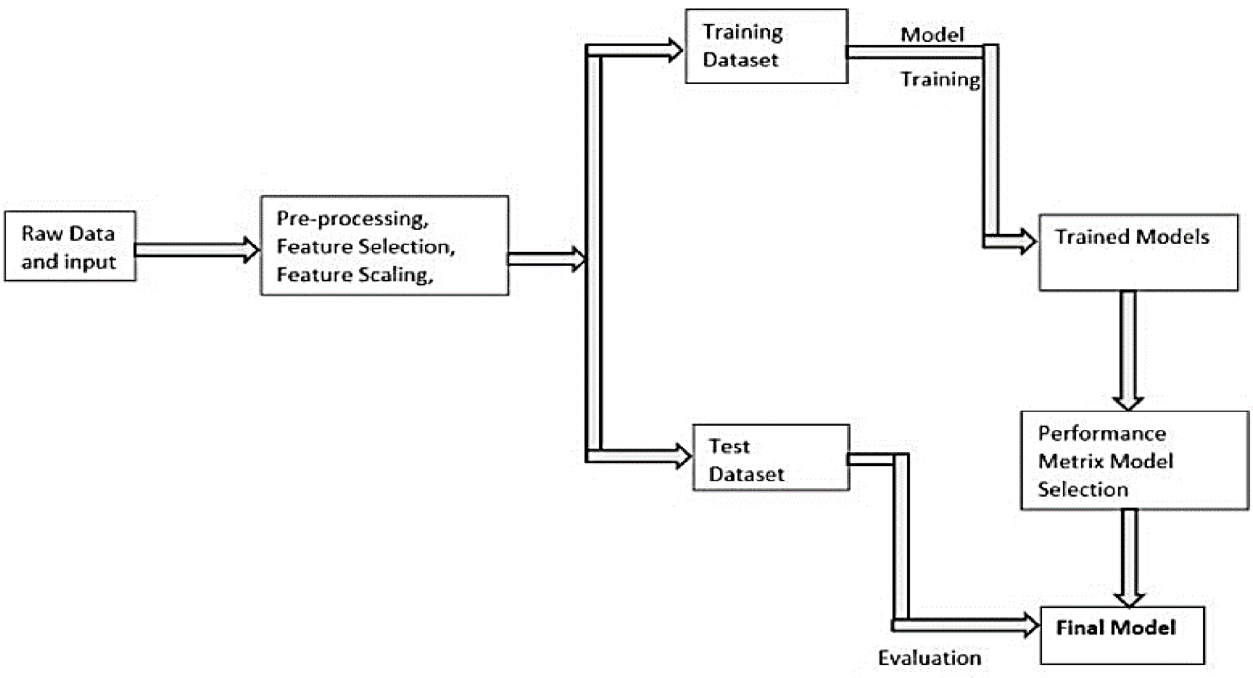

Fig 3.1: Block Diagram of Methodology

FLOW DIAGRAM:

Crop suitability:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

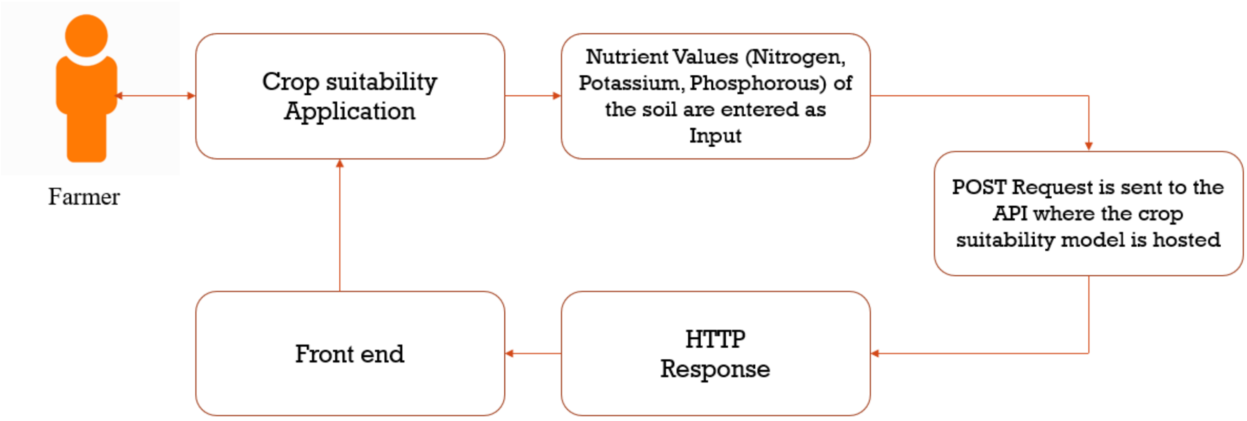

Fig 3.2.1: Flow of Crop suitability

Fertilizer Recommendation:

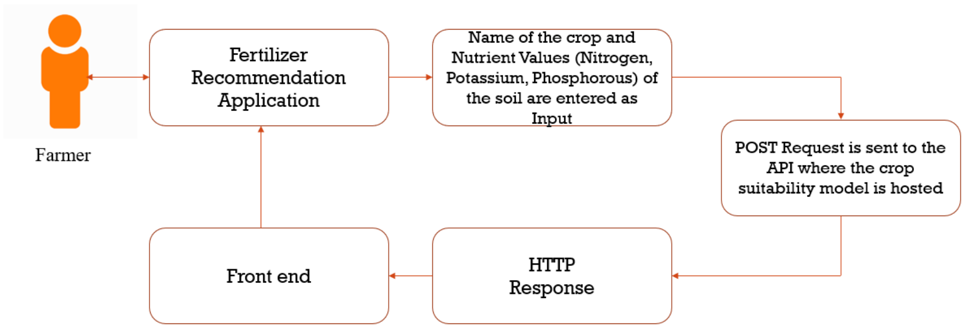

Fig 3.2.2: Flow of fertilizer recommendation

Disease prediction:

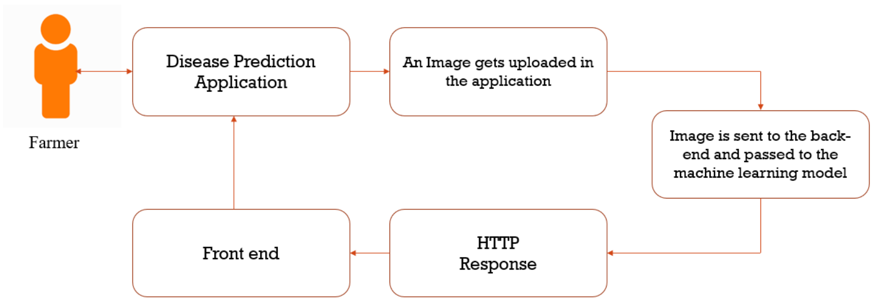

Fig 3.2.3 Flow of Disease prediction

PROCESS FLOW

- Dataset

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

- Data pre-processing.
- Splitting dataset.
- Build the model
- Training the model.
- Evaluating the model.
- Testing the model

PERFORMANCE EVALUATION:

This module's main goal is to assess how well-trained machine learning models perform. Several performance evaluation metrics, including the F1 score, accuracy, and classification error, are used to do this. Making ensuring the model is dependable and operating properly for the intended use is the aim. Metrics like accuracy, precision, and recall are frequently used to assess categorization models. The frequency with which the model predicts the outcome accurately is measured by accuracy, and the frequency with which it predicts a positive outcome is measured by precision. Recall quantifies the frequency with which the model accurately forecasts a favorable result, independent of the prediction's accuracy. To make sure that the crops recommended are suitable for the specified soil conditions, performance evaluation is essential when it comes to crop recommendations. It is feasible to pinpoint any flaws or potential areas for development in the machine learning model's performance and then optimize the algorithms to provide better outcomes. Thus, the model's dependability and efficacy for the intended purpose are further enhanced.

ALGORITHM USED:

Random Forest Algorithm

Random Forest Algorithm used for supervised learning that has significantapplications in both regression and classification tasks. The ensemble learning concept is employed in Random Forest which involves combining various classifiers to enhance model performance and solve complicated issues. It consists of multiple decision trees that operate on distinct subsets of the dataset providedIn order to improve the whole dataset's forecast accuracy, the model then averages the predictions of each decision tree.The Random Forest algorithm overcomes the problem of overfitting which is faced by single decision trees by considering

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

majority votes from the ensemble of trees to generate the final output. Accuracy is improved and 1the likelihood of the model overfitting is decreased as the number of trees in the forest increases.

Convolutional Neural Network:

Convolutional Neural Network (CNN) is a popular type of neural network for performing image classification and recognition tasks. It is commonly used in various fields, such as scene labeling, object detection, and face recognition.

To process an image in CNN, the computer first sees it as an array of pixels with dimensions h x w x d, where h stands for height, w for width, and d for dimension. The actual values of h, w, and d depend on the resolution of the image, such as a 6 x 6 x 3 array for an RGB image or a 4 x 4 x 1 array for a grayscale image.

In CNN, the input image is processed through a series of convolution layers, pooling layers, and fully connected layers using filters or kernels. Finally, the Softmax function is applied to classify the object with probabilistic values ranging from 0 to 1.

RESNET:

By adding more layers to their architecture, deep neural networks can learn ever more complicated information, which enhances their accuracy and performance in handling complex problems. Plotting the error % for training and testing data in a 20-layer network and a 56-layer network illustrates how adding more layers to conventional convolutional neural network models can result in a decrease in performance. Rather than overfitting, problems with the optimization function, network initialization, and the vanishing gradient problem could be the cause of this decline in performance. To avoid declining performance, it is therefore essential to set a maximum threshold for depth in deep neural network models. The introduction of Residual Blocks-based ResNets, or residual networks, is a solution to this problem. One of the main distinctions between Residual Blocks and other models is the existence of a direct connection that bypasses certain layers. The core of residual blocks is this link, which is referred to as the "skip connection". This skip connection is incorporated, changing the layer's output. In its absence, the input 'x' is sent via the activation function f(), multiplied by the layer weights, and added to a bias term to produce the output H(x). But the output becomes H(x)=f(x)+x when the skip connection is introduced.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

When the input and output dimensions are different, as they are for convolutional and pooling layers, this method has a small problem. In this case, two ways can be used: either the projection method (which matches the dimension by adding 1×1 convolutional layers to the input) or padding the skip connection with extra zero entries to expand its dimensions. While the second way adds an extra parameter w1 to the output, resulting in H(x)=f(x)+w1. x, the first approach requires no more parameters.

RESULT

The implementation of machine learning algorithms in crop suitability and disease prediction has shown promising results. The developed web application provides farmers with recommendations for crops that are best suited for their location and nutritional requirements, along with recommendations for the optimal fertilizer to be used. The results showed that Random Forest achieved the highest accuracy of 95%, followed by Naive Bayes at 91%, SVM at 87%, Decision Tree at 82%, and Logistic Regression at 67%. The application also utilizes machine learning algorithms to predict and detect crop diseases, helping farmers take timely action to prevent the spread of the disease and minimize crop damage. The incorporation of the LIME interpretability method enhances the transparency and interpretability of the model, enabling users to understand why certain recommendations are being made.

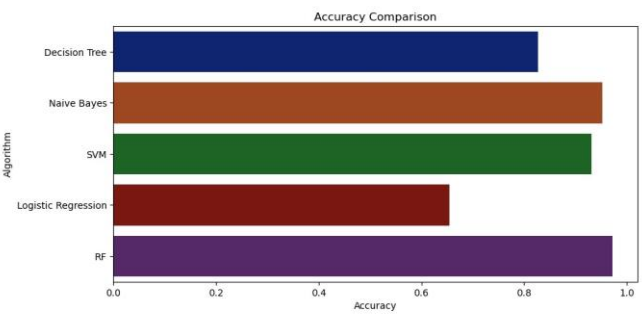

Fig 4.1 Accuracy comparison

According to the comparative analysis performed, Random Forest Algorithm shows maximum efficiency with 95%.

CONCLUSION

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 4, 2019

The proposed system consists of a machine learning and web-scraping-based web application which offers user-friendly features such as crop recommendation using the Random Forest algorithm, fertilizer recommendation using a rule-based classification system, and crop disease detection utilizing the EfficientNet model on leaf images. Our system provides an easy-to-use interface for users to input their data and quickly receive accurate results. To enhance the transparency and interpretability of our model, we incorporated the LIME interpretability method, which can explain the predictions of the disease detection image. This additional feature can potentially help users understand why the model makes certain predictions, leading to further improvements in datasets and models.

REFERENCES

[1] Liheng Zhonga, Lina Hub, Hang Zhou, “Deep Learning Based Multi-Temporal Crop Classification” https://Doi.Org/10.1016/J.Rse.2018.11.032

[2] Sharada P. Mohanty, David P. Hughes and Marcel Salathé on “Using Deep Learning for Image-Based Plant Disease Detection” https://doi.org/10.3389/fpls.2016.01419.

[3] Xihai Zhang, Yue Qiao , Fanfeng Meng, Chengguo Fan , And Mingming Zhang, “Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks”,DOI:10.1109/ACCESS.2018.2844405.

[4] Changjian Zhou, Sihan Zhou, Jinge Xing , And Jia Song, “Tomato Leaf Disease Identification By Restructured Deep Residual Dense Network”, DOI: 10.1109/Access.2019.3058947.

[5] G. Mariammal, A. Suruliandi, S. P. Raja, And E. Poongothai, “Prediction of Land Suitability for Crop Cultivation Based on Soil And Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers” IEEE Transactions On Computational Social Systems

[6] Manish Kumar, Ahlad Kumar, And Vinay S. Palaparthy, “Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning”, Ieee Sensors Journal, Vol. 21, No. 16, August 15, 2012.

[7] Pradeep Kurup, and Pijush Samui and ET all(2015). Examining Efficacy of Metamodels in predicting Ground Water Table, International Journal of Performability Engineering, Vol. 11, No. 3, pp. 275-281..

[8] Rishi Gupta, Akhilesh Kumar Sharma, Oorja Garg, Krishna Modi, Shahreen Kasim, Zirawani Baharum, Hairulnizam Mahdin, And Salama A. Mostafa, “Wb-Cpi: Weather Based Crop Prediction in India Using Big Data Analytics”, DOI: 10.1109/Access.2012.311724.

[9] V Anjana Devi, R Hemalatha, R Venkateshwar,J Naren and G Vidhya on “A framework for the Diagnosis of Diabetic Retinopathy Using Deep Learning Techniques”, International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 01, 2019 ISSN: 1475-7192.

[10] Mirza Muhammad Waqar, Faiza Rehman, Muhammad Ikram3 on “Land Suitability Assessment for Rice Crop Using Geospatial Techniques”, IGARS 2013, IEEE 978-1-4799-1114-1/13.

[11] Norasmanizan Binti Abdullah, Ayu Wazira Azhari, Mohd Sabri Hussin, Mahmad Nor Jaafar on “Land Suitability Mapping for Implementation of Precision Farming”,2011 National Postgraduate Conference, DOI: 10.1109/NatPC.2011.6136375.

[12] N. Deepa, R.Chandrasekar, R. Sathiyaseelan and ET all (2014)Dimension Reduction using Multivariate Statistical Model, International Journal of Applied Engineering Research, Volume 9, Number 14 (2014) pp. 2531-2538 .

[13] V. Anjana Devi (2011), "Agent Based Cross Layer Intrusion Detection System for MANET", International conference on Network Security CNSA 2011, ISBN 978-3-642-22539-0, pp 427-440. (Springer CCIS).