https://doi.org/10.33472/AFJBS.6.6.2024.5639-5650

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Rhizome Rot Disease Classification Using Hybrid Randomforest and Adaboost

V.Janani M.E[1], Dr.N.M.Siva Mangai[2]

[1]Research Scholar, ECE, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

[2]Associate Professor, ECE, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu

Email: [1]jananidharshan@gmail.com

Article Info

Volume 6, Issue 6, June 2024

Received: 22 April 2024

Accepted: 24 May 2024

Published: 17 June 2024

doi: 10.33472/AFJBS.6.6.2024.5639-5650

ABSTRACT:

One such crop that is widely farmed around the world is turmeric. Crop output is significantly impacted by the plant disease. The most typical plant diseases are rhizome rot, leaf spot, and leaf blotch. Rhizome rot is one of the diseases that has been documented to be particularly harmful to the production of turmeric, with a 60% crop loss rate. In order to address the multi-class unbalanced rhizome rot dataset classification, adaptive boosting (AdaBoost) in conjunction with a potent ML classifier (Random Forest) is suggested in this study. AdaBoost builds a strong classifier by integrating many sub-classifiers based on weights. Using Random Forests as the first stage classifier and AdaBoost as the second stage classifier to determine which class the illness sample falls into is a novel, robust, and more accurate method offered. To assess the robustness of the suggested hybrid AdaBoost approach, three different datasets are used. The suggested technique enhances both accuracy and stability for the unbalanced rhizome, according to experimental findings, which are compared to various state-of-the-art algorithms (kNN, SVM, RF, kNN-AdaBoost, and SVM-AdaBoost).RFAdaBoost has a 92% f1_score and yields the lowest root mean squared value. When the hybrid algorithm's stability is assessed using the variance, g-means metric value, and Matthew's correlation coefficient (MCC), RFAdaBoost performs better than the other cutting-edge methods.

Keywords: Machine Learning; Random Forest; Support vector machine; kNN; AdaBoost

1. INTRODUCTION

The goal of this agricultural research is to raise food quality and productivity while lowering costs and increasing profit. By 2050, the global food demand will have increased by 50%. The worldwide agriculture sector, valued at $5 trillion, is currently transitioning to artificial intelligence (AI) technology. AI makes it simpler for farmers to analyse real-time data, such as temperature, soil conditions, or weather, in order to maximize agricultural output. AI is also used in precision agriculture to help identify low-nutrient plants and detect illnesses and pests. AI-enabled sensors are able to identify weeds and determine which pesticides should be used in the field.In order to boost agricultural output, seasonal forecasting models are also made using AI. AI-enabled cameras mounted on drones may take pictures of farms and use the data to identify issues and provide recommendations for possible fixes.

Although the herb turmeric is widely grown in South and East India, it is mostly grown as an annual crop in Malaysia and India.Many diseases, including rhizome rot, leaf spot, and leaf blotch, can reduce crop output. To raise the manufacturing rate, suitable detection and prevention techniques must be found. Pesticide use will be decreased and environmental conditions will be protected by early disease diagnosis and control strategies.

In large-scale planting, conventional methods for detecting leaf diseases are not very effective. Compared to manual diagnosis and conventional identification techniques, machine learning and image processing technologies provide several advantages. Initially, photos are taken in agricultural fields. Infected leaf sections are removed from the picture during pre-processing. Features are retrieved from segmented pictures. Plant leaf diseases are detected by using these traits in conjunction with appropriate categorization procedures. Training with a large number of datasets increases memory requirements and training time. These conventional issues are handled with the use of artificial intelligence (AI) models. The creation of AI models has made it possible to gather enormous volumes of data, create and program algorithms, extract rules from data, and build models using contemporary hardware.

2. LITERATURE SURVEY

India's effective production of turmeric (Curcuma longa L.) is hampered by the deadly disease known as rhizome rot. There are theories on the connection between rhizome rot caused by turmeric and several fungal diseases. The goal of Chenniappan, C. et al. [1] was to discover the fungus linked to the illness rhizome rot. From symptomatic rhizomes taken from different turmeric farms in Tamil Nadu, India, 51 fungal strains were discovered. Using the turmeric variety "Erode Local" in a greenhouse experiment, 18 morphologically distinct isolates were examined for pathogenicity. Based on morphology, eleven fungal isolates that might produce symptoms with $\geq$30% disease severity were found, and molecular methods were used to confirm them. The present study's results highlighted the variety of fungi linked to rhizome rot and their pathogenicity, which may be utilized to create a successful management plan for the condition.

Although a valuable crop, turmeric is susceptible to a number of illnesses that can seriously affect its yield. Preventing crop failure and losses requires early diagnosis of these diseases. Rout, S. et al. [2] used a single-phase machine learning-based detection model to develop a unique method for precisely diagnosing illnesses of turmeric plants. In contrast to the conventional YOLOV3-Tiny model, they presented the Improved YOLOV3-Tiny model, which uses convolutional neural networks and a residual network structure to increase detection accuracy. Using both day and night photos, the authors evaluated their approach and discovered that it performed better than other methods like YOLO and Quicker R-CNN with the VGG16 algorithm. Additionally, they discovered that adding Cycle-GAN to the turmeric leaf dataset increased detection accuracy, especially for smaller samples. Their model is a useful tool for identifying illnesses of turmeric leaves since it provides both high accuracy and quick identification speed.

In villages studied by Tejas Kumar, M.K. et al. [3], the prevalence of rhizome rot complex was greater in ginger crop monocultures or repeated cultivations than in ginger crops cultivated just after rice or maize cultivation. When compared to the ginger crop planted in between plantation crops, the frequency of rhizome rot complex was greater in certain of the villages examined. Rhizome rot complex is more common in Karnataka's ginger-growing regions for a variety of reasons, including improper planting material selection, poor field drainage, monocropping, a lack of crop rotation, careless fertilizer usage, and a lack of disease management expertise.

The rhizome of ginger usually consists of water, proteins, lipids, carbs, fibers, and aromatic oils. In addition to its nutritional and flavourful qualities, ginger may have several therapeutic benefits in traditional therapy. Because of the immense relevance of ginger, there have been periodic expansions in the total area for agricultural production; nevertheless, the productivity of ginger has unfortunately diminished over time due to the soft rot sickness. Soft rot is one of the most common diseases that damage ginger and is usually caused by the bacteria Ralstonia spp. and the fungus Pythium and Fusarium spp. The rhizome's juicy portions, collars, and roots are all impacted by the microbes that cause the disease. Soft rot is the most dangerous and destructive ginger disease known to science. Since then, the soft rot of ginger has been linked to other Pythium species. Following an investigation of the 11 Pythium varieties connected to ginger soft rot, Dohroo et al. discovered that P. myriotylum and P. aphanidermatum were the most prevalent variants.

In order to maximize the output of turmeric oil, Abduh, M.Y. et al. [4] set out to ascertain the function of Aspergillus awamori, Aspergillus niger, and Aspergillus oryzae in starch degradation on turmeric rhizome substrate. To satisfy the nutritional requirements of fungal development, an additional 10% weight per volume of yeast extract was added to the substrate, which is the rhizome of turmeric. The inoculation employed 5x107 cells/ml of fungal concentration. The solid-state fermentation process was run at 25–28 ºC, 99% humidity, dark conditions (~0 W), and 3.5 L/min of aeration. Steam distillation was used for three hours to extract the turmeric oil, with a substrate moisture content of 68–71% and a substrate–water ratio of 1:5. For eleven days, the biodegradation process took place. Specifically on days 7, 9, and 11, the starch content and production of turmeric oil were measured throughout the fermentation process. The outcomes demonstrated that boosting the output of turmeric oil by solid-state fermentation's starch biodegradation process was successful. On the eleventh day of fermentation, Aspergillus awamori demonstrated the most favourable starch breakdown activity, 62.5% to 2.9% wet weight. After the eleventh day of fermentation, Aspergillus oryzae had the greatest beneficial impact, almost tripling the production of turmeric oil to 3.17% dry weight. Turmeric oil primarily consists of three compounds: ar-, α-, and β-turmerone.

Rhizome rot in ginger is caused by bacterial infections (Zingiber officinale). It is unknown who the important members of the endophytic microbial population are in ginger rhizomes and how they affect rhizome deterioration when adventitious bud growth is activated. Inoculation assays and high-throughput 16S rRNA amplicon sequencing were utilized to examine the pathogenicity, community makeup and structure, and microbial diversity of the isolated bacteria. Huang, K., et al [5] suggested that as the rhizome rot illness progressed, the endophytic microbiota's composition changed. The bacterial community of rhizomes displaying bacterial decay symptoms included Enterobacteriaceae, Lachnospiraceae, and the genera Clostridium, Bacteroides, Acrobacter, Dysgonomonas, Anaerosinus, Pectobacterium, and Lactococcus. These genera were also present in rhizomes that showed no symptoms.

The three main illnesses associated with turmeric are taphrinamaculans-induced leaf blotch, leaf spot produced by species of Colletotrichum, and rhizome rot caused by several Pythium species. Rhizome rot is the most damaging disease, causing the crop to suffer financial losses (Rathiah, 1982). The Telengana area of Andhra Pradesh, where the crop is farmed on a vast scale, has recorded crop losses of up to 50%. A thorough analysis of the literature showed that the disease is complicated, including a variety of fungal species as well as other creatures like nematodes, and that the true cause of the illness remains a matter of debate.

3. METHODOLOGY

To get the best results using ML classifiers, AdaBoost can be applied. Programming AdaBoost is rapid, easy, and requires little work, and there are very few tweaking restrictions. Because it can be used with a variety of techniques, like SVM, neural networks, random forests, etc., this is quite flexible.

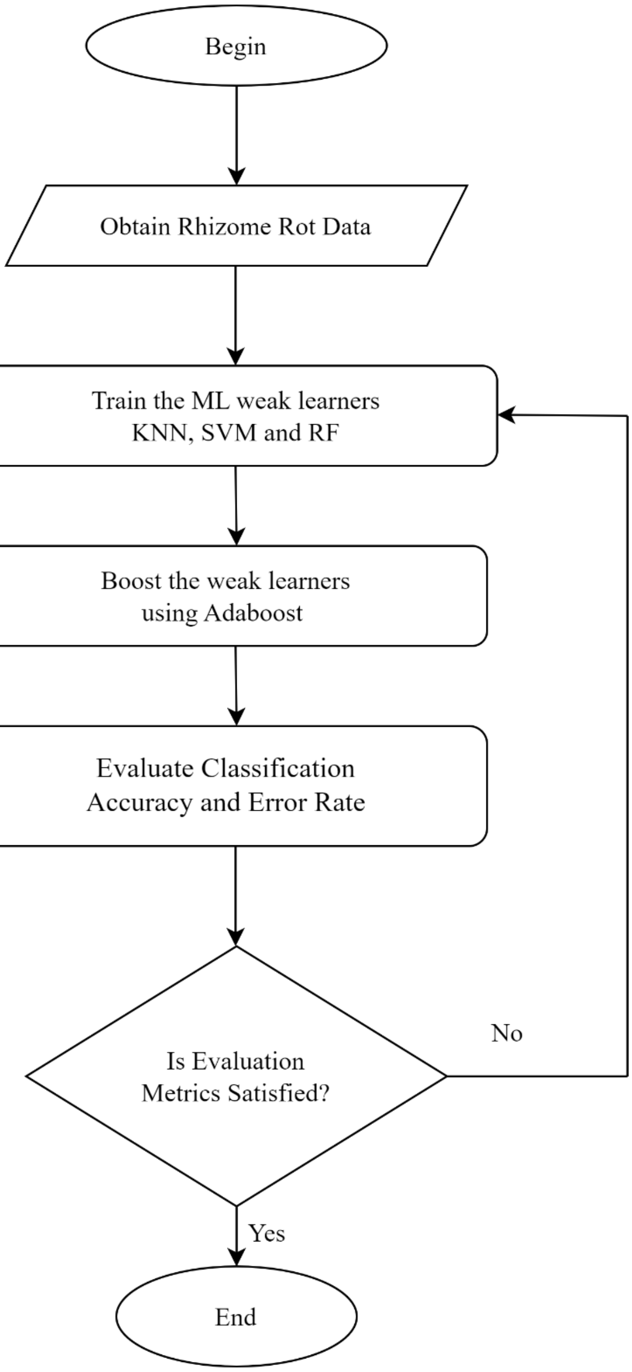

Fig 3.1 Flow Diagram

Boosting is a typical method of combining somewhat correct hypotheses (weak classifiers) to improve the performance of a learning system. The Boosting strategy reduces the mistakes caused by the weak hypotheses. Adaptive Boosting is an iterative training technique created by Freund and Schapire. The next weak learner system is trained with newly created data points after the original data points have been applied to the weak learner and the poorly categorized points have been weighted with different probabilities, V.

AdaBoost iterates the T cycles weak trainer algorithm repeatedly. Subsequently, Adaboost updates the distribution V for each T cycle by decreasing the weights assigned to successfully categorized sample points and increasing the weights assigned to erroneously classified data points. Ultimately, a single, highly accurate classifier is formed by combining the moderately correct possibilities.The primary characteristic of AdaBoost is that when the accuracy of the weak classifiers is more than half, the learning error of the final strong classifier gradually approaches zero. It implies that component classifications will be much higher than arbitrary guesses.

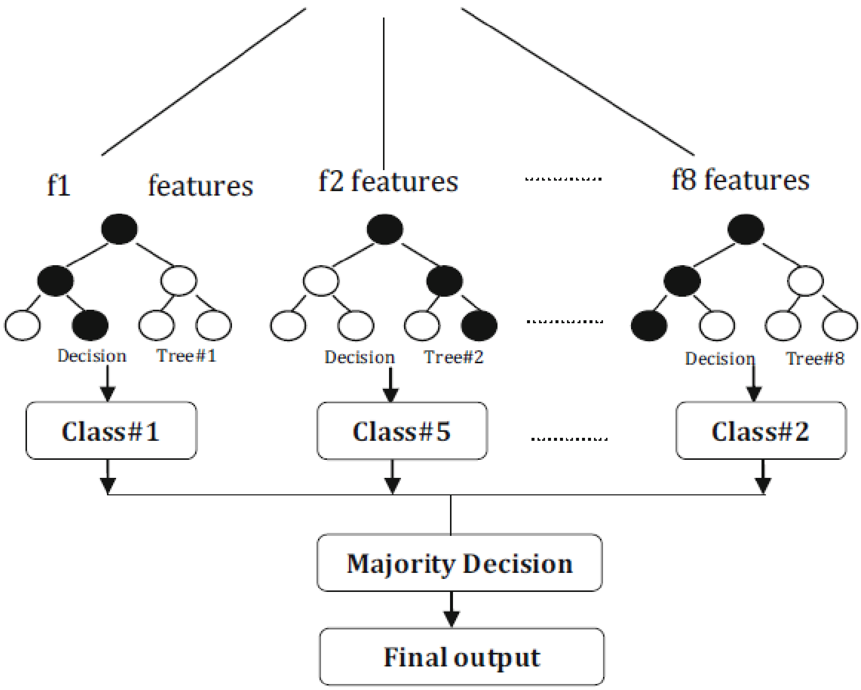

Fig 3.2 Random Forest

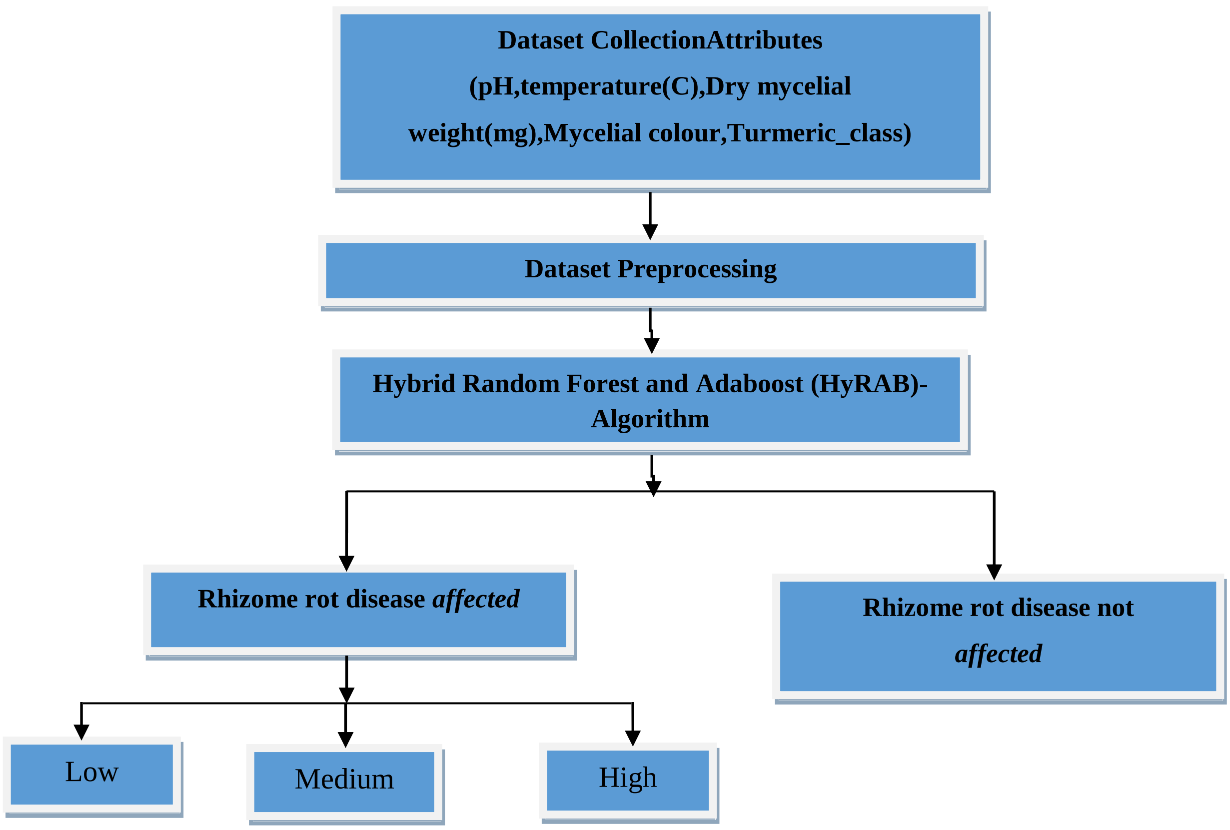

Fig 3.2 Proposed Methodology Frameworks

Step 1: Train a Random Forest Classifier:

1. Start by training a Random Forest classifier on your dataset. Random Forest is known for its ability to handle complex, high-dimensional data and reduce overfitting.
2. Tune the hyperparameters of the Random Forest, such as the number of trees (n_estimators), maximum depth of trees (max_depth), minimum samples per leaf (min_samples_leaf), and feature selection method to optimize its performance.

Step 2: Train an AdaBoost Classifier on Random Forest's Output:

1. Use the predictions of the Random Forest as input for an AdaBoost classifier. AdaBoost will focus on the instances that the Random Forest had difficulty classifying.
2. You can initialize the AdaBoost model with the Random Forest's output as weights or probabilities to guide the boosting process. AdaBoost will adapt to the challenging cases identified by the Random Forest.
3. Tune the hyperparameters of the AdaBoost algorithm, such as the number of weak learners (base models), learning rate, and maximum depth of the base models.

Step 3: Combine Predictions:

1. Combine the final predictions from the Random Forest and AdaBoost. You can use weighted averaging, voting, or any other combination method. You may assign different weights to the predictions from the Random Forest and AdaBoost based on their individual performance.

Benefits of a Hybrid Random Forest-AdaBoost Classification Model:

- Random Forest can capture complex patterns and handle high-dimensional data well.
- AdaBoost can further improve classification performance by focusing on instances that the Random Forest struggles with, effectively reducing errors.

Considerations:

- Proper feature engineering and preprocessing are essential for both the Random Forest and AdaBoost components. Ensure that the input data is appropriately prepared for each algorithm.
- Hyperparameter tuning is crucial for both Random Forest and AdaBoost. Experiment with different settings to find the optimal configuration for each.
- Regularization parameters in the Random Forest, such as the depth of trees, play an important role in controlling overfitting and model performance.
- Evaluate the computational cost and training time, especially for large datasets. Hybrid models can be more computationally intensive.
- Proper cross-validation and assessment of model performance are necessary to determine whether this hybrid approach provides significant improvements over using either algorithm individually.

By combining Random Forest and AdaBoost, you can create a robust hybrid classification model that benefits from the strengths of both ensemble techniques. This approach can be particularly effective in scenarios where the individual models alone may not provide optimal results.

Algorithm Steps for Hybrid Random Forest with AdaBoost

Hybrid Random Forest with AdaBoost involves combining the strengths of both algorithms. The idea is to use Random Forest as a base estimator for AdaBoost, where AdaBoost will adaptively boost the performance of Random Forest. Below are the steps for a hybrid Random Forest with AdaBoost algorithm:

1. **Initialize Weights:** Assign equal weights to all samples in the training dataset.
2. **For each iteration (t):** a. **Initialize Random Forest:** Train a Random Forest classifier on the training data with the weights assigned in the current iteration. Random Forest builds multiple decision trees using bootstrapped samples and feature randomization.

b. **Compute Error:** Calculate the weighted error rate of the Random Forest predictions on the training data, considering the misclassifications and their respective weights.
c. **Update Weighted Importance:** Update the weights of the samples. Increase the weights of the misclassified samples so that the next Random Forest iteration focuses more on the previously misclassified samples.
d. **Compute Classifier Weight:** Calculate the weight of the current Random Forest iteration in the final ensemble. This weight is based on the Random Forest's performance.
e. **Combine Random Forest Learner:** Add the Random Forest model to the ensemble, considering its classifier weight in the final prediction.

3. **Final Ensemble:** Combine the Random Forest models with their respective weights to form a strong ensemble classifier through AdaBoost.

```
# Random Forest with AdaBoost Pseudo-code


# Step 1: Import necessary libraries and modules


fromsklearn.ensembleimportRandomForestClassifier, AdaBoostClassifier
fromsklearn.datasetsimportmake_classification
fromsklearn.model_selectionimporttrain_test_split
fromsklearn.metricsimportaccuracy_score


# Step 2: Generate a sample dataset (you can replace this with your dataset)
X, y = make_classification(n_samples=1000, n_features=20, n_classes=2, random_state=42)


# Step 3: Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


# Step 4: Initialize Random Forest and AdaBoost classifiers
random_forest = RandomForestClassifier(n_estimators=100, random_state=42)
adaboost = AdaBoostClassifier(base_estimator=random_forest, n_estimators=50, random_state=42)


# Step 5: Train the AdaBoost with Random Forest as the base estimator
adaboost.fit(X_train, y_train)


# Step 6: Predict using the combined model
predictions = adaboost.predict(X_test)


# Step 7: Evaluate the model
accuracy = accuracy_score(y_test, predictions)
print(f"Accuracy: {accuracy}")
```

The machine learning classifier models' performance with the unbalanced dataset is assessed using two key metrics: (i) accuracy and (ii) f1_score. To help determine the overall impact of the hybrid models for Rhizome rot datasets, these two-performance metrics are calculated using a confusion matrix.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Sensitivity\ (Or)\ Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Classification\ Accuracy = \frac{Number\ of\ Accurately\ predicted\ Rhizome\ Rot\ Dataset}{Total\ Number\ of\ Infected\ \ dataset} \times 100\%$$

$$Classification\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

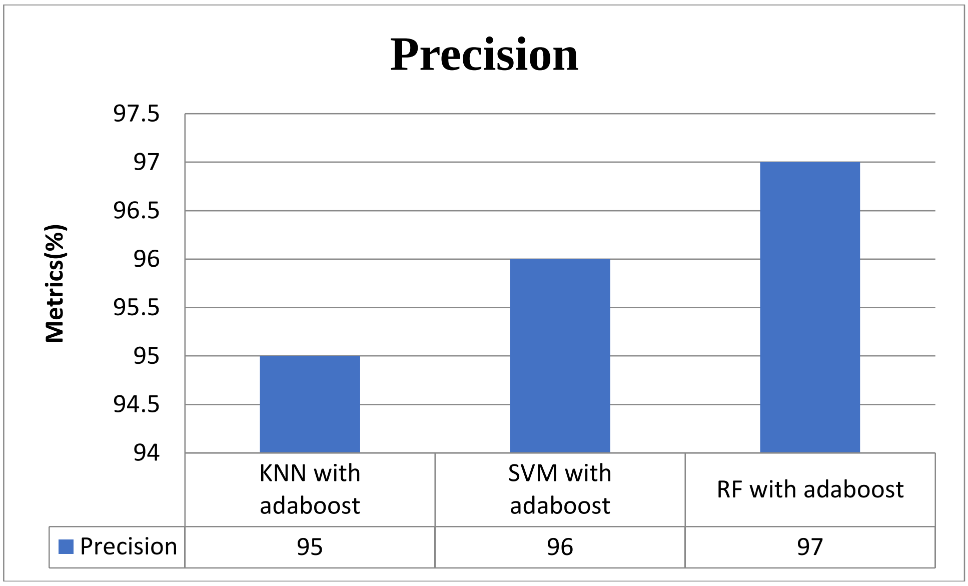

Fig 4.1 Performance Metrics Comparison_ Precision

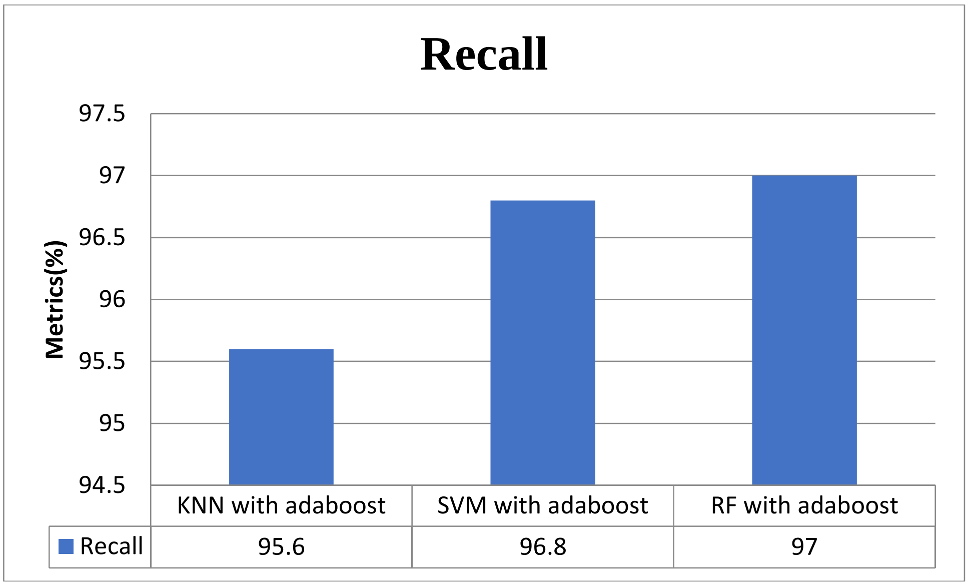

Fig 4.2Performance Metrics Comparison_ Recall

The reliability of the model is determined by measuring the degree to which it recognizes exact samples from the datasets; the f1_score is estimated using precision and recall parameters. When evaluating a study area with an unbalanced dataset, the classifier accuracy shouldn't be the default statistic used. F1_score is utilized for the stable performance evaluation, which is required to comprehend the results of the hybrid boosting methods on the minority class.

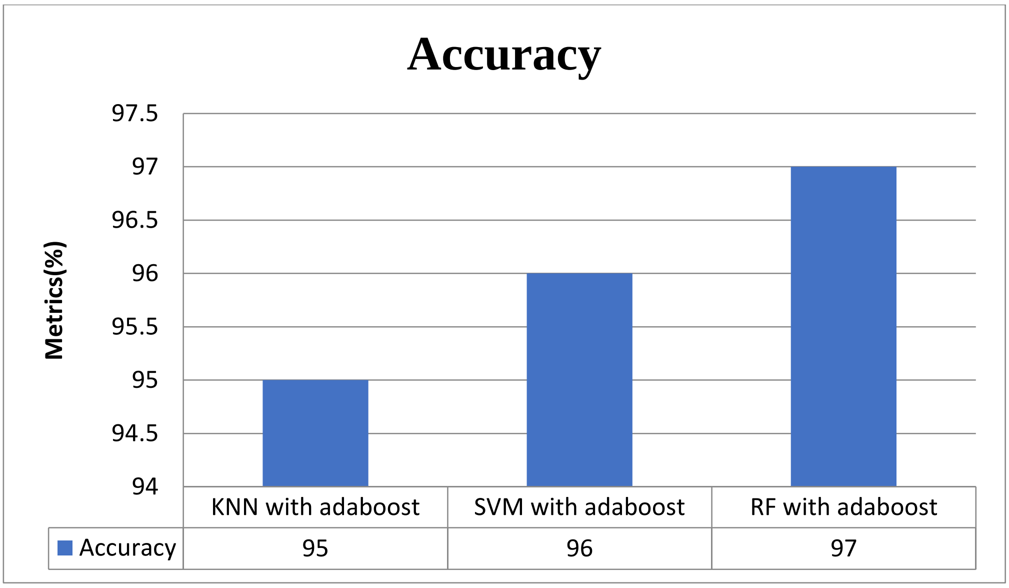

Fig 4.3Performance Metrics Comparison_ Accuracy

A number of experimental findings are showcased in order to evaluate the effectiveness of the suggested hybrid AdaBoost algorithm. Two analyses are made of the experiments. The first investigation looks into the possibility of improving the classification results by combining ML with adaptive boosting. In light of the hybrid models, three single algorithms—kNN, SVM, and RF—are contrasted. This evaluation's primary goal is to demonstrate the hybrid RFAdaBoost's resilience.

5. CONCLUSION

Due to its many uses, turmeric is widely utilized in many different sectors. It finds use in many areas such as healthcare, food, cosmetics, and textiles. Technology should be used to boost the production of turmeric. Rhizome rot may be accurately identified with the Hybrid Random Forest and Adaboost Technique. No other conventional approach can achieve error-free outcomes than deep learning. As the values in the suggested technique will be acquired in real time, the amount of the dataset will expand and maximum accuracy will result. The accuracy increases with the number of datasets we train. For rhizome rot, there isn't any disease control software or complete monitoring available right now. Consequently, we have put up a creative strategy to get over the present obstacles.

6. REFERENCES

1. Chenniappan, C., Daniel, G.M., Ponnusamy, P. and Narayanasamy, M., 2020. Association of multiple fungal pathogens to rhizome rot disease of turmeric. Archives of Phytopathology and Plant Protection, 53(9-10), pp.440-458.
2. Rout, S. and Venu, N., Machine Learning Based Analysis And Classification Of Rhizome Rot Disease In Turmeric Plants.
3. Tejas Kumar, M.K., Ravikumar, M.R., Lokesh, B.K. and Ganpathi, T., 2022. Prevalence of rhizome rot complex of ginger in major ginger growing parts of Karnataka.
4. Abduh, M.Y., Ramawati, A.I., Komariyah, T. and Theda, Y., 2023. Solid State Fermentation of Turmeric Rhizomes with Aspergillus SP. To Improve Yield and Composition of Extracted Turmeric Oil. *IIUM Engineering Journal*, *24*(1), pp.1-12.
5. Huang, K., Sun, X., Zou, Y., Li, H., Xu, P., Zhang, W., Zhang, Y., Li, H., Sun, H., Wang, W. and Pang, M., 2022. Comparison of the Endophytic Bacterial Microbiota of Asymptomatic and Symptomatic Ginger Rhizomes During the Activation of Adventitious Bud Development. *Plant Disease*, *106*(9), pp.2470-2479.
6. Daniel-Ogbonna, C.I. and Okoye, A.C., 2023. UNLOCKING THE BENEFITS OF TURMERIC (CURCUMA LONGA L) ROOT CROPS FOR AGRIBUSINESS DEVELOPMENT IN NIGERIA. CENTENNIAL, p.549.
7. Vinayarani, G., Madhusudhan, K.N. and Prakash, H.S., 2019. Induction of systemic resistance in turmeric by rhizospheric isolate Trichoderma asperellum against rhizome rot disease. *Journal of Plant Pathology*, *101*, pp.965-980.
8. Li, S., Long, M., Guo, L., Luo, Y., Sun, X., Qian, K., He, L. and Ma, G., 2023. The Identification and Characterization of a New Species, Neopestalotiopsis Strain CP1, a Pathogen Causing Leaf Blight on Curcuma phaeocaulis in Guangxi, China.
9. Surajit, K. and Chowdhury, A.K., 2008. Biological control of rhizome rot disease of turmeric. *Journal of Mycopathological Research*, *46*(1), pp.127-128.
10. Brijesh, H. and Ajjappala, B., 2023. Micropropagation strategies in medicinally important turmeric (Curcuma sp): Current research and future challenges. *Journal of Applied Biology and Biotechnology*, *11*(3), pp.1-8.
11. Wang, H., Su, J., Liu, Z., Lin, S., Guo, J., Gao, Q. and Zhou, H., 2023. Curvularialingshanensis sp. nov., a New Pathogen Causing stem and leaf rot on Curcuma kwangsiensis and Its Sensitivity to Five Fungicides.
12. Abduh, M.Y., Ramawati, A.I., Komariyah, T. and Theda, Y., 2023. Solid State Fermentation of Turmeric Rhizomes with Aspergillus SP. To Improve Yield and Composition of Extracted Turmeric Oil. *IIUM Engineering Journal*, *24*(1), pp.1-12.

13. Singh, A., Chandra, P., Bahadur, A., Debnath, P., Subbaiyan, G.K., Ellur, R.K., Siddappa Tadasanahaller, P., Hosahalli Manjunath, A.K., Yadav, G.K. and Bashyal, B.M., 2023. Assessment of morpho-cultural, genetic and pathological diversity of Rhizoctonia solani isolates obtained from different host plants. *Journal of Plant Pathology*, pp.1-16.
14. Gohel, N.M., Mistry, S., Rathava, A. and Dhaduk, H.L., 2022. Management of leaf blotch (Taphrina maculans Butler) and leaf spot (Colletotrichum capsici (Syd.) Butler & Bisby) diseases in turmeric through ready-mix fungicides under field conditions. *Indian Phytopathology*, *75*(2), pp.487-491.
15. Ibrahim, N.N.A., Wan Mustapha, W.A., Sofian-Seng, N.S., Lim, S.J., Mohd Razali, N.S., Teh, A.H., Rahman, H.A. and Mediani, A., 2023. A Comprehensive Review with Future Prospects on the Medicinal Properties and Biological Activities of Curcuma caesiaRoxb. *Evidence-Based Complementary and Alternative Medicine*, *2023*.
16. Huang, K., Sun, X., Zou, Y., Li, H., Xu, P., Zhang, W., Zhang, Y., Li, H., Sun, H., Wang, W. and Pang, M., 2022. Comparison of the Endophytic Bacterial Microbiota of Asymptomatic and Symptomatic Ginger Rhizomes During the Activation of Adventitious Bud Development. *Plant Disease*, *106*(9), pp.2470-2479.
17. Thammakijjawat, P., Thaveechai, N., Paradhonuwat, A., Wannakairoj, S. and Suthirawut, S., 1999. Bacterial rhizome rot of patumma and detection of seed-borne rhizome. In *The 37th Kasetsart University Annual Conference, 3-5 February, 1999.* (pp. 295-302). Text and Journal Publication Co, Ltd.
18. Horita, M., Kobara, Y., Yano, K., Hayashi, K., Nakamura, Y., Iiyama, K. and Oki, T., 2023. Comprehensive Control System for Ginger Bacterial Wilt Disease Based on Anaerobic Soil Disinfestation. *Agronomy*, *13*(7), p.1791.
19. Rashedy, A., Shaban, A., El-Wakil, A.F., El-Hamid, A., Abd El-Hamid, A. and Abdel Moniem, E.A.H., 2023. Efficiency of Turmeric, Hibiscus and Beet Root Extracts as an Eco-Friendly Edible Coating to Extend storage Life of Hass Avocado Fruit. *Egyptian Journal of Chemistry*, *66*(6), pp.41-48.
20. Talpur, M.S.A., Abro, M.A., Jatoi, G.H., Marri, J.M., Poussio, G.B., Qayyum, M.M., Rajer, F.U. and Shafique, S., 2023. Exploring the effects of plant extracts, biological agents and essential oils on the non-chemical management of banana leaf spot disease. *Plant Protection*, *7*(2), pp.321-330.
21. Dethoup, T. and Jantasorn, A., 2023. Potential application of dry medicinal plant powders in controlling black spot on Chinese kale and chili anthracnose diseases. *Agriculture and Natural Resources*, *57*(2), pp.289-300.
22. Waskel, S., Dhuware, S.R., Agashe, R., Waskel, R. and Agashe, D.R., 2023. Impact of frontline demonstration on yield enhancement of turmeric in Balaghat district of Madhya Pradesh.