

ORIGINAL RESEARCH ARTICLE

Advancements in apple disease classification: Machine learning models, IoT integration, and future prospects

Amit Kumar¹, Neha Sharma¹, Rahul Chauhan², Kamalpreet Kaur Gurna³, Abhineet Anand^{4,*},
Meenakshi Awasthi⁵

¹ Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura 140401, India

² Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun 248002, India

³ Department of Computer Science and Engineering, Baba Banda Singh Bahadur Engineering College, I.K.Gujral
Punjab Technical University, Jalandhar 144603, India

⁴ Department of Computer Science and Engineering, Chandigarh University, Mohali 140413, Punjab, India

⁵ Department of Electrical, Electronics & Communication Engineering, Galgotias University, Greater Noida 203201,
India

* Corresponding author: Abhineet Anand, abhineet.e13847@cumail.in

ABSTRACT

Apple orchards are of significant importance in the global agricultural sector, but they are vulnerable to a range of diseases that have the potential to cause diminished crop productivity and financial hardships. This manuscript investigates the utilization of machine learning methodologies, such as Logistic Regression, Neural Networks, and Random Forest, to classify three prevalent apple diseases: Blotch, Normal, and Rot Scab. The performance of these models is assessed using several assessment criteria, and confusion matrices are presented to aid in the prompt and precise detection of these diseases. This supports the implementation of efficient disease control strategies in apple orchards. By utilizing these ML models for the detection and treatment of diseases, not only augment agricultural productivity but also make a valuable contribution to sustainable agricultural practices by diminishing the necessity for excessive pesticide application. The experimental results indicates that Logistic Regression reflects the best performance as compared to other machine learning models taken into consideration using the different parameters. it obtained 90.6% of AUC and 65.7% of classification accuracy as compared to NN and Random Forest, which has achieved, 89.3%, 65.1%, 80.9% and 52.2.%, respectively.

Keywords: apple disease classification; machine learning; sustainable agriculture; environment-friendly practices; developing countries; food security

ARTICLE INFO

Received: 9 October 2023

Accepted: 24 October 2023

Available online: 12 March 2024

COPYRIGHT

Copyright © 2024 by author(s).

Journal of Autonomous Intelligence is
published by Frontier Scientific Publishing.

This work is licensed under the Creative
Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).
<https://creativecommons.org/licenses/by-nc/4.0/>

1. Introduction

The apple cultivation industry holds significant importance in the global agricultural sector, since it contributes significantly to food production and economic sustainability^[1]. However, apple orchards have persistent issues associated with a variety of illnesses that can greatly reduce both the quantity and quality of the crop. Fungal pathogens are responsible for the occurrence of prevalent maladies such as Blotch, Normal, and Rot Scab among crops^[2]. These conditions have garnered substantial attention from agricultural specialists and cultivators. Blotch, characterized by the manifestation of black lesions on leaves and fruit, is mostly ascribed to the fungal disease *Venturia inaequalis*^[3]. On the other hand, it should be noted that Rot Scab is accountable for the formation of brown scab-like lesions on apples, since it is derived from the fungus *Venturia pirina*^[4]. Conversely,

apples categorized as “Normal” exhibit no discernible indications of illness. The prompt and accurate identification of these illnesses plays a crucial role in enabling effective disease management approaches in apple orchards^[5]. The rapid application of efficient detection technologies facilitates timely intervention, hence reducing the spread of diseases and minimising agricultural productivity losses^[6]. Throughout the course of history, the primary approach to disease identification has largely focused upon the visual inspection performed by trained specialists. The aforementioned approach has been distinguished by its propensity for requiring a significant amount of time and its inherent subjectivity.

In contemporary times, significant progress has been made in the domains of machine learning and image analysis, offering a potentially feasible alternative. This study examines the application of machine learning techniques, including Logistic Regression, Neural Networks, and Random Forest, in the classification of apple diseases through the analysis of high-resolution pictures^[7]. **Figure 1** illustrates the method of categorizing apple diseases. The flowchart shown illustrates the comprehensive procedure for classifying Apple Diseases. This procedure encompasses various stages, namely data collection, annotation, preprocessing, model selection, training, evaluation, and deployment^[8]. The procedure described above offers a structured approach for the methodology employed in the investigation, which aims to properly classify apple diseases. This research endeavor has significant implications for the improvement of orchard management practices and the advancement of sustainable apple production^[9]. The study project aims to provide valuable insights for farmers and academics, resulting in benefits for the agricultural industry and consumers worldwide.

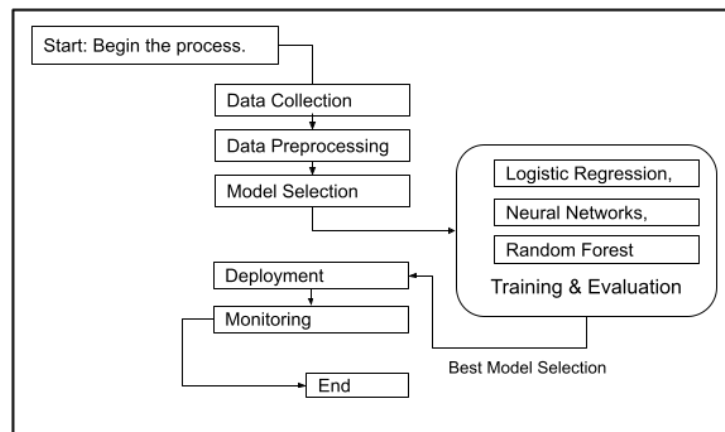


Figure 1. Apple disease classification process.

The research is based on previous research conducted in the domain of disease identification and categorization. Sharma et al.^[10] discuss the contemporary and future intelligent computationally techniques for multimedia data such as images. Shrivastava and Pradhan^[11] explore the rice leaf disease using different ML and DL models. Gupta et al.^[12] illustrate the stack ensemble approach to classify the post-COVID-19 complications. However, this paper sets itself apart through several significant aspects. Initially, it is important to note that the manuscript coverage transcends the limitations imposed by a certain crop or disease category. This work encompasses a range of agricultural diseases, with a specific emphasis on apple diseases such as Blotch, Normal, and Rot Scab. This expanded viewpoint enables a complete examination of disease detection methodologies that can be implemented across a wide range of crops. Secondly, it incorporates recent research and upcoming technologies that have not been addressed fully in earlier surveys. Moreover, this research expands beyond conventional methods of apple disease identification. The utilization of knowledge derived from contemporary research in the field of agricultural pathology is evident in this work, as exemplified by the incorporation of findings from studies, Aggarwal et al.^[13] discuss the federated transfer learning techniques employed to classify the rice-leaf disease. Aggarwal et al.^[14] illustrate the use of federated learning for classification of rice-leaf using the non-independent and identically distributed images. Additionally, this work

incorporates insights from Sethy et al.^[15] which discusses the pre-trained deep neural network and features selection machine learning models for the classification of rice leaf. The goal is to offer a holistic viewpoint on the utilization of traditional machine learning algorithms for the identification of apple diseases by integrating interdisciplinary results. This approach surpasses the limitations of conventional methodologies. This would offer readers valuable perspectives on the most recent advancements and patterns in disease diagnosis.

The motivation behind exclusive focus on “Apple” is rooted in its significant presence within the worldwide agriculture industry and the high incidence of diseases observed in apple orchards. The particularity of this characteristic allows for a thorough analysis of disease detection methods pertaining to a widely grown and economically important fruit, with the possibility of extending its relevance to other areas within the field of agriculture. The paper is organized as: Section 1 highlights the introduction; Section 2 illustrates the data sets and preprocessing of the data; Section 3 discusses the previous and current literature; Section 4 evaluates the experimental results obtained; Section 5 concludes and presents the scope of future work.

2. Data collection and preprocessing

The research involved the diligent collection of a complete dataset of high-resolution photographs of apples displaying different levels of three specific conditions:

2.1. Blotch apple

Blotch, caused by the fungal pathogen *Venturia inaequalis*, is a prevalent disease that affects apple trees. Apples affected by the condition known as Blotch may display a spectrum of symptoms, with the intensity of the ailment ranging from minor to significant. Certain apples may have minor, dispersed black lesions on their external surface, whilst others may display larger, more widespread lesions. **Figure 2** presents a visual representation of apples exhibiting varying degrees of Blotch severity, illustrating the diverse manifestations of the disease.

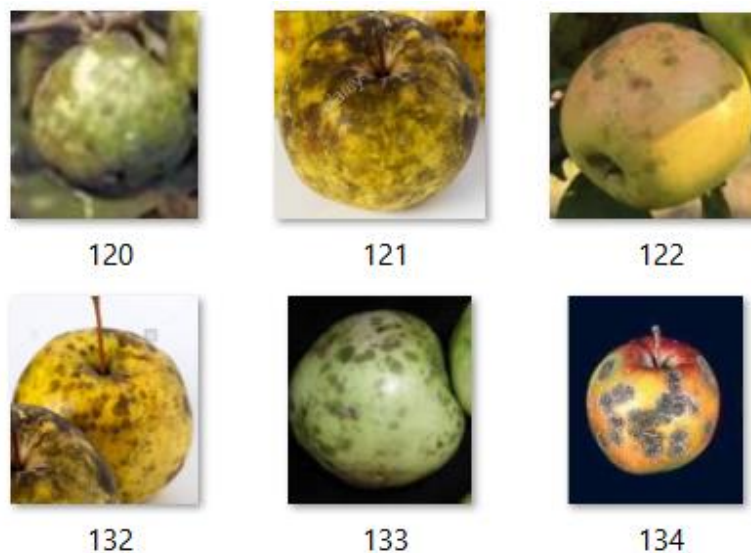


Figure 2. Blotch apple images.

2.2. Rot Scab apple

Rot Scab, a prevalent apple disease, is attributed to the fungal pathogen *Venturia pirina*^[16]. Similar to Blotch, the severity of Rot Scab can demonstrate fluctuation.

Apples afflicted with Rot Scab may display observable manifestations, characterized by the presence of brown scab-like lesions on their external surface. **Figure 3** displays a graphical depiction of the varying degrees

of Rot Scab observed on different apple specimens, hence showcasing the diverse symptoms of this disease.

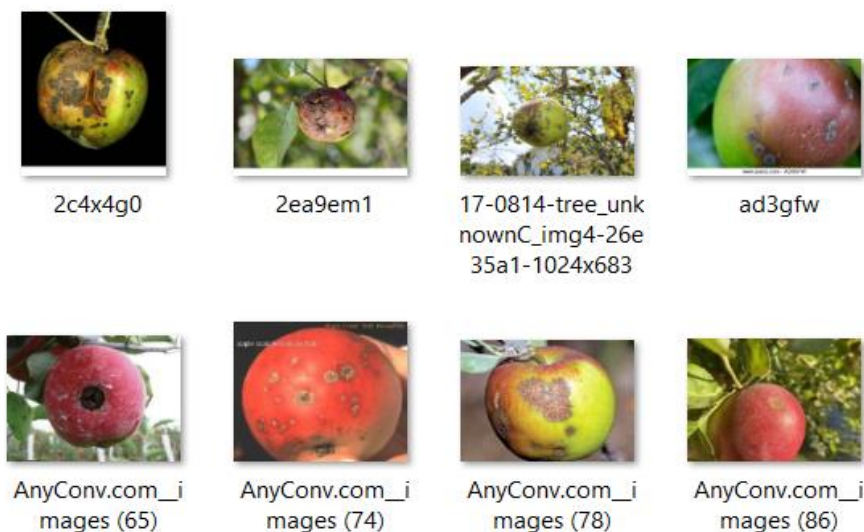


Figure 3. Rot/scab apple images.

2.3. Normal apple

In contrast to apples that are afflicted with diseases such as Blotch and Rot Scab, “Normal” apples are ones that do not manifest any symptoms of disease. The apples in question are widely regarded as being nutritious and devoid of any discernible imperfections. **Figure 4** typically comprises depictions of apples that exhibit a state of immaculacy, devoid of any signs of disease.



Figure 4. Normal apple images.

The application of a standardized area diagram is a widely used technique in the discipline of plant pathology to assess the magnitude of disease severity on plants or fruits. The present methodology presents a systematic framework for assessing the extent to which a particular disease affects the overall surface area. The aim of this specific situation is to produce standardized area diagrams in order to assess the severity of two apple diseases, specifically sooty blotch and flyspeck.

The objective of this research is to investigate how a standardized area diagram may be used to assess the severity of sooty blotch and flyspeck infections on apple fruit. Using a typical area diagram in the lateral perspective, the severity of the sooty blotch and flyspeck infections in apple fruit was estimated. The image is made up of a collection of reference photos showing various levels of disease severity, from 0% (which indicates the absence of the disease) to 100% (which represents total covering). Observers compare the physical surface of the fruit with a series of reference images in order to assess the level of harm that each

illness has caused to the lateral surface of the fruit. This methodology offers a numerical assessment of the severity of the illness within this particular context. When assessing the severity of sooty blotch and flyspeck diseases in apple fruit from a top perspective, a standardized area diagram is used (see **Figure 5**). The illustration provided in this context includes visual cues that serve as references and demonstrate different stages of disease coverage. In order to estimate the part of the fruit's upper surface area that is affected by each ailment, observers do a visual assessment by comparing the physical appearance of the fruit's upper surface to specified benchmarks. This methodology enables a thorough evaluation of the severity of the disease from an alternative standpoint. Area diagrams, which adhere to established standards, are valuable tools for researchers and agricultural specialists to precisely measure the severity of sooty blotch and flyspeck infections on apple fruits. The implementation of standardized procedures enables the constant evaluation of disease severity, hence enhancing disease management practices within apple orchards. The objective was to curate a dataset that encompasses the entire range of disease severity by gathering photos of apples at different phases of disease progression, as well as disease-free apples. The utilization of this methodology facilitated the acquisition of correct discernment and classification of various disease stages and healthy apples by our machine learning models. Additionally, it offered a more accurate depiction of the difficulties encountered in actual apple orchards, where the intensity of diseases might exhibit significant variation across individual trees. The procedure of acquiring the information played a crucial part in the training and evaluation of our machine learning models for disease classification.

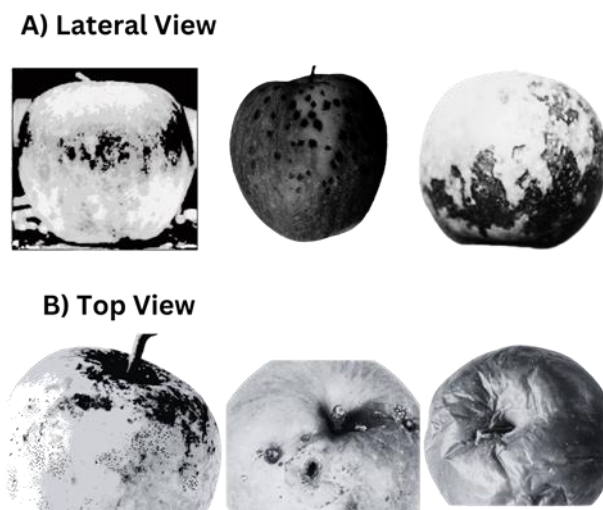


Figure 5. Sooty blotch and flyspeck severity in two fruit views: (a) lateral view, (b) top view.

3. Literature review

This paper employed a thorough methodology to address the classification of four prevalent apple diseases, namely Blotch, Normal, Rot and Scab. In order to achieve this objective, three independent machine learning algorithms^[17] and techniques^[18] are used: Logistic regression Neural Network and Random forest, each possessing its own unique strengths and traits.

3.1. Logistic regression

Logistic regression is a commonly employed and foundational classification technique renowned for its straightforwardness and interpretive capabilities. Notwithstanding its straightforwardness, this approach can yield unexpectedly favorable outcomes, especially in cases when the data can be easily separated into distinct categories or when the number of characteristics is relatively limited. Within the scope of our investigation, the utilization of Logistic Regression was employed as the foundational model, offering a direct and comprehensible resolution to the task of classifying apple diseases^[19,20]. The present model utilizes the logistic function in order to assess the likelihood associated with an apple being classified under a specific illness

category. Logistic Regression is a suitable method for binary and multiclass classifications due to its ability to determine the appropriate coefficients for the features involved in the task.

3.2. Neural networks

Neural networks, specifically convolutional neural networks (CNNs), have significantly transformed the field of image classification. Convolutional neural networks (CNNs) demonstrate exceptional proficiency in capturing complex patterns and hierarchical structures of features inside images, rendering them the preferred option for tasks pertaining to image analysis. The current study employed convolutional neural networks (CNNs) to automatically extract and obtain relevant visual information. The utilization of transfer learning was implemented to harness the capabilities of pre-trained convolutional neural network (CNN) architectures. The previously mentioned structures were then modified and enhanced using the dataset primarily pertaining to diseases that impact apple harvests. The application of this methodology enables the model to effectively discern intricate disease patterns, borders, and textures observed in images of apples. Consequently, it emerges as a potent instrument for classifying disorders.

3.3. Random Forest

The Random Forest methodology is widely recognised in the domain of machine learning owing to its robustness and capacity to effectively handle datasets characterized by a high number of dimensions. The system consists of a collection of decision trees, wherein each tree is trained on a randomly selected subset of both the dataset and the features. The utilization of Random Forest^[21], a technique that combines predictions from many decision trees, is implemented in order to mitigate overfitting and enhance the model's capacity for generalization. In our investigation, the Random Forest technique was employed, which involved the generation of a forest consisting of a collection of decision trees. The grouping of trees together determines the classification of illness for each apple. Although Random Forest may not possess the same level of intricacy as deep learning models such as Convolutional Neural Networks (CNNs), its capacity to effectively manage intricate and nonlinear associations within the data renders it a viable model worth investigating. The objective of this study was to conduct a detailed analysis on the efficacy of three distinct machine learning algorithms in the categorization of apple illnesses. The study examined various methodologies for attaining precise disease identification in apple orchards, leveraging the straightforwardness of Logistic Regression, the image analysis skills inherent in Convolutional Neural Networks (CNN), and the resilience of Random Forest. Hence, this research study provides a significant contribution towards the development of effective techniques for disease management.

4. Experimental results

The assessment of the performance of each model entailed the application of diverse metrics to ascertain its ability to differentiate between the distinct disease categories, as illustrated in **Table 1**. The assessment results indicate that Logistic Regression demonstrated superior performance when compared to the other models. The aforementioned findings are substantiated by the metrics acquired, which consist of an Area Under the Curve (AUC) value of 0.906, a Classification Accuracy (CA) value of 0.657, and an F1 Score of 0.642. The model was found to have an accuracy value of 0.632 and a recall value of 0.657, respectively. Based on these findings, it appears that the categorization abilities of the model are very precise. The Matthews Correlation Coefficient (MCC) received a score of 0.585, which indicates that it is effective in effectively resolving imbalanced data and is efficient in doing so. The findings indicate that the performance of Neural Networks was equivalent to that of Logistic Regression, as demonstrated by an Area Under the Curve (AUC) value of 0.893, a Classification Accuracy (CA) value of 0.651, and an F1 Score of 0.641. Additionally, the results demonstrate that the accuracy of Neural Networks was higher than that of Logistic Regression. The figures for Precision and Recall, in that order, were 0.632 and 0.651 respectively. In addition, the Matthews

Correlation Coefficient, also known as the MCC, was calculated and found to have a value of 0.579. The findings of this inquiry suggest that a high level of accuracy can be achieved when classification tasks are carried out. In contrast, the Random Forest model demonstrated somewhat lower performance metrics, as seen by an Area Under the Curve (AUC) value of 0.809, a Classification Accuracy (CA) value of 0.522, and an F1 Score of 0.494. These metrics measure how well the model is able to classify the data. The system's Precision and Recall values were 0.478 and 0.522, respectively, while its Matthews Correlation Coefficient (MCC) was observed to be the lowest at 0.416. The results presented in this study highlight the remarkable effectiveness of Logistic Regression and Neural Networks in accurately classifying apple illnesses, hence confirming their importance as excellent tools for disease management in apple orchards. This research uses Adams optimizer for refining model parameters and elevating classification performance. The usage of this technique is pivotal in attaining the reported outcomes.

Table 1. Model performance metrics.

Model	AUC	CA	F1 Score	Precision	Recall	MCC
Logistic Regression	0.906	0.657	0.642	0.632	0.657	0.585
Neural Networks	0.893	0.651	0.641	0.632	0.651	0.579
Random Forest	0.809	0.522	0.494	0.478	0.522	0.416

5. Discussion

5.1. Confusion matrices: Logistic regression (LR)

Within the context of the confusion matrix shown in **Figure 6**, the rows in the dataset correspond to the distinct illness groups, namely Blotch, Normal, Rot and Scab. The columns in the Logistic Regression model represent the disease types that have been predicted. The numerical values within the individual cells of the table correspond to the frequency of samples that have been classified into specific categories according to the predictions made by the model.

		Predicted								Σ
		Test\Bl...	Test\N...	Test\Ro...	Test\Sc...	Train\B...	Train\...	Train\R...	Train\S...	
Actual	Test\Bl...	12	1	0	0	13	0	0	4	30
	Test\N...	1	13	0	0	1	8	1	0	24
	Test\Ro...	1	0	5	0	1	0	29	2	38
	Test\Sc...	0	0	0	2	2	3	4	17	28
	Train\B...	5	0	0	0	102	0	3	6	116
	Train\...	0	5	0	1	2	57	1	1	67
	Train\R...	0	1	25	1	1	1	84	1	114
	Train\S...	1	2	2	11	6	2	6	55	85
Σ		20	22	32	15	128	71	128	86	502

Figure 6. Confusion matrices: logistic regression (LR).

5.2. Neural network confusion matrix

The confusion matrix depicted in **Figure 7** represents the dataset's rows, which are associated with distinct sickness groups, specifically Blotch, Normal, Rot, and Scab. The columns within the Neural Network model correspond to the disease categories that have been forecasted. The elements contained within the matrix denote the numerical representation of the samples that have been classified and grouped into their appropriate classes. The confusion matrix offers a comprehensive analysis of the classification performance of the model for each individual class. This analysis aids in the identification of classes that are more prone to confusion by

the model, hence offering valuable insights into the model's strengths and limitations.

		Predicted								Σ
		Test\Bl...	Test\N...	Test\Ro...	Test\Sc...	Train\B...	Train\...	Train\R...	Train\S...	
Actual	Test\Bl...	12	1	0	0	13	0	0	4	30
	Test\N...	1	13	0	0	1	8	1	0	24
	Test\Ro...	1	0	5	0	1	0	29	2	38
	Test\Sc...	0	0	0	2	2	3	4	17	28
	Train\B...	5	0	0	0	102	0	3	6	116
	Train\...	0	5	0	1	2	57	1	1	67
	Train\R...	0	1	25	1	1	1	84	1	114
	Train\S...	1	2	2	11	6	2	6	55	85
Σ		20	22	32	15	128	71	128	86	502

Figure 7. Confusion Matrices: Neural Network (NN).

5.3. Random Forest confusion matrix

The Confusion Matrix shown in **Figure 8** provides an evaluation of the Random Forest model's efficacy in accurately classifying various apple illnesses. The illness categories (Blotch, Normal, and Rot Scab) are shown by the rows, while the anticipated classifications by the model are represented by the columns. The numerical values within the matrix represent the frequency of samples that belong to specific categories.

		Predicted								Σ
		Test\Bl...	Test\N...	Test\Ro...	Test\Sc...	Train\B...	Train\...	Train\R...	Train\S...	
Actual	Test\Bl...	8	0	1	0	18	1	1	1	30
	Test\N...	0	4	0	2	1	8	7	2	24
	Test\Ro...	0	0	3	0	4	4	24	3	38
	Test\Sc...	0	0	0	0	7	2	4	15	28
	Train\B...	4	1	0	0	88	2	11	10	116
	Train\...	1	3	1	0	3	44	7	8	67
	Train\R...	1	0	15	1	9	8	71	9	114
	Train\S...	2	0	2	7	16	8	9	41	85
Σ		16	8	22	10	146	77	134	89	502

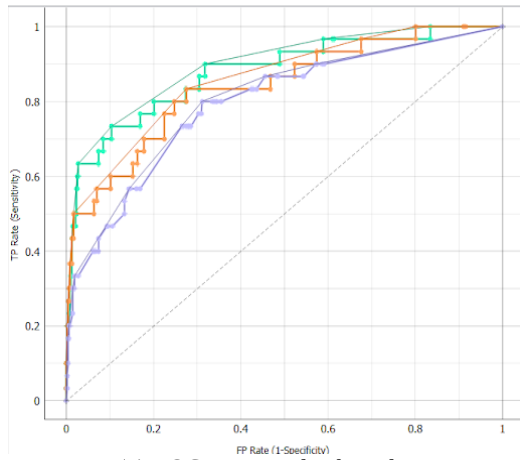
Figure 8. Confusion matrices: Random Forest (RF).

The Confusion Matrix offers significant insights into the performance of the model, namely in terms of accurately categorizing different disease types and the occurrences of misclassification.

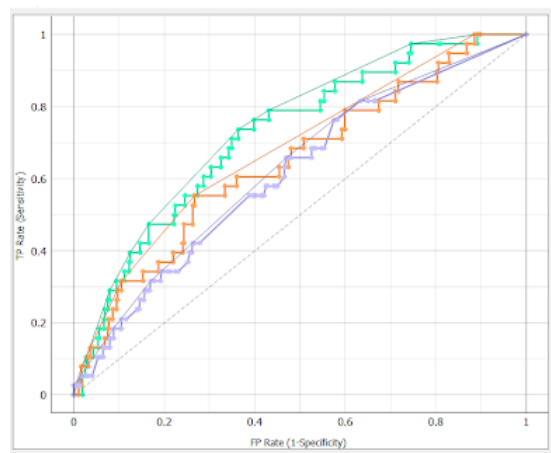
5.4. ROC curve—Test/train (Blotch, Rot, Scub, Normal) apple classification

The Receiver Operating Characteristic (ROC) curves in **Figures 9** and **10** illustrate the discriminatory performance of the Blotch, normal Rot, and scab apple test and train model in distinguishing between samples of Blotch, normal Rot, and scab apples. In **Figure 9a–d**, the graph illustrates the relationship between the true positive rate, also known as sensitivity, and the false positive rate, which is the complement of specificity, when the classification threshold undergoes variation. A greater value of the area under the receiver operating characteristic curve (AUC) signifies enhanced discriminatory ability. In the present scenario, the AUC value for blotch, rot, scab, and normal classification indicates that the model demonstrates robust performance in distinguishing blotch-affected apples from healthy ones. The capacity to effectively discriminate is of utmost

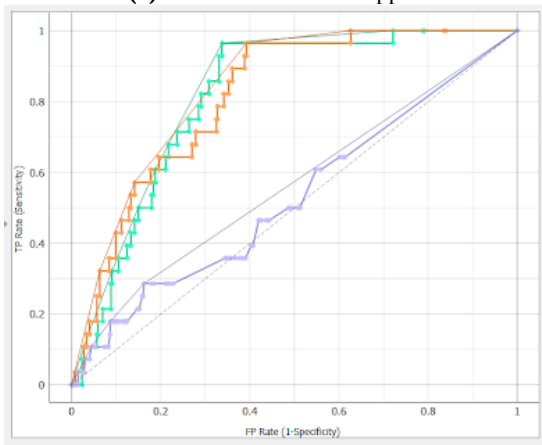
importance in the timely management of diseases in apple orchards. The lift curve is a visual tool employed in predictive modeling and classification tasks to evaluate the performance of a model relative to random chance. The evaluation process aids in assessing the efficacy of a model in effectively discerning between good and negative consequences.



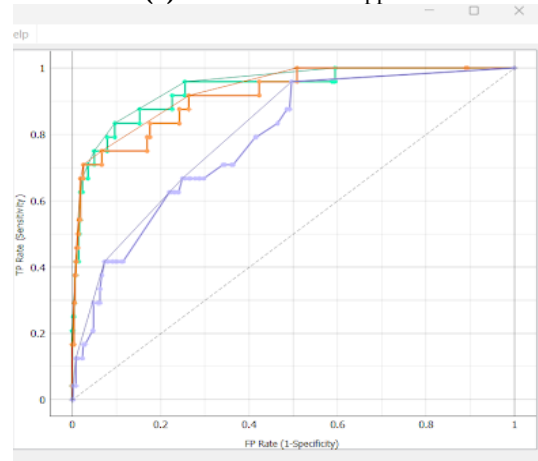
(a) ROC curve—Blotch apple.



(b) ROC curve—Rot apple.

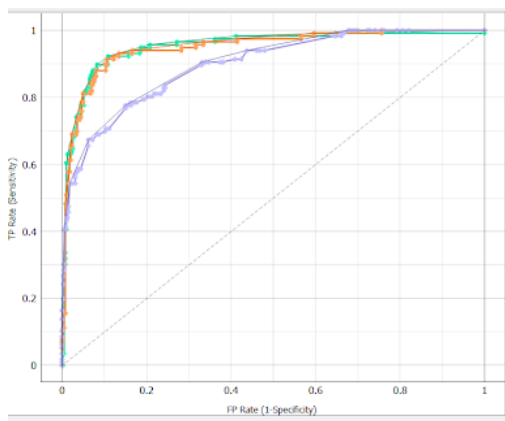


(c) ROC curve—Scab apple.

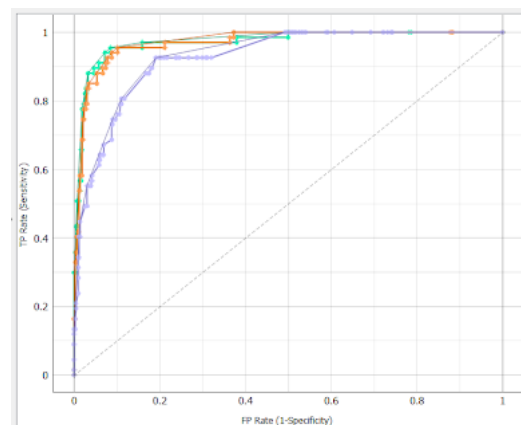


(d) ROC curve—Rot apple.

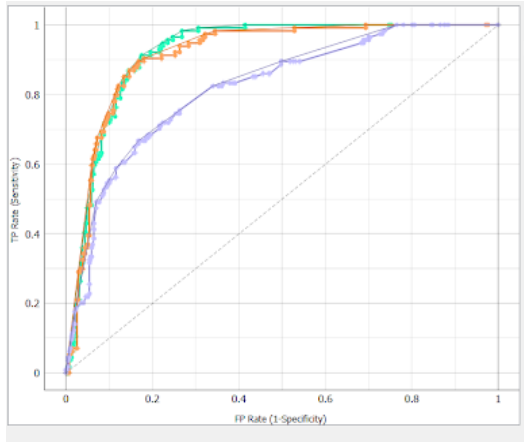
Figure 9. ROC Curve—Test/Blotch, Rot, Scub, normal apple classification.



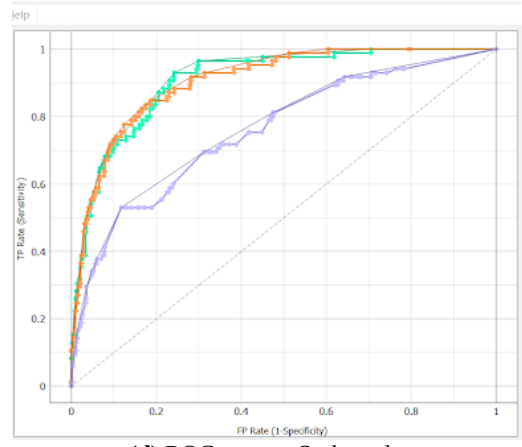
(a) ROC curve—Blotch apple.



(b) ROC curve—Normal apple.



(c) ROC curve—Rot apple.

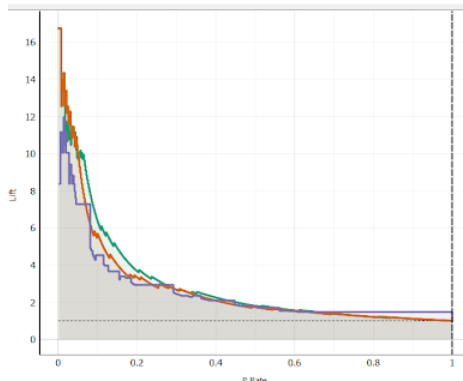


(d) ROC curve—Scab apple.

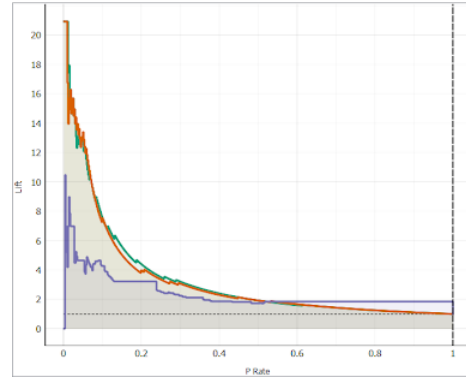
Figure 10. ROC curve—train/Blotch, Rot, Scub, normal apple classification.

5.5. Performance curve: Lift curve—Test/Train (Blotch, Rot, Scub, Normal) apple

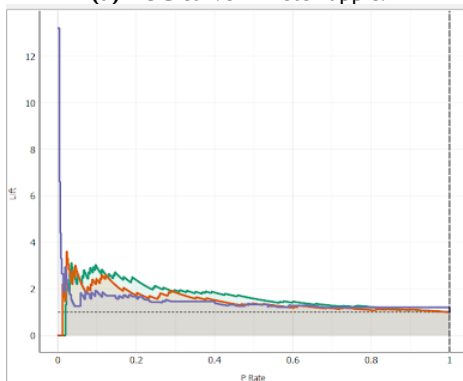
The lift curve provides a visual representation of the model's performance **Figure 11a–d** and **Figure 12a–d**, illustrating the ratio of the model's performance to that of random chance at various thresholds. A lift curve characterized by a pronounced upward slope signifies a robust model, whereas a flat curve implies that the model's performance is no more effective than random chance. The utilization of this curve holds significant importance in comprehending the model's capacity to effectively prioritize and differentiate positive cases. This, in turn, aids decision-making processes and facilitates the allocation of resources across diverse domains, including but not limited to marketing and fraud detection.



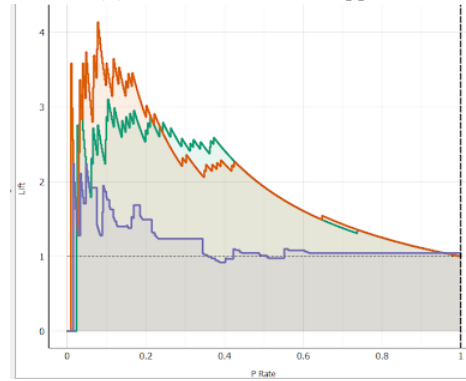
(a) ROC curve—Blotch apple.



(b) ROC curve—Blotch apple.

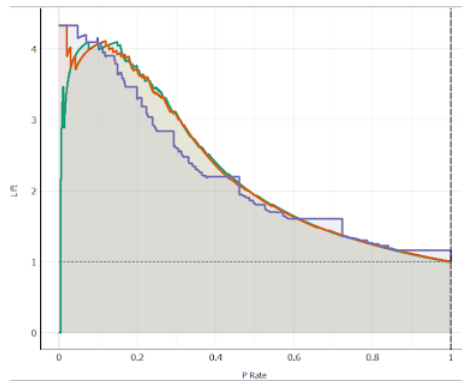


(c) ROC curve—Blotch apple.

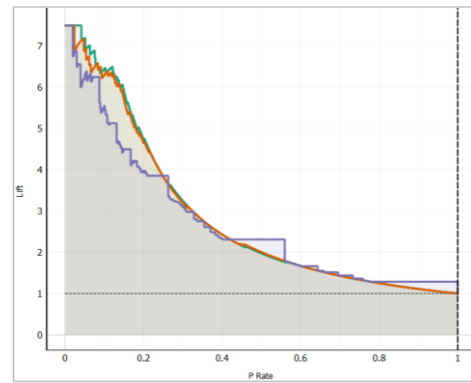


(d) ROC curve—Blotch apple.

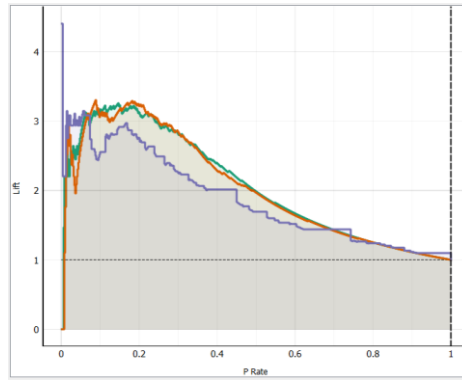
Figure 11. Performance curve: lift curve (Test/Blotch, Rot, Scub, Normal apple classification).



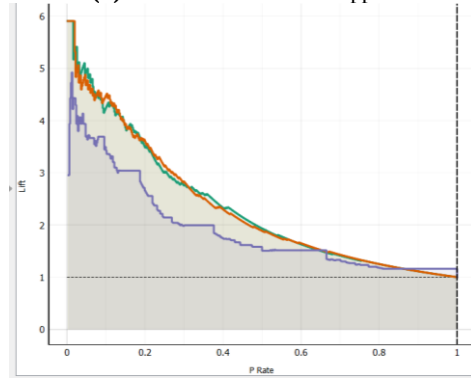
(a) ROC curve—Blotch apple.



(b) ROC curve—Normal apple.



(c) ROC curve—Rot apple.

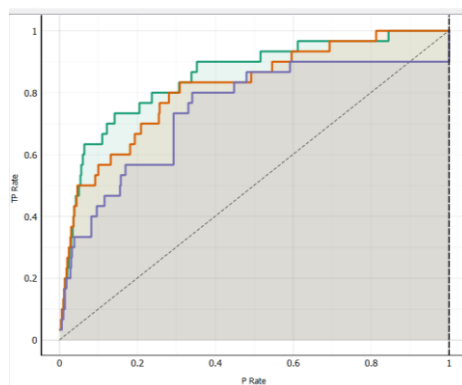


(d) ROC curve—Scab apple.

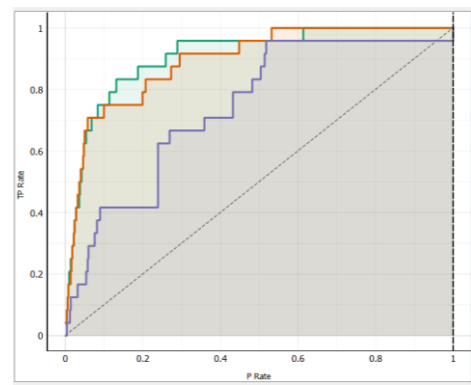
Figure 12. Performance curve: lift curve (Train/Blotch, Rot, Scub, Normal apple classification).

5.6. Cumulative gains

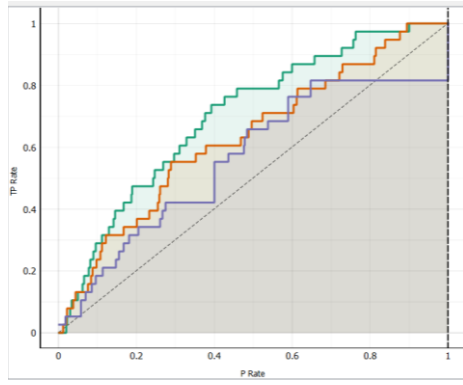
The Cumulative Gains technique is a visual aid employed in the paper of data and predictive modeling to evaluate the effectiveness of models, specifically in the context of binary classification tasks. **Figure 13a–d** and **Figure 14a–d** demonstrate the superior predictive performance of a model in comparison to random chance. By graphing the cumulative proportion of positive outcomes versus the cumulative percentage of total samples, one can assess the model's capacity to effectively prioritize or rank positive examples. This information possesses significant value in the context of decision-making, particularly in scenarios such as marketing campaigns, where the identification of prospects with high potential is of utmost importance.



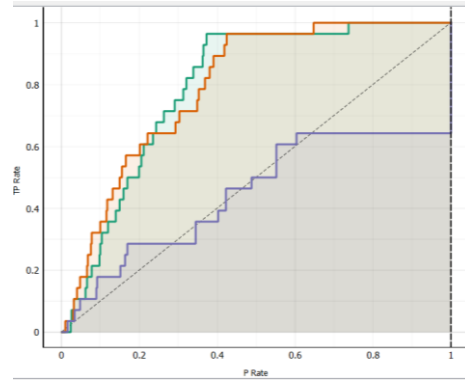
(a) Blotch apple.



(b) Normal apple.

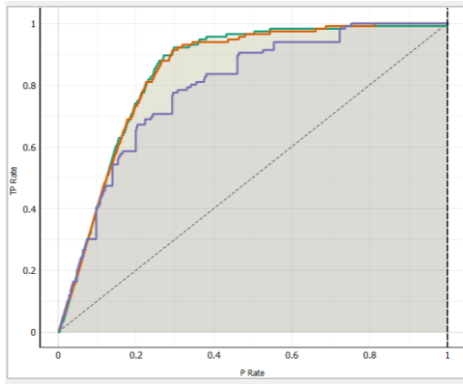


(c) Rot apple.

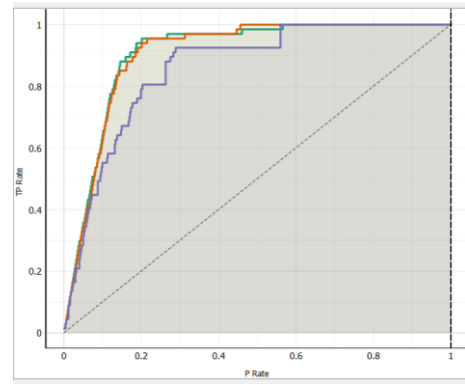


(d) Scab apple.

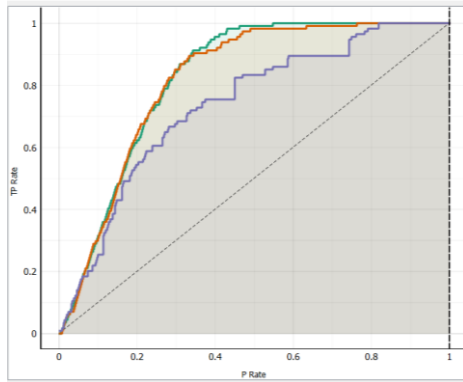
Figure 13. Cumulative gains (Test data/Blotch, Rot, Scub, Normal apple classification).



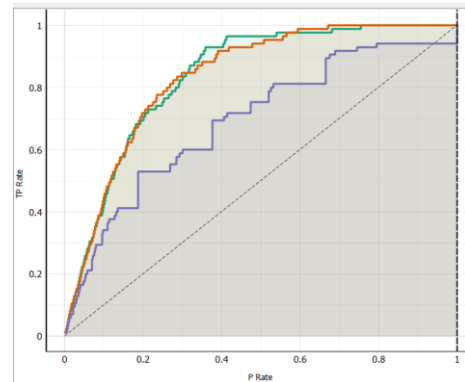
(a) Blotch apple.



(b) Normal apple.



(c) Rot apple.

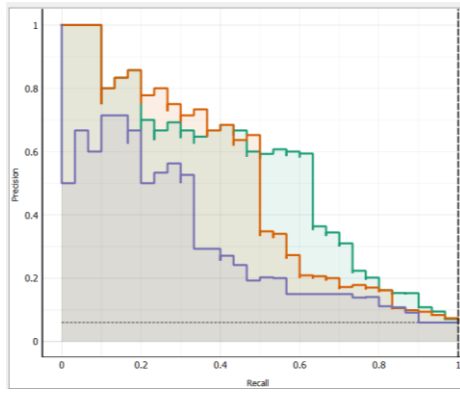


(d) Scab apple.

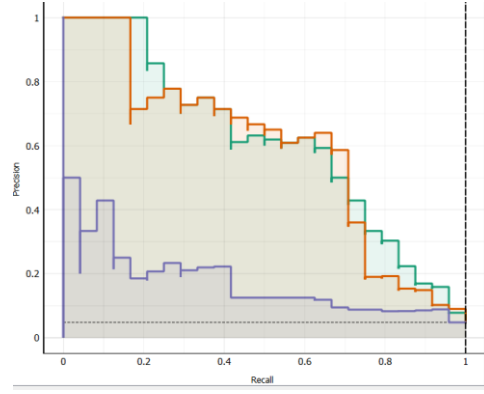
Figure 14. Cumulative gains (Train Data/ Blotch, Rot, Scub, Normal apple classification).

5.7. Precision-recall

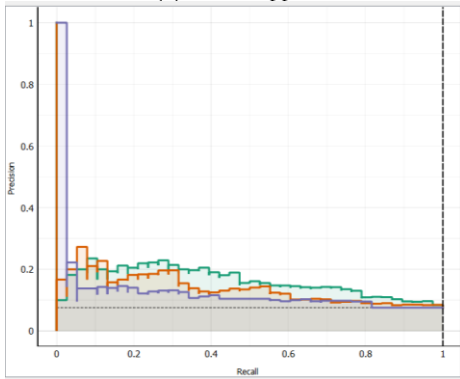
During the precision-recall analysis conducted for the classification of apple diseases, specifically Blotch, Rot Scab, and Normal, the following observations were made shown in **Figure 15a–d** and **Figure 16a–d**. The precision and recall of the Blotch, Normal, Rot and Scab categories were found to be relatively balanced, suggesting that the models demonstrated a satisfactory capacity to accurately detect and classify occurrences of these diseases. The category labeled as “Normal” had similar levels of precision and recall, indicating successful categorization of apples that are in a healthy state. The findings of this paper highlight the overall effectiveness of the models in accurately differentiating between various disease classes and identifying healthy apples. This is crucial for effectively managing diseases in apple orchards.



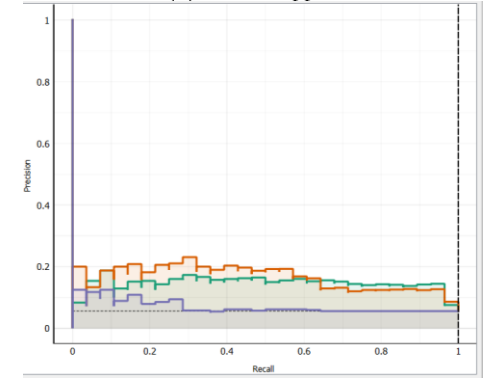
(a) Blotch apple.



(b) Normal apple.

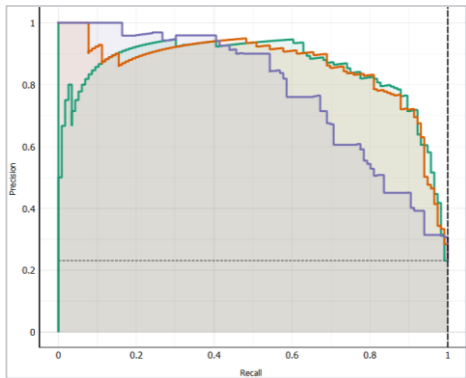


(c) Rot apple.

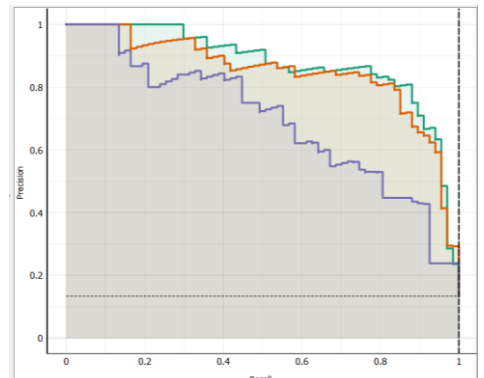


(d) Scab apple.

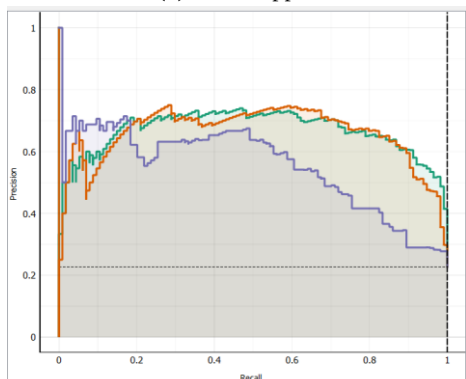
Figure 15. Precision-recall (Test Data/ Blotch, Rot, Scub, Normal apple classification).



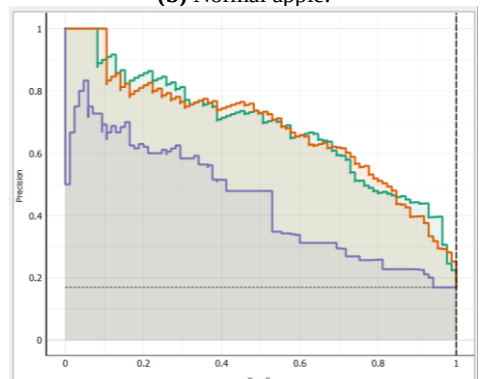
(a) Blotch apple.



(b) Normal apple.



(c) Rot apple.



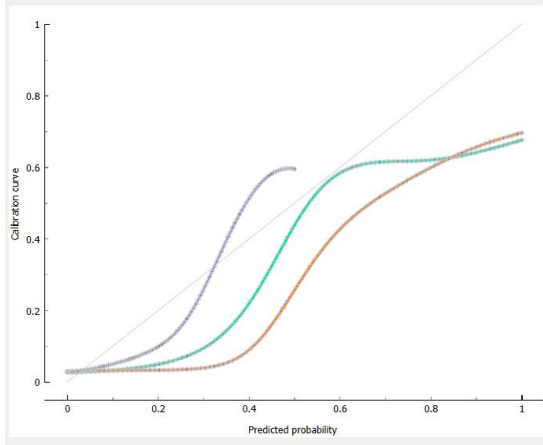
(d) Scab apple.

Figure 16. Precision-recall (Train Data/ Blotch, Rot, Scub, Normal apple classification).

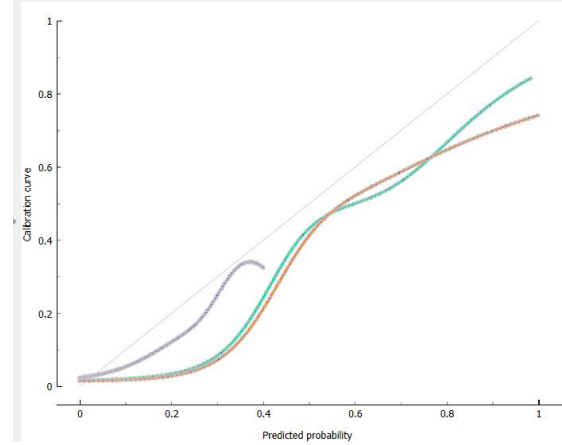
5.8. Calibration plot (Sigmoid Calibration)

A calibration plot, commonly known as a Sigmoid Calibration plot, **Figure 17a–d** and **Figure 18a–d** is

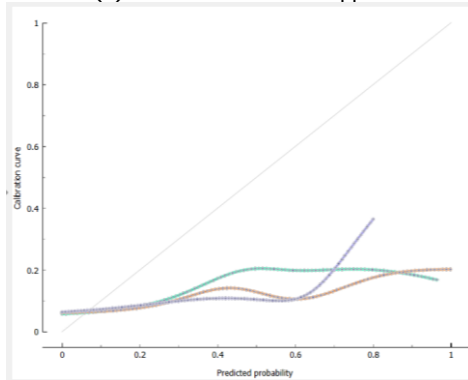
a visual depiction utilized for evaluating the efficacy of a binary classification algorithm. This aids in the visualization of the degree to which the expected probability corresponds with the observed outcomes. The x-axis in this plot normally denotes the anticipated probability, while the y-axis represents the observed results, which are commonly binary, with 0 indicating negative outcomes and 1 indicating positive outcomes. The optimal calibration curve should have a linear pattern, demonstrating a precise correspondence between anticipated and actual probability. Deviances from the aforementioned diagonal line can potentially indicate discrepancies in the model's calibration and offer valuable insights into the required modifications for enhanced performance.



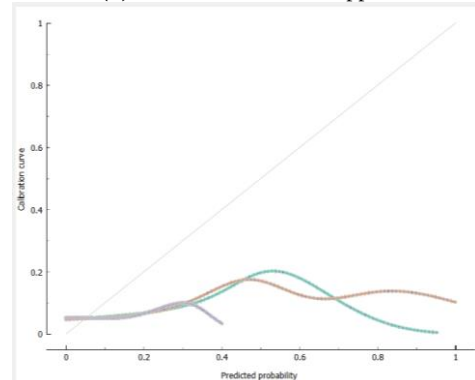
(a) Calibration: Normal apple.



(b) Calibration: Normal apple.

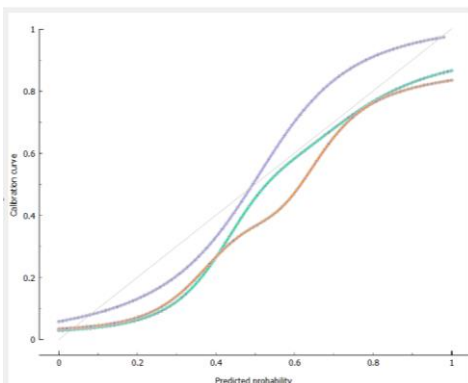


(c) Calibration: Normal apple.

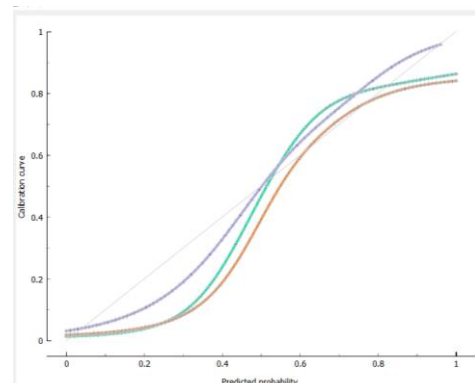


(d) Calibration: Normal apple.

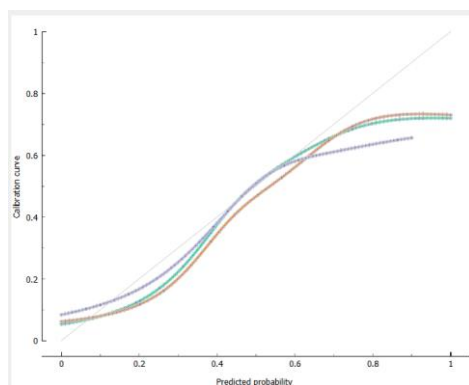
Figure 17. Calibration—Test/Blotch, Rot, Scub, Normal apple classification.



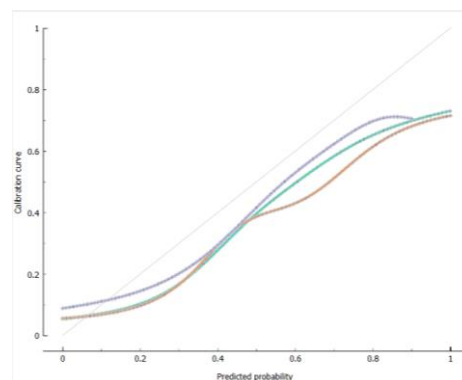
(a) Calibration: Normal apple.



(b) Calibration: Normal apple.



(c) Calibration: Normal apple.



(d) Calibration: Normal apple.

Figure 18. Calibration—Train/Blotch, Rot, Scub, Normal apple classification.

6. Conclusion

This research investigated the categorization of apple diseases by the utilization of Logistic Regression, Neural Networks, and Random Forest models. The Logistic Regression model had superior performance, as evidenced by its high Area Under the Curve (AUC), classification accuracy, and Matthews Correlation Coefficient. Neural networks are closely accompanied by competitive metrics. Although the Random Forest algorithm demonstrated promise, additional optimisation is required. Looking forward, the potential future avenues for this research appear to be promising. The incorporation of these models into real-time orchard monitoring systems will enhance the capacity for proactive disease management. Furthermore, the use of ensemble approaches to exploit the advantages of many algorithms holds the potential to significantly augment the accuracy of classification. Expanding the dataset to incorporate a wider diversity of disease kinds and species will boost model adaptability, supporting the development of comprehensive methods for apple disease management. Furthermore, the convergence of Internet of Things (IoT) and unmanned aerial vehicle (UAV) technologies for the purpose of data acquisition and analysis presents significant potential in enhancing disease detection and control measures inside apple orchards. The exploration of these prospective research directions presents promising opportunities for the progression of disease identification and the establishment of sustainable apple cultivation practices.

Author contributions

Conceptualization, AK and NS; methodology, AK and NS; software, RC; validation, KKG; formal analysis, AA; investigation, AK; resources, MA; data curation, AA and KKG; writing—original draft preparation, AK; writing—review and editing, NS; visualization, AK and RC; supervision, NS and AA; project administration, AK and AA; funding acquisition, AA. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Khalifa SAM, Elshafiey EH, Shetaia AA, et al. Overview of Bee Pollination and Its Economic Value for Crop Production. *Insects*. 2021, 12(8): 688. doi: 10.3390/insects12080688
2. Shuping DSS, Eloff JN. THE USE OF PLANTS TO PROTECT PLANTS AND FOOD AGAINST FUNGAL PATHOGENS: A REVIEW. *African Journal of Traditional, Complementary and Alternative Medicines*. 2017, 14(4): 120-127. doi: 10.21010/ajtcam.v14i4.14
3. Le Cam B, Sargent D, Gouzy J, et al. Population Genome Sequencing of the Scab Fungal Species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata*. *G3 Genes|Genomes|Genetics*. 2019, 9(8): 2405-2414. doi: 10.1534/g3.119.400047

4. Cooke IR, Jones D, Bowen JK, et al. Proteogenomic analysis of the *Venturia pirina* (pear scab fungus) secretome reveals potential effectors. *Journal of Proteome Research*. 2014, 13(8), 3635-3644.
5. Jwo DJ, Chiu SF. Deep Learning Based Automated Detection of Diseases from Apple Leaf Images. *Computers, Materials & Continua*. 2022, 71(1): 1849-1866. doi: 10.32604/cmc.2022.021875
6. Malhotra K, Firdaus M. Application of Artificial Intelligence in IoT Security for Crop Yield Prediction. *ResearchBerg Review of Science and Technology*. 2022, 2(1): 136-157.
7. Lutimath N, Sharma N, Byregowda K. Prediction of heart disease using biomedical data through machine learning techniques. *EAI Endorsed Transactions on Pervasive Health and Technology*. 2021, 7(29).
8. Dubey SR, Jalal AS. Apple disease classification using color, texture and shape features from images. *Signal, Image and Video Processing*. 2015, 10(5): 819-826. doi: 10.1007/s11760-015-0821-1
9. Whitehead SR, Wisniewski ME, Droby S, et al. The Apple Microbiome: Structure, Function, and Manipulation for Improved Plant Health. *The Apple Genome*. Published online 2021: 341-382. doi: 10.1007/978-3-030-74682-7_16
10. Sharma N, Chakraborty C, Kumar R. Optimized multimedia data through computationally intelligent algorithms. *Multimedia Systems*. 2023, 29(5), 2961-2977.
11. Shrivastava VK, Pradhan MK. Rice plant disease classification using color features: A machine learning paradigm. *Journal of Plant Pathology*. 2020, 103(1): 17-26. doi: 10.1007/s42161-020-00683-3
12. Gupta A, Jain V, Singh A. Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications. *New Generation Computing*. 2021, 40(4): 987-1007. doi: 10.1007/s00354-021-00144-0
13. Aggarwal M, Khullar V, Goyal N, et al. Federated Transfer Learning for Rice-Leaf Disease Classification across Multiclient Cross-Silo Datasets. *Agronomy*. 2023, 13(10): 2483. doi: 10.3390/agronomy13102483
14. Aggarwal M, Khullar V, Goyal N, et al. Lightweight Federated Learning for Rice Leaf Disease Classification Using Non Independent and Identically Distributed Images. *Sustainability*. 2023, 15(16): 12149. doi: 10.3390/su151612149
15. Sethy PK, Barpanda NK, Rath AK, Behera SK. Deep feature based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture*. 2020, 175, 105527.
16. Al-Wesabi FN, Albraikan AA, Hilal AM, et al. Artificial Intelligence Enabled Apple Leaf Disease Classification for Precision Agriculture. *Computers, Materials & Continua*. 2022, 70(3).
17. Spolti P, Schneider L, Sanhueza RMV, et al. Improving sooty blotch and flyspeck severity estimation on apple fruit with the aid of standard area diagrams. *European Journal of Plant Pathology*. 2010, 129(1): 21-29. doi: 10.1007/s10658-010-9636-7
18. Raj A, Kumar N, Sheikh TH. Empirical Evaluation of Software Testing Techniques. *International Journal of Recent Technology and Engineering (IJRTE)*. 2019, 7(6s5): 152-153.
19. Singh S, Singh NP. Machine Learning-Based Classification of Good and Rotten Apple. *Recent Trends in Communication, Computing, and Electronics*. 2018, 377-386. doi: 10.1007/978-981-13-2685-1_36
20. Samajpati BJ, Degadwala SD. Hybrid approach for apple fruit diseases detection and classification using random forest classifier. 2016 International Conference on Communication and Signal Processing (ICCSP). Published online April 2016. doi: 10.1109/iccsp.2016.7754302
21. Lutimath NM, Sharma N, Byregowda BK. Prediction of Heart Disease using Random Forest. 2021 Emerging Trends in Industry 40 (ETI 40). Published online May 19, 2021. doi: 10.1109/eti4.051663.2021.9619208