Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January-June- 2024 2024 5(1), 2205–2215

EXPLORATORY DATA ANALYSIS FOR LEAF DISEASE DETECTION

Revathy M[1], Dr. Karthikeyan Elangovan[2]

[1]Research Scholar (Part time Internal), Department of Computer and Information Science, Annamalai University, Chidambaram, India
[2]Research Supervisor, Assistant Professor/Programmer, (Deputed as Assistant Professor and Head in Government Arts and Science College, Gingee, Villupuram), Department of Computer and Information Science, Faculty of Science, Annamalai University, Annamalai Nagar, Tamilnadu, India

Corresponding Author
Revathy M,
revathimphil1213@gmail.com

DOI
10.29121/shodhkosh.v5.i1.2024.2277

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Agriculture plays a significant part in India due to their population growth and increased food demands. Hence, there is a need to enhance the yield of crop. One of these important effects on low crop yields is diseases caused by bacteria, fungi and viruses. This can be prevented and handled by means of applying plant disease detection approaches. Machine learning techniques will be employed in the process of disease identification on plants as it mostly applies information themselves and offers fabulous techniques for detection of plant diseases. Analysis of the sickness should be done appropriately and proper movements must be taken at the correct time. A correct detection of leaf disorder is crucial for plant culture as well as the rural financial system. Even though many works were executed for identifying leaf disease, due to the inadequate strategies additionally the obligations about classifying leaf disorder is difficult to be expecting This paper explores the Ada Boost with Naïve Bayes perform well as well it showing an efficient outcome. It has the greatest accuracy result of 85.75%. The Ada Boost with Naïve Bayes produces the greatest precision result of 0.86. The Ada Boost with Naïve Bayes and Stochastic Gradient Boost with Naïve Bayes produce the maximum recall of 0.86. The Ada Boost with Naïve Bayes has the greatest F-Measure result of 0.86. The Ada Boost with Naïve Bayes model has the highest MCC value of 0.65. The Ada Boost with Naïve Bayes model has the greatest kappa value of 0.66. The Ada Boost with Naïve Bayes model has an optimal results compare with other models.

Keywords: Ada Boost, F-Measure, Naïve Bayes, Stochastic Gradient Boost

1. INTRODUCTION

In recent years, studies associated with identification of sicknesses in crops have been an increasing number of pretentious a key position in enhancing the overall performance as well as competitiveness of trending agribusiness [1]. Plant sicknesses have a large impact on top of the rural efficiency. They are able to effortlessly corrupt the high-satisfactory of the merchandise [2]. Thus, it's far very vital to cure the plant sicknesses at early on otherwise better phases so that correct as well as appropriate achievement can be taken with the aid of using the farmers to keep away from in addition loss [3]. The detection is historically done with the aid of using human professionals. Human professionals become aware of sicknesses visually however they face a few problems which can damage their efforts [4,25]. Due to this difficulty as well as toward the huge range of cultivated flowers along with their present phytopathological problems, even skilled agronomists and plant pathologists frequently fail to efficaciously diagnose precise sicknesses, and are therefore caused flawed conclusions and treatments [5]. In the agriculture area one of the vital studies subjects is the popularity of plant disease. To keep away from the fatalities in the amount of agriculture merchandise and within side the yield, a vital key's the popularity of plant sicknesses [6].

There are distinctive kinds of sicknesses which subsist within the plants like fungal, bacterial, and viral and so forth. It's been located 85% plant life are suffering from fungal like organisms. Following are the a few primary statistics on bacterial, viral, fungal illnesses. Bacterial sicknesses: bacterial diseases named as microorganism reasons [7] distinctive varieties of signs that include overgrowths of plants, leaf spots, scabs and cankers. Bacterial infection signs are almost about comparable like fungal disorder. The most common form of signs found in bacterial sickness is leaf spot [8]. Viral illnesses: in the case of viral sicknesses, it's far little difficult to become aware of and analyze. Symptoms of viral disease are Mosaic leaf sample, Plant stunting etc. A few of the primary viral disorder Cauliflower mosaic virus and so forth. Fungal sicknesses: those are the sicknesses that are normally observed on wide variety of vegetables.

Fungal diseases are liable for a substantial harm on plant. Some of principal fungal illnesses are Downy mildews, Powdery mildews, Rusts, etc [9, 10]. Newly, deep learning has gained super achievement in lots of fields, inclusive of specifically within the agriculture area, like pest identification weed detection, fruit detection, tomato sickness' identity, cucumber sicknesses crop species and illnesses identity [11]. ResNet 50 (Residual Network) CNNs represent the maximum effective strategy for modeling complex strategies as well as performing image extraction in programs with large quantity of records, like the one in every of sample recognition in images [12,24].

1.1 TYPES OF LEAF DISEASES

Leaf diseases can be broadly classified into three categories based on the pathogens that cause them: fungal, bacterial, and viral diseases. Each of these categories has its own set of characteristics, symptoms, and modes of transmission.

- **FUNGAL DISEASES**: Fungal pathogens are the most common cause of leaf diseases, thriving in humid environments where they can spread quickly through spores. Some well-known fungal diseases include **powdery mildew**, **downy mildew**, **rusts**, and **leaf spot diseases**. Fungal infections usually present as spots, discoloration, or powdery coatings on the leaf surface, leading to wilting, yellowing, or premature leaf drop. If left unchecked, fungal diseases can spread rapidly across crops, devastating large areas of cultivation.

Figure 1: Unhealthy fungal diseased Leaves

- **BACTERIAL DISEASES**: Bacteria also infect leaves, typically through wounds, insect bites, or water. **Bacterial leaf blight** and **bacterial leaf spots** are common bacterial diseases that affect crops such as rice, citrus, and tomatoes. Bacterial infections often appear as water-soaked spots, which may enlarge and turn brown or black. In severe cases, bacterial pathogens can clog the plant's vascular system, disrupting the flow of nutrients and water.

Figure 2: Unhealthy Bacteria diseased Leaves

Figure 3: Healthy Leaves

- **VIRAL DISEASES**: Viruses infect plants and cause leaf diseases that often result in mosaic patterns, yellowing, or stunted growth. **Tobacco mosaic virus (TMV)** and **tomato yellow leaf curl virus (TYLCV)** are examples of viruses that can cause severe damage to crops. Viral infections are generally spread by insect vectors, such as aphids and whiteflies, or through contaminated tools.

1.2 CAUSES AND SYMPTOMS OF LEAF DISEASES

Leaf diseases can arise from various causes, including improper growing conditions, lack of nutrients, and the presence of pathogens. Environmental factors such as high humidity, overcrowding of plants, and poor air circulation create favorable conditions for the development and spread of pathogens.

The symptoms of leaf diseases vary depending on the pathogen but typically include spots, blisters, streaks, or necrotic (dead) areas on the leaves. These symptoms can interfere with the plant's photosynthetic processes, leading to reduced energy production and overall poor health. Diseased leaves often curl, wilt, or drop prematurely, weakening the plant and making it more susceptible to other stresses.

1.3 IMPACT ON AGRICULTURE

Leaf diseases can have devastating economic and ecological impacts, especially in agriculture where crops are grown on a large scale. A severe leaf disease outbreak can reduce crop yields significantly, affecting food supply and leading to financial losses for farmers. For instance, rice bacterial blight and wheat rusts have historically caused substantial damage to staple crops. Moreover, managing and controlling these diseases involves significant expenses related to the application of pesticides and fungicides.

1.4 DETECTION AND MANAGEMENT

Accurate and early detection of leaf diseases is vital to mitigate their impact. Traditional methods involve manual inspection of plants, which can be time-consuming and prone to errors. However, modern technology, particularly in machine learning and image processing, has revolutionized leaf disease detection. Automated systems now use digital imaging and AI algorithms to detect and classify diseases based on visual symptoms, enabling faster and more accurate diagnoses. Early detection allows farmers to implement effective control measures, such as targeted pesticide use, biological control agents, or cultural practices like crop rotation and proper spacing to reduce disease spread.

In conclusion, leaf disease detection and management are critical for maintaining healthy plants, ensuring food security, and reducing the economic burden on agriculture. Advances in technology, particularly in precision farming, are playing a crucial role in improving plant health monitoring and disease management strategies.

What follows is the rest of the paper's outline: The associated work is described in Section 2. The proposed technique is introduced in Section 3, followed by a brief overview of the results and discussion in Section 4. In Section 5, we wrap up the paper and its findings.

2. LITERATURE SURVEY

S. Ramesh et al. proposed system is used to detect the healthy and diseased leaf. The datasets is created and trained under the Random Forest to classify the healthy and diseased leaves. Once the image is inserted it undergoes several process and display the result. The final output shows both the healthy and diseased leaf of the particular image given. It displays the 70% accuracy level [13].

Arsenovic, Marko et al. executed in this system, a new dataset is created that containing 769,265 images was introduced with aim to create one of the largest dataset containing the leaf images. The trained model gives the accuracy level 93.67 of the image compared with the the dataset created.

Machine learning algorithms are used to provide various optimal decisions [14]. Ramcharan, Amanda, et al. presents the system detects the virus diseases of the cassava plant leaf. Image recognition gives both the cost effective and scalable technology for the disease detection. Thus, it displays the accuracy level for the leaflet of cassava dataset, then overall accuracy was higher and then ranged from 80% (20-70 split) KNN to 93.0% (80-10 split) SVM [15].

Kulkarni, Pranesh, et al. implemented system is a smart and efficient technique for the detection of the crop disease which uses the computer vision and machine learning algorithms. The proposed system can able to detect the 20 different diseases of the 5 common plants with the 93% accuracy.

Thus, the system can be used to detect the diseased plants in agricultural site as real time plant disease detection. It gives the average 93% accuracy and the 0.93 F1 score [16].Chohan, Murk, et al. performed a system proposes a deep learning based model named the plant disease detector and able to detect the several diseases from the plants using the pictures of their leaves. Plant Village is a dataset that is used to train the model. Thus 15% of the data from Plant Village data is used for the testing purpose that contains the images of healthy as well as the diseased plants. The proposed system has achieved the 98.3% of testing accuracy [17]. Pantazi, Xanthoula Eirini, Dimitrios Moshou, and Alexandra A. Tamouridou. demonstrated an approach of an automated way of crop disease identification on the various leaf sample images. A success rate of 95% is achieved for total of the 46 plant-condition combinations tested. The implemented application is capable of identifying the healthy conditions in the plant species [18]. Kumar, P. Manoj, C. M. Surya, and Varun P. Gopi. executed a system that explores the feature vectors from both front and back side of the green leaf along with the morphological features to improve at a unique optimum combination that maximizes the identification rate. Using classifier or matching algorithm the performance analysis is executed and display the name of the medicinal plant with its accuracy[19,26,27].

De Luna, Robert G., et al. proposed a system that involves the image processing techniques to extract features related to leaf in coexistence with using the artificial neural network in order to identify and detect some of the Philippine herbal plants. The experimental results gives a 98.61% accuracy of the herbal plant identification. From the system, the results obtained from the 72 new leaf samples, complete system provides the herbal leaf identification of 98.61% accuracy [20]. Ibrahim, Zaidah, Nurbaity Sabri, and NN Abu Mangshor. executed system gives the application of texture features for the leaf recognition of herbal plant identification. Then pre-processing is done on image for resizing and converting the image into grey-scale image. The features are extracted and leaf recognition is executed. Finally, it displays the name of the herbal plant with its accuracy level [21]. Kumar, Santhosh S., and B. K. Raghavendra. describes that they have done survey on the different plant's disease and the various advance techniques to detect the plant diseases. Disease in the plants that takes place when some of the virus, bacteria infects a plant and disturb its normal growth. Farmers need the fast and efficient techniques to detect all the types of diseases of the plants that can save time. These systems that can reduce the efforts and use of pesticides [22].

3. MATERIALS AND METHODS

we used public dataset for plant leaf disease detection called PlantVillage curated by Sharada P. Mohanty et Al. [23]. The dataset consists of 87000 RGB images of healthy and unhealthy plant leaves having 38 classes out of which We have selected only 25 classes for experimentation of our algorithm

4. METHODS

The following method are applied in this research work

- Borrowed dataset
- Data preprocessing
- Apply for Ensemble machine learning algorithms:
- Gradient Boosting Machine with Naïve Bayes (GBM with NB)
- Stochastic Gradient Boosting with Naive Bayes (SGD with NB)
- AB with NB(Ada with NB)
- Extreme Gradient Boost with Naive Bayes (XGB with NB)
- Light Gradient Boosting Machine with Naive Bayes(LGBM with NB)
- To get Optimal results
- Find a best Model

To produce an efficient result, these strategies were applied in python API. This study uses only 10% of the total dataset and uses tenfold cross validation for all categories.

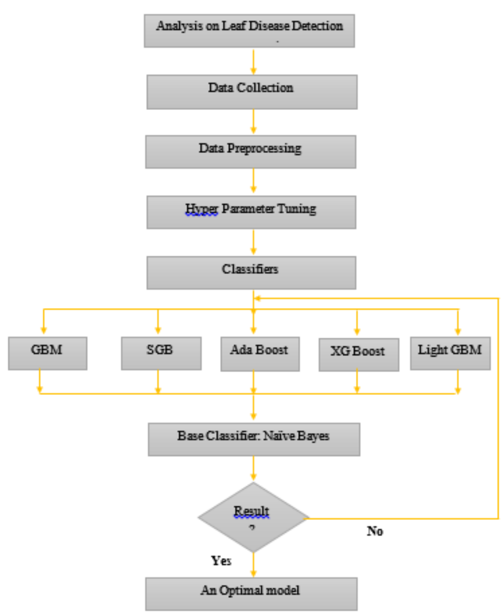

Figure 4: Proposed System

5. RESULTS AND DISCUSSIONS

Table 1: Models and their outcome

S.No	**Classifiers**	**Accuracy**	**Precision**	**Recall**
1. 1	**2. AB with NB**	**3. 85.75%**	**4. 0.86**	**5. 0.86**
6. 2	**7. XGB with NB**	**8. 84.10%**	**9. 0.84**	**10. 0.84**
11. 3	**12. LGB with NB**	**13. 83.85%**	**14. 0.83**	**15. 0.83**
16. 4	**17. GDM with NB**	**18. 79.02%**	**19. 0.81**	**20. 0.8**
21. 5	**22. SGB with NB**	**23. 85.00%**	**24. 0.85**	**25. 0.86**

The above table shows that the various selected ensemble classifiers.

The AB with NB has an accuracy level of 85.75%, a precision value of 0.86, a recall value of 0.86, an F-Measure value of 0.86, an MCC value of 0.65 and a kappa statistic value of 0.66.

The XGB with NB results in an accuracy level of 84.10%, a precision value of 0.84, a recall value of 0.84, an F-Measure value of 0.83, an MCC value of 0.56 and a kappa statistic value of 0.55.

The LGBM with NB produces accuracy level 83.85%, a precision value 0.83, recall value 0.83, an F-Measure value 0.84, an MCC value 0.56 and a kappa statistic value 0.58.

The GBM with NB produces a yield of 79.02% an accuracy, a precision value of 0.81, a recall of 0.80, an F-Measure of 0.79, an MCC of 0.54 and a kappa statistic of 0.54.

The SGB with NB results in an accuracy level of 85%, a precision value of 0.86, a recall value of 0.86, an F-Measure value of 0.84,an MCC value of 0.58 and a kappa statistic value of 0.59.

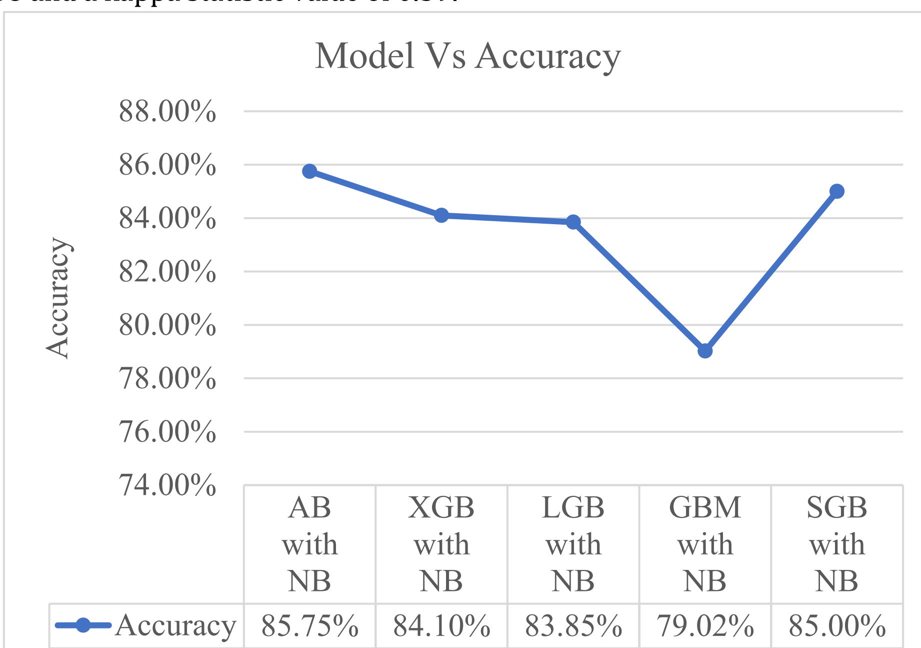

Figure 5: Performance of Ensemble classifiers with their accuracies

The above diagram shows that the accuracy performances of selected models. The AB with NB has the greatest accuracy result of 85.75%. The GBM with NB produces the lowest accuracy result of 79.02%. The accuracy of the LGBM with NB, XGB with NB, and SGB with NB is 83.85%, 84.10%, and 85%, respectively.

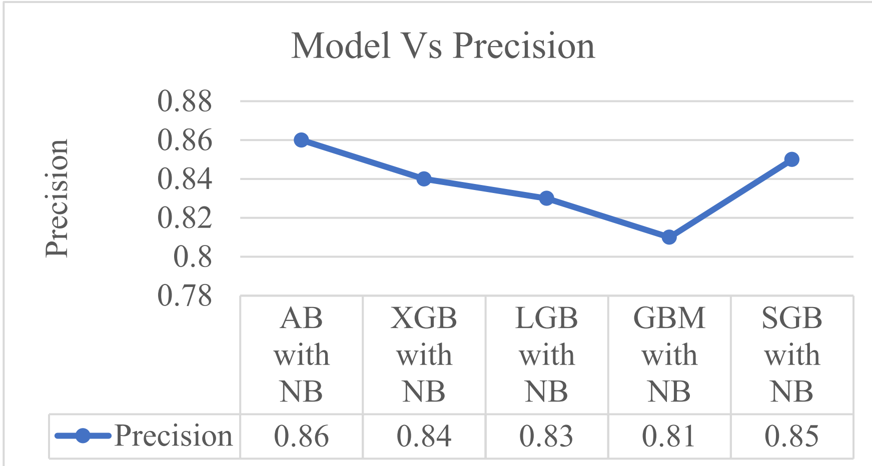

Figure 6: Performance of Ensemble Classifiers with their Precision values

The precision performances of selected models are depicted in the diagram above. The AB with NB produces the greatest precision result of 0.86. GBM with NB produces the lowest accuracy result of 0.81. The accuracy levels of the LGBM with NB, XGB with NB, and SGB with NB are 0.83, 0.84, and 0.85, respectively.

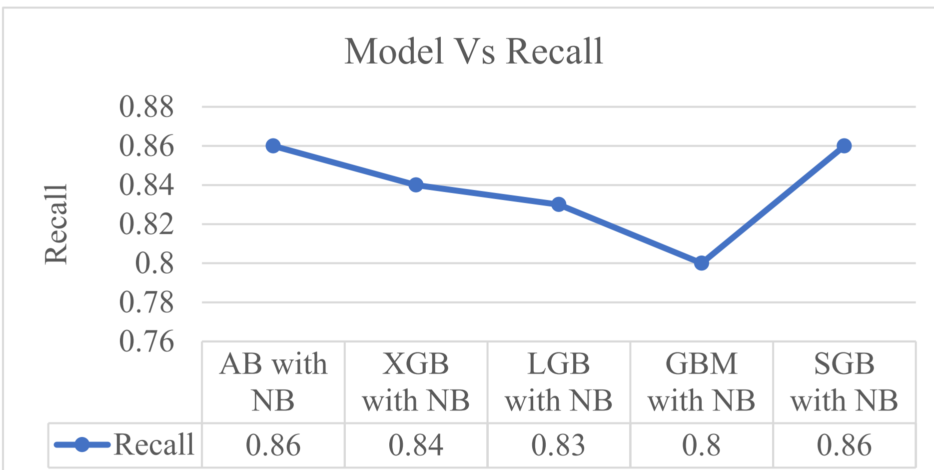

Figure 7: Performance of Ensemble Classifiers with their Recall values

The graph above depicts the recall performances of selected models. The AB with NB and SGB with NB produce the maximum recall of 0.86. GBM with NB produces the lowest recall result of 0.80. The recall levels for the LGBM with NB and the Extreme GBM with NB are 0.83 and 0.84, respectively.

Table 2: Models and their statistical outcome

S. No	Classifiers	F-Measure	MCC	Kappa
1	AB with NB	0.86	0.65	0.66
2	XGB with NB	0.83	0.56	0.55
3	LGB with NB	0.84	0.56	0.58
4	GDM with NB	0.79	0.54	0.54
5	SGB with NB	0.84	0.58	0.59

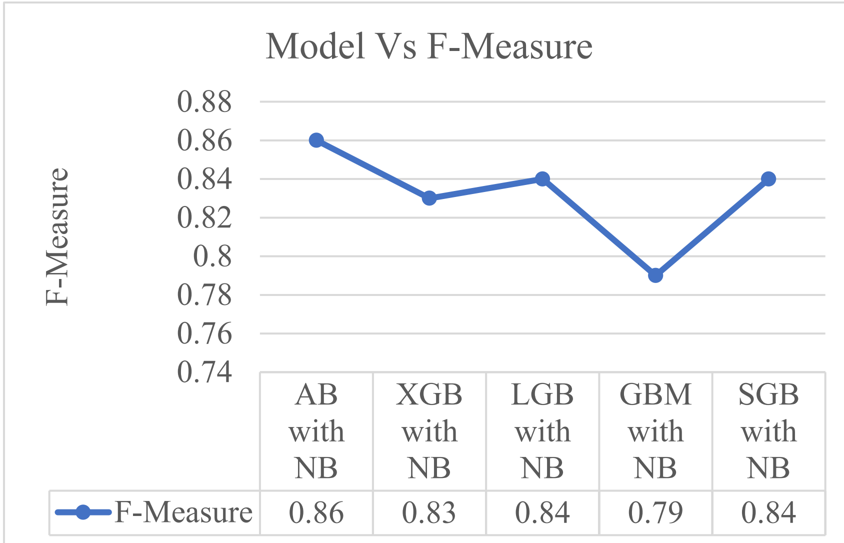

Figure 8: Performance of Ensemble Classifiers with their F-Measure values

The graph above depicts the F-Measure performances of selected models. The AB with NB has the greatest F-Measure result of 0.86. The GBM with NB produces the lowest F-Measure result of 0.79. The XGB with NB has an F-Measure of 0.83, whereas the LGB with NB and SGB with NB have the same value of 0.84.

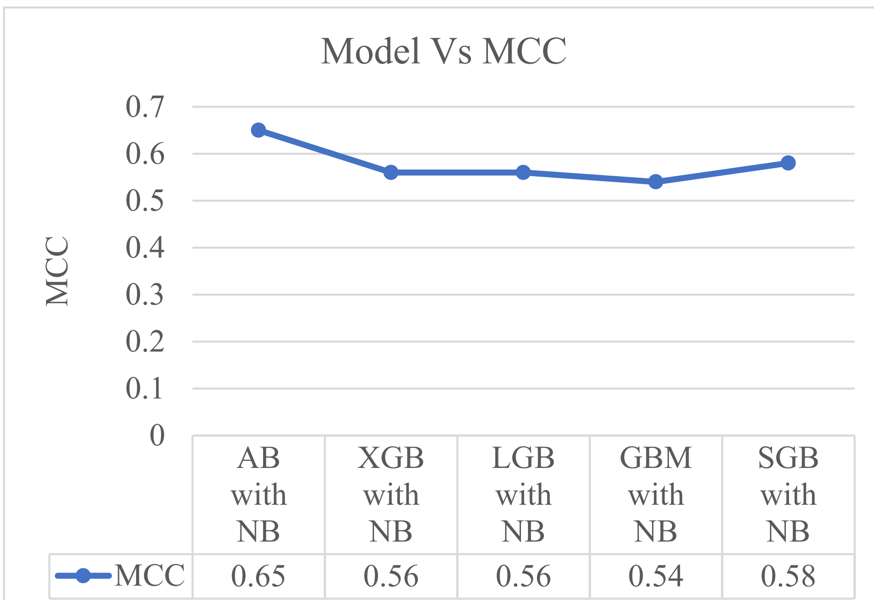

Figure 9: Performance of Ensemble Classifiers with their MCC values

The graphic above depicts the MCC performance of selected models. The AB with NB model has the highest MCC value of 0.65. GBM with NB produces the lowest MCC result (0.54). The remainder of the models, such as the XGB with NB model and the Light Gradient Boosting Machine with NB Decision Trees model, have the same MCC value of 0.56. The MCC value for SGB with NB 0.58.

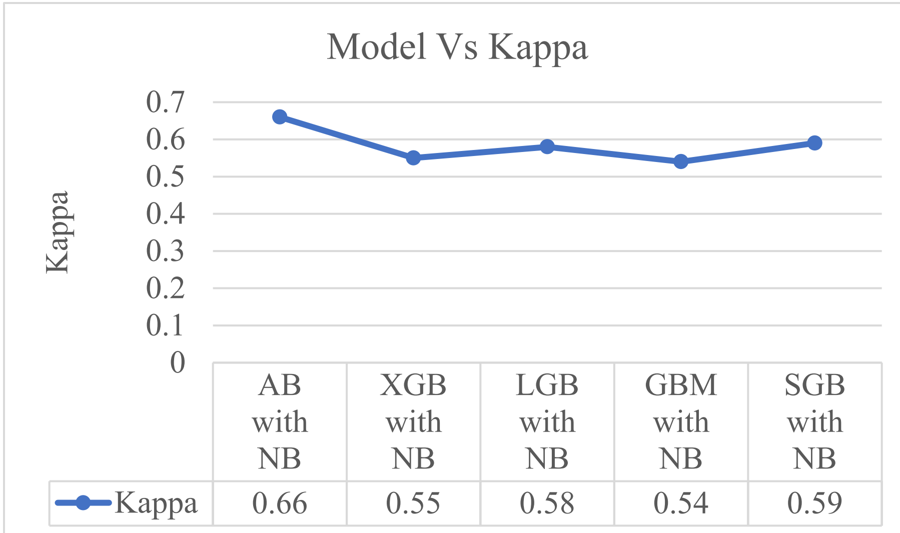

Figure 10: Performance of Ensemble classifiers with their Kappa statistic values

The graph above depicts the kappa value performances of selected models. The AB with NB model has the greatest kappa value of 0.66. The GBM with NB produces the lowest kappa result of 0.54. Other models with kappa values between 0.55 and 0.59 are Extreme Gradient Boost with NB, Light Gradient Boosting Machine with NB, and Stochastic Gradient Boosting with NB.

6. CONCLUSIONS

This work finds that the AB with NB has an F-Measure of 0.86, an MCC of 0.65, a kappa statistic of 0.66, a recall of 0.86, a precision of 0.86, and a precision of 0.86. The combined XGB and NB achieves an impressive 84.10% accuracy, 0.84 precision, 0.84 recall, 0.83 F-Measure, 0.56 MCC, and 0.55 kappa statistic. Accuracy of 83.85%, precision of 0.83%, recall of 0.83%, F-Measure of 0.84%, mean correlation coefficient of 0.56%, and kappa statistic of 0.58% are all generated by the LGBM with NB. Using NB, the GBM achieves a return of 79.02% accurate predictions, 0.81 precision, 0.80 recall, 0.79 F-Measure, 0.54 MCC, and 0.54 kappa. Accuracy is at 85%, precision is 0.86, recall is 0.86, the F-Measure is 0.84, the MCC is 0.58, and the kappa statistic is 0.59 when using the SGB with NB. The AB with NB has the greatest accuracy result of 85.75%. a precision result of 0.86, a recall of 0.86, an F-Measure result of 0.86, an MCC value of 0.65 and a kappa value of 0.66.This model recommends the AB with NB compare with other models.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

Araujo, J.M.M. and Peixoto, Z.M.A, “A new proposal for automatic identification of multiple soybean diseases”. Computers and electronics in agriculture, Vol. 167, pp. 1-9, 2019.

Barburiceanu, S., Meza, S., Orza, B., Malutan, R. and Terebes, R, “Convolutional Neural Networks for Texture Feature Extraction”. Applications to Leaf Disease Classification in Precision Agriculture. IEEE Access, Vol.9, pp.160085-160103, 2021.

Jadhav, S.B., Udup, V.R. and Patil, S.B, “Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier”. International Journal of Electrical and Computer Engineering, Vol. 9(5), pp.4077-4092, 2019.

Sardogan, M., Tuncer, A. and Ozen, Y., 2018, September. Plant leaf disease detection and classification based on CNN with LVQ algorithm. In 2018 3rd international conference on computer science and engineering (UBMK) (pp. 382-385). IEEE.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture, 145, pp.311-318.

Ramesh, S. and Vydeki, D., 2020. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. Information processing in agriculture, 7(2), pp.249-260.

Too, E.C., Yujian, L., Njuki, S. and Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture, 161.

Mukti, I.Z. and Biswas, D., 2019, December. Transfer learning based plant diseases detection using ResNet50. In 2019 4th International conference on electrical information and communication technology (EICT) (pp. 1-6). IEEE.

Fang, Y. and Ramasamy, R.P., 2015. Current and prospective methods for plant disease detection. Biosensors, 5(3), pp.537-561.

Jagadeesh d.Pujari, Rajesh Yakkundimdath,Abdulmunaf S.Byadgi, "Image processing Based detection of fungal diseases in plants.In proceedings on Information and Communication Technologies.Proceedia computer science,Vol,46, 2015.

Wu, Q., Zhang, K. and Meng, J., 2019. Identification of soybean leaf diseases via deep learning. Journal of The Institution of Engineers (India): Series A, 100(4), pp.659-666.

Kaushik, M., Prakash, P., Ajay, R. and Veni, S., 2020, June. Tomato leaf disease detection using convolutional neural network with data augmentation. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1125-1132). IEEE.

S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018, pp.41-45, doi: 10.1109/ICDI3C.2018.00017.

Arsenovic, Marko, et al. "Solving current limitations of deep learning-based approaches for plant disease detection." Symmetry 11.7 (2019): 939.85

Ramcharan, Amanda, et al. "Deep learning for image-based cassava disease detection." Frontiers in plant science 8 (2017): 1852.

Kulkarni, Pranesh, et al. "Plant Disease Detection Using Image Processing and Machine Learning." arXiv preprint arXiv:2106.10698 (2021).

Chohan, Murk, et al. "Plant disease detection using deep learning." International Journal of Recent Technology and Engineering 9.1 (2020): 909-914.

Pantazi, Xanthoula Eirini, Dimitrios Moshou, and Alexandra A. Tamouridou. "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers." Computers and electronics in agriculture 156 (2019): 96-104.

Kumar, P. Manoj, C. M. Surya, and Varun P. Gopi. "Identification of ayurvedic medicinal plants by image processing of leaf samples." 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2017.

De Luna, Robert G., et al. "Identification of philippine herbal medicine plant leaf using artificial neural network." 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM). IEEE, 2017.

Ibrahim, Zaidah, Nurbaity Sabri, and NN Abu Mangshor. "Leaf recognition using texture features for herbal plant identification." Indonesian Journal of Electrical Engineering and Computer Science 9.1 (2018): 152-156

Kumar, Santhosh S., and B. K. Raghavendra. "Diseases detection of various plant leaf using image processing techniques: A review." 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019.

Mohanty SP, Hughes DP and Salathé M (2016) Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7:1419. doi: 10.3389/fpls.2016.01419

Raj, S.., Shankar, G., Murugesan, S., Raju, M. N. ., Mohan, E., & Rani, P. J. I. (2023). Exploratory Data Analysis on Blueberry yield through Bayes and Function Models. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11s), 634–641. https://doi.org/10.17762/ijritcc.v11i11s.8299

Thivakaran, T. K. ., Priyanka, N. ., Antony, J. C. ., Surendran, S. ., Mohan, E. ., & Innisai Rani, P. J. . (2023). Exploratory Data Analysis for Textile Defect Detection. International Journal on Recent and Innovation Trends in Computing and Communication, 11(9s), 121–128. https://doi.org/10.17762/ijritcc.v11i9s.7403

Rani, P. J. I. ., Venkatachalam, K. ., Sasikumar, D. ., Madhankumar, M. ., A., T. ., Senthilkumar, P. ., & Mohan, E. . (2024). An Optimal Approach on Electric Vehicle by using Functional Learning . International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 197–206.

P.Jona Innisai Rani, Dr.C.Suresh Kumar , Dr. S.Venkatakrishnan. (2020). Brown Spot and Narrow Brown Spot Detection in paddy Using SVM Classifier. International Journal of Advanced Science and Technology, 29(7), 3615-3622.