Performance Analysis of Machine and Deep Learning Techniques for Pomegranate Fruit Disease Detection and Classification

Pakruddin B
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
fakrubasha@gmail.com

Mayur R
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
mayur23ramesh@gmail.com

Vikranth Reddy Y
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
vikranthreddy237@gmail.com

Anant chikmurge
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
anantchikmurge2003@gmail.com

Dasari Varshith Narayana
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
varshithdas12@gmail.com

Lokesh T
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
lokesht26022004@gmail.com

Abstract— **Pomegranate (Punica granatum) is a fruit crop with great nutritional and economic values; it has been touted in various parts of the world, especially the Mediterranean, Middle East, and South Asia, for health benefits. However, with more intense effort to increase yield and production, the production of the pomegranate crop is threatened by many diseases caused by pathogens such as fungi, bacteria, and viruses. Traditional diagnosis methodologies are based on the visual inspection of farmers and experts in agriculture, which takes a longer time and also rely on human judgment and may lead to some mistakes. To overcome these challenges this research utilizes the ML and DL methods like SVM, KNN, and CNN for the automatic identification and classification of diseases related to pomegranate. Dataset contains five classes such as Healthy, Anthracnose, Bacterial Blight, Cercospora, and Alternaria. We performed many task in this research like pre-processing, feature extraction to improve the model performance. Each model is trained and tested for detection and classification. KNN got 82% accuracy, which is pretty good but struggles a bit with similar disease traits. SVM performed the best and outwitted KNN with an outstanding accuracy of 93%. The SVM model classified the diseases efficaciously with high sensitivity and specificity in the ROC curve. The CNN model achieved about 98% after it had passed over 10-50 epochs, which had improved incrementally with every epoch. The results confirm that DL is a promising solution for pomegranate disease detection, offering more efficiency and reliability than traditional methods. It supports early disease identification, aiding farmers and experts in better crop management and productivity.**

Keywords—Disease detection, Classification, Pomegranate Fruit, Deep learning, Machine learning.

I. INTRODUCTION

Pomegranates contain nutrients, vitamins C and K, antioxidants, dietary fibers, and anti-inflammatory compounds that improve immunity, health, and prevent chronic diseases. They are used in culinary dishes, traditional medicine, and skincare products because of their health-benefiting properties. Economically, pomegranates form an important crop, especially in countries like India, Iran, Turkey, and Spain, generating considerable agricultural export earnings and income coupled with providing employment. However, successful farming is only possible if the climate ranges from semi-arid to tropical conditions, together with well-drained soil, and proper management of water, pests, and diseases to maintain quality and yield. Pomegranates are among the highly cultural fruits, symbolizing fertility and abundance in harvests; their demand increases over time because of the health benefits bestowed upon consumers as well as versatility in products from fresh to processed produce. Pomegranate is deemed a crop of importance despite climatic, pest, and infrastructure challenges facing the crop [1].

Several major issues plague the production of pomegranate crops, which include effective disease management, uniform quality control, and infrastructure deficiency, which act as liabilities to realize the potential of the crop. Accessibility to the market is also constrained, and climate change only exacerbates the problems for growers. Government initiatives, corresponding dynamic industry responses, and investment strategies in technology and infrastructure shall help in addressing such challenges. These, therefore, need to be addressed for the pomegranate trade to be stronger in terms of production and fruit quality, hence exposing it to a global marketplace [8]. Yielding diseases have especially threatened pomegranate cultivation since they do affect yield and fruit quality. There is a vast scope for better management and detection of diseases with the use of more sophisticated machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN). Fungal Diseases Including Alternaria Fruit Rot, Cercospora Fruit Spot, Aspergillus Fruit Rot, Botrytis Fruit Rot, and Fusarium Wilt can be identified and classified efficiently by employing these methods. For example, CNNs would best be used to classify images, such as scanning images of pomegranate plants for visible signs of symptoms, thus making the detection early and reportable at the soonest possible time [9].

For example, Bacterial Diseases such as Bacterial Blight caused by Xanthomonas axonopodis pv. punicae, can be predicted with great accuracy by SVM and KNN algorithms. These models can analyze the features from images of plants and environmental data on which the risks of infection to various diseases could be determined. The farmer would receive an early warning by training a labeled dataset containing healthy and infected plants through SVM and KNN models in order to take proper measures. Among the factors that can influence the impact of diseases on pomegranate production is effective disease management. Incorporation of these algorithms in existing systems may make it easier to enhance disease detection in the plants, such as KNN, SVM, and CNN algorithms. The algorithms scan images of the pomegranate plants for symptoms of discoloration, wilting, and leaf spots. CNNs are very sensitive to image features and can be useful in the identification of putative fungal, bacterial, or viral infections that otherwise cannot easily be seen. KNN classifies in terms of proximity to what is known about disease patterns, which makes possible fast recognition and action. SVMs may be constructed such that strong classifiers are developed that separate healthy plants from diseased ones by analysis of various characteristics, both plant images and environmental information. Integrating such machine learning techniques with conventional practices to handle diseases, which include better air circulation, farm hygiene, and the cultivation of disease-resistant varieties, farmers will be more proactive about pomegranate disease management. The technique also integrates early detection of disease along with overall health and productivity of pomegranate crops with better fruit quality and yields.

This research paper focuses on developing continuously improved accuracy and speed in pomegranate disease detection systems using advanced algorithms of machine learning. Specifically, the research finds out how KNN, SVM, and CNN help in the identification of fruit diseases based on classification.

II. LITERATURE REVIEW

Mangena et al. [10] suggested framework shows good results in the classification of leaf diseases and was created utilizing MATLAB and a graphical user interface. According to experimental results, the framework has a 98.39% accuracy rate in distinguishing between healthy and diseased leaves. Furthermore, it particularly classifies illnesses on pomegranate leaves with an accuracy of 98.07%. M. D. Nirmal et al. [11] proposed framework, incorporating machine learning technologies, achieves 95.54% accuracy in distinguishing between healthy and unhealthy leaves and 96.43% accuracy in categorizing pomegranate leaf diseases. Using a dataset from Mendeley Data with 559 images (287 healthy and 272 unhealthy), the data was split 80:20 for training and testing. The sorting machine, using proposed algorithms, accurately distinguished sunburned from healthy pomegranates. The brightness intensity distribution algorithm achieved 98% accuracy with a response time of 0.88 seconds [12]. In ref [13] study utilized Suppression Subtractive Hybridization (SSH) to identify potential genes associated with pomegranate resistance to bacterial blight. In ref [1] authors aims to examine agricultural research on topics including pomegranate disease detection and examining algorithms, diseases, datasets, accuracy benefits, and drawbacks of each technique. This review study will be essential for other researchers to comprehend the state-of-the-art in image processing for the identification and categorization of leaf and fruit diseases. P. Wakhare et al.[14] study presents a deep learning method using convolutional neural networks (CNNs) to identify and classify bacterial blight in pomegranates. With a labeled dataset of healthy and diseased fruit, the model achieved 92% accuracy in detection and 89% in classification. M. D. Nirmal et al.[15] proposed machine learning framework achieved 95.54% accuracy in identifying healthy vs. unhealthy pomegranate leaves and 96.43% in disease classification, using a dataset of 559 images (287 healthy, 272 unhealthy) split 80/20 for training and testing. Manisha Bhange et al.[16] study introduces a web-based application that allows farmers to upload fruit images for disease identification. The system, trained on a pomegranate image dataset, processes the uploaded images to assess disease severity.

III. PROPOSED METHODOLOGY

This section outlines the methodology used for detecting pomegranate diseases using machine learning models. The process includes the dataset used, the feature extraction techniques applied, and a detailed explanation of the models: K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). The methodology is designed to accurately classify five pomegranate disease classes: Healthy, Anthracnose, Bacterial Blight, Cercospora, and Alternaria.

A. Dataset used

In this experiment, the dataset is given, with a set of images of pomegranate fruits or leaves from five classes based on the disease type [2]. Healthy: The images of a pomegranate fruit or its leaves which show no signs of disease. Anthracnose: The dark lesions occur on the leaves, stem, and fruits due to fungal invasion. Bacterial Blight: As a result of bacterial origin, the disease appears with water-soaked lesions and black spots on leaves and fruits. Cercospora: A kind of fungal disease that manifests characteristically with small, round spots appearing on the leaves. It provides high-resolution images in RGB, and the dataset was divided using an 80%-20% split to ensure it is of robust evaluation. The image sizes are standardized by resizing so that all images in the dataset appear with the same size, for example, 128 x 128 pixels in size. Normalization applies to scale the pixel values to between 0 and 1 in this approach, which facilitates better convergence during training time.

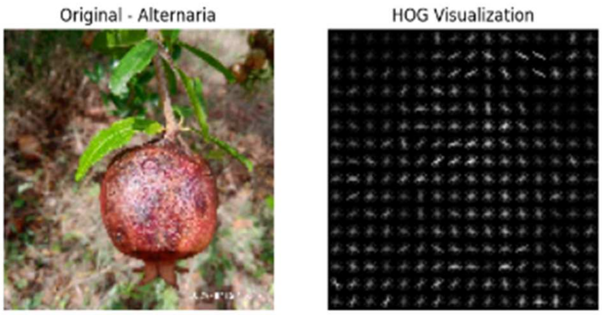
Fig.1. Sample images from dataset

B. *Feature Extraction Techniques*

Feature extraction is crucial for enhancing the performance of machine learning models by converting raw image data into meaningful patterns that can be classified. In this project, various feature extraction techniques were applied:

Histogram of Oriented Gradients (HOG): This technique captures the structure and shape of the object by computing the gradients of image intensity.

(A)



(B)

Fig. 2. HOG feature extraction (a) Healthy (b) Alternaria

Colour Features: The discoloration on the fruit surface due to disease is an important indicator. Colour histograms were used to capture the distribution of colours within each image, allowing the models to differentiate between healthy and infected areas.

Texture Features: The Gray-Level Co-occurrence Matrix (GLCM) was used to extract texture features, which are particularly useful for identifying fungal diseases like Anthracnose and Alternaria that lead to specific rough textures on the fruit or leaves.

Edge Detection: Using techniques like Canny edge detection, we extracted the edges in images to highlight the boundaries of lesions, spots, and other disease manifestations.

Resize: Resizing images ensures that all images in the dataset have a uniform dimension, which is critical for feeding them into machine learning models like CNN, KNN, and SVM. Inconsistent image sizes can lead to errors during training and evaluation.

(A)

(B)

Fig. 3. Sample Pre-processing for (A) Backterial_Blight (B) Cercospora

C. *Types of models used*

The models used for classification were K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). Since models are approached using somewhat different methodologies in classification, this has brought a very interesting comparison which can be drawn through assessing performance on the pomegranate disease dataset.

a) K-Nearest Neighbours (KNN):

KNN is one of the simplest yet powerful algorithms for classification. It classifies the image based on the majority of the vote of its nearest neighbours in the feature space. The working of the KNN algorithm works by computing Euclidean distance between the input image and all images in a training set, and the image is then assigned to the class most common among the K-nearest neighbours. KNN requires manually extracted features-for example, HOG, GLCM, colour features. The extracted features are given to the model to calculate distances and classify images about. KNN is highly easy to implement and works well with small data sets. It does not require a prolonged training session, thereby making it efficient in a few cases. KNN degrades with large size datasets as it is computationally expensive. It is also sensitive to irrelevant features and noise in the dataset [3].

b) Support Vector Machine (SVM)

SVM is a pretty strong learning algorithm in supervised learning and can be used for any kind of classification. The SVM model strives to find that hyperplane which has the maximum margin that separates classes with max-margin samples from other classes. In this case, the model needs to find an optimum hyperplane that separates images of healthy and diseased pomegranates based on the learned features. Very similar to KNN SVM uses pre-extracted features (HOG, GLCM, and colour histograms) for the classification task.

SVM performs satisfactorily in high-dimensional space; especially, for the binary classification tasks, SVM is very robust. Even nonlinear relations of features are also feasible with kernel functions, such as, for example, the RBF kernel. SVM has slower time at training in comparison to other classifiers, especially on large datasets. It does not scale well for multi-class classification. Fine-tuning the hyper parameters of the kernel of choice is necessary, and this proves critical to SVM [4].

c) Convolutional Neural Network (CNN)

CNN is a very strong deep learning model suited only for image classification. Such models automatically learn the relevant features without human intervention; no feature extraction from raw images. A CNN consists of several layers including convolutional layers that extract features from the image and pooling layers that down-sample the image and finally, the fully connected layers to do the final classification. The CNN model applied in this case consists of repeated convolutional layers followed by max pooling layers. Finally, the fully connected layers classify the output into one of the five disease categories [7]. CNNs have proved particularly good for image classification tasks and may automatically learn the most important features from the data. These were generally better than traditional models (KNN, SVM) for the problem of complex image recognitions especially for large sets. CNNs require huge volumes of training data. They are very computationally expensive. They require huge amounts of time and computational resources to tune for hyperparameters [5].

d) Performance Metrics

Accuracy: This is the proportion of correctly predicted instances (both true positives and true negatives) to the total number of instances [17].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: This metric indicates the proportion of true positive predictions to the total predicted positives. High precision means the model is making fewer false positive errors.

$$Precision = \frac{TP}{TP+FP}$$

Recall: (Sensitivity or True Positive Rate): Recall measures the model's ability to correctly identify actual positives. High recall indicates fewer false negatives.

$$Recall = \frac{TP}{TP + FN}$$

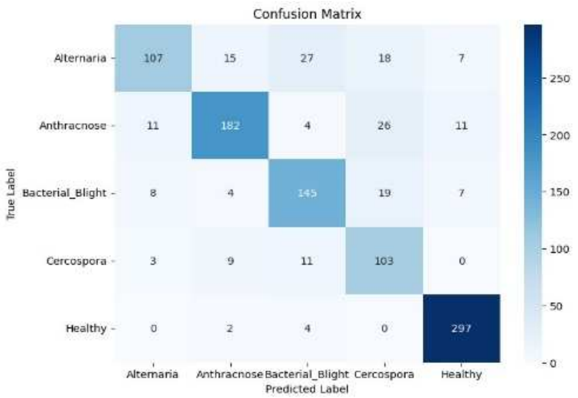
F1-Score: This is the harmonic mean of precision and recall, providing a balanced measure of the model's performance, especially when there is an uneven class distribution [6].

$$F_{Score} = \frac{2 * Recall * Precision}{Recall + Precision}$$

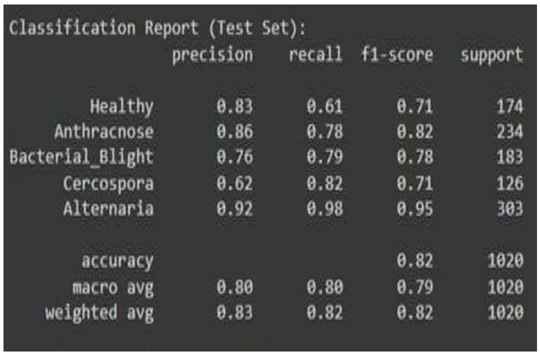
IV. EXPERIMENTAL RESULTS

A. Experimental results for KNN

The K-Nearest Neighbors (KNN) model showed moderate performance on the test set for pomegranate disease detection. It performed well for "Alternaria" (F1-score: 0.95) and "Anthracnose" (F1-score: 0.82) but struggled with "Healthy" (F1-score: 0.71) and "Cercospora" (F1-score: 0.71) due to lower recall and precision. "Bacterial Blight" had balanced metrics with an F1-score of 0.78. The overall accuracy was 0.82, with macro and weighted averages around 0.80, suggesting room for improvement in some classes.



(A)



Classification Report (Test Set):

	precision	recall	f1-score	support
Healthy	0.83	0.61	0.71	174
Anthracnose	0.86	0.78	0.82	234
Bacterial_Blight	0.76	0.79	0.78	183
Cercospora	0.62	0.82	0.71	126
Alternaria	0.92	0.98	0.95	303
accuracy			0.82	1020
macro avg	0.80	0.80	0.79	1020
weighted avg	0.83	0.82	0.82	1020

(B)

Fig.4. Confusion matrix (A) and Classification report (B) for KNN

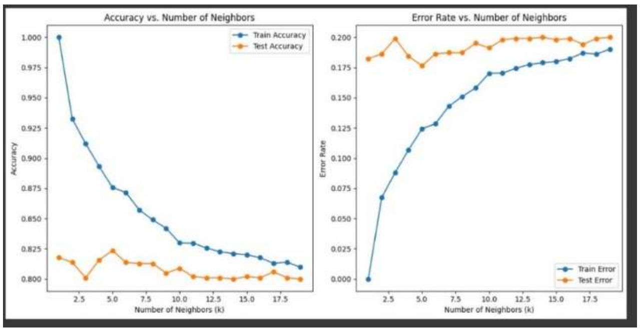
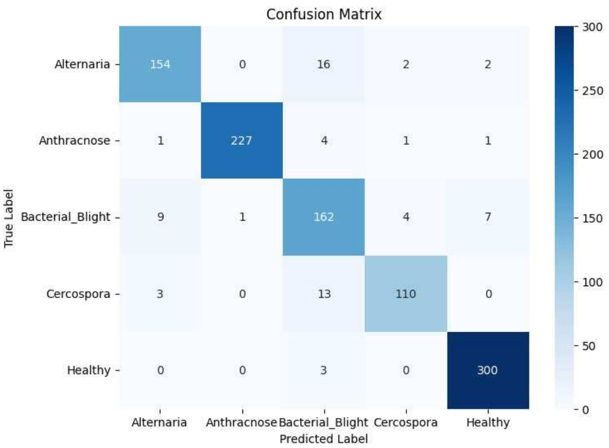


Fig.5. KNN Accuracy and Error rate graph

B. Experimental results for SVM

The test accuracy for the SVM model was 0.9343, indicating good performance. Both "Healthy" and "Anthracnose" exhibited nearly perfect recall (0.99) and high

F1-scores of 0.98. Despite having a somewhat lower recall (0.89), "Alternaria" demonstrated dependable detection (F1-score: 0.90). "Cercospora" achieved 0.91 with a great precision (0.94) but a lower recall (0.87), while "Bacterial Blight" obtained an F1-score of 0.85. With weighted and macro averages of almost 0.93, the model's balanced performance shows that it is reliable for pomegranate disease identification.



(A)

Test Accuracy: 0.9343

	precision	recall	f1-score	support
Alternaria	0.92	0.89	0.90	174
Anthracnose	1.00	0.97	0.98	234
Bacterial_Blight	0.82	0.89	0.85	183
Cercospora	0.94	0.87	0.91	126
Healthy	0.97	0.99	0.98	303
accuracy			0.93	1020
macro avg	0.93	0.92	0.92	1020
weighted avg	0.94	0.93	0.93	1020

(B)

Fig.6. Confusion matrix (A) and classification report (B) for SVM

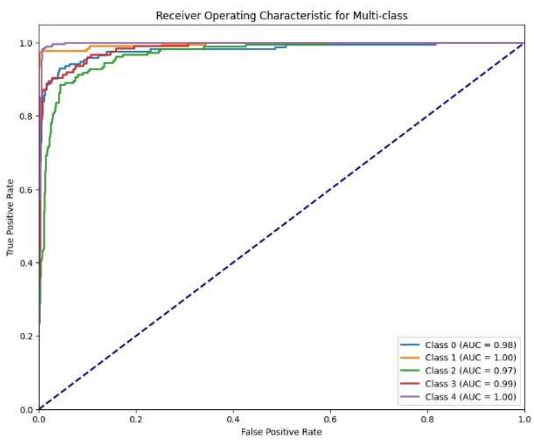
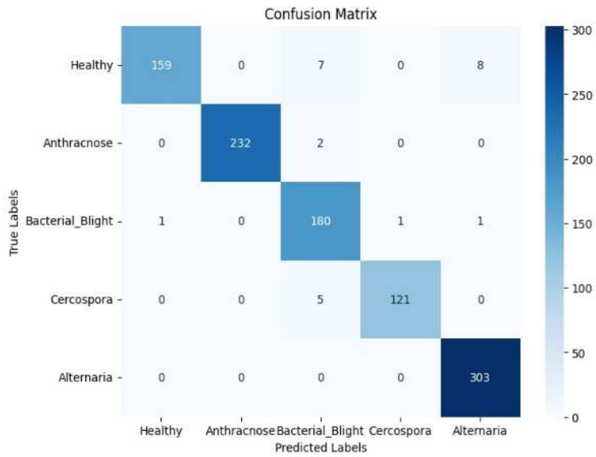


Fig. 7. ROC curve for SVM

C. *Experimental results for CNN*

The CNN model performed exceptionally well, with an accuracy of 0.98. The F1-scores for "Anthracnose" and "Alternaria" were 0.99 and 1.00, respectively, which were almost ideal. With few misclassifications, "Healthy" and "Bacterial Blight" received scores of 0.95 and 0.98, respectively, while "Cercospora" had a score of 0.98. The model's balanced and dependable performance across all illness categories is confirmed by macro and weighted averages of about 0.98.



(A)

Classification Report:

	precision	recall	f1-score	support
Healthy	0.99	0.91	0.95	174
Anthracnose	1.00	0.99	1.00	234
Bacterial_Blight	0.93	0.98	0.95	183
Cercospora	0.99	0.96	0.98	126
Alternaria	0.97	1.00	0.99	303
accuracy			0.98	1020
macro avg	0.98	0.97	0.97	1020
weighted avg	0.98	0.98	0.98	1020

(B)

Fig. 8. Confusion matrix (A) and classification report (B) for CNN

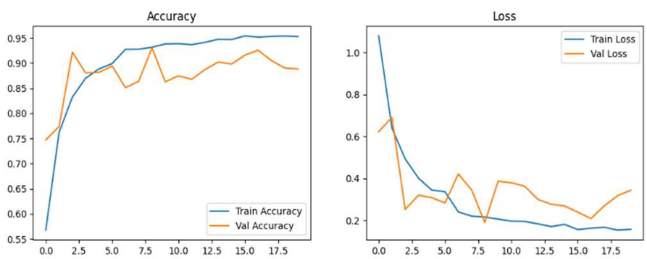


Fig. 9. Accuracy & Loss Graph for CNN

V. CONCLUSION

This research focused on the detection of pomegranate diseases through the development and evaluation of ML and DL models utilizing a dataset comprising five distinct classes: Healthy, Anthracnose, Bacterial Blight, Cercospora, and Alternaria. Through a systematic approach involving data pre-processing, feature extraction, and model training, we successfully implemented three different algorithms KNN, SVM and CNN. Our proposed work demonstrate that the CNN model outperformed both KNN and SVM, achieving an overall accuracy of approximately 98% during testing. The experimental results also showed that KNN and SVM achieved satisfactory accuracies of 82% and 93%, respectively. The successful implementation of these models suggests that similar approaches can be extended to other crops and disease types, thereby enhancing crop yield and minimizing economic losses. Future work should explore the integration of more diverse datasets and additional feature extraction techniques to further improve classification accuracy and model robustness.

REFERENCES

[1] Pakruddin, B., Hemavathy, R. (2024). A Systematic Review of Pomegranate Fruit Disease Detection and Classification Using Machine Learning and Deep Learning Techniques. In: Joshi, A., Mahmud, M., Ragel, R.G., Karthik, S. (eds) ICT: Innovation and Computing. ICTCS 2023. Lecture Notes in Networks and Systems, vol 879. Springer, Singapore. https://doi.org/10.1007/978-981-99-9486-1_13

[2] Pakruddin B., Hemavathy R., A comprehensive standardized dataset of numerous pomegranate fruit diseases for deep learning, Data in Brief Journal Q2 by Elsevier, Volume 54, 2024, 110284, ISSN 2352-3409,https://doi.org/10.1016/j.dib.2024.110284

[3] Uddin, S., Haque, I., Lu, H. et al. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep 12, 6256 (2022). https://doi.org/10.1038/s41598-022-10358-x

[4] Manisha Bhange, H.A. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Pro-cessing", Procedia Computer Science,Volume 58,2015,Pages 280-288.

[5] Vasumathi MT, Kamarasan M. (2021), "An Effective Pomegranate Fruit Classification Based On CNN-LSTM Deep Learning Models", Indian Journal of Science and Technology. 14(16): 1310-1319.

[6] Nirgude, V. ., & Rathi, S. (2021). A Robust Deep Learning Approach To Enhance The Accuracy Of Pomegranate Fruit Disease Detection Under Real Field Conditions. Journal of Experimental Biology and Agricultural Sciences, 9(6), 863–870. https://doi.org/10.18006/2021.9(6).863.870.

[7] Pakruddin, B. and Hemavathy, R. (2024).Development of a Pomegranate Fruit Disease Detection and Classification Model using Deep Learning. Indian Journal of Agricultural Research. doi: 10.18805/IJARe.A-6281.

[8] Pooja Kantale , Shubhada Thakare, 2020, Pomegranate Disease Classification using Ada-Boost Ensemble Algorithm, International Journal Of Engineering Research & Technology (IJERT) Vol-ume 09, Issue 09 (September 2020).

[9] Jadhav, S.B., Udupi, V.R. & Patil, S.B. Identification of plant diseases using convolutional neu-ral networks. Int. j. inf. tecnol. 13, 2461–2470 (2021). https://doi.org/10.1007/s41870-020-00437-5.

[10] Mangena, Venu & Thanh, Dang & Khamparia, Aditya & Pande, Sagar & Malik, Rahul & Gupta, Deepak. (2020). Recognition and Classification of Pomegranate Leaves Diseases by Image Processing and Machine Learning Techniques. Computers, Materials & Continua. 66. 2939–2955. 10.32604/cmc.2021.012466.

[11] M. D. Nirmal, P. Jadhav and S. Pawar, "Pomegranate Leaf Disease Classification using Feature Extraction and Machine Learning," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 619-626, doi: 10.1109/ICOSEC54921.2022.9951907.

[12] Parnian Rezaei, Abbas Hemmat, Nima Shahpari, Seyed Ahmad Mireei, Machine vision-based algorithms to detect sunburn pomegranate for use in a sorting machine, Measurement, Volume 232, 2024, 114682, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2024.114682.

[13] B.Tanuja Priya, B.N.S. Murthy, Anand C. Reddy, Raghavendra Gunnaiah, D.C.Lakshmana Reddy, Identification and expression analysis of candidate genes imparting resistance to pomegranate bacterial blight, South African Journal of Botany, Volume 171, 2024, Pages 485-496, ISSN 0254-6299, https://doi.org/10.1016/j.sajb.2024.06.018.

[14] P. Wakhare, S. Neduncheliyan and P. B. Mane, "Deep Learning-Based Approach for Pomegranate Fruit Bacterial Blight Disease Detection using Convolutional Neural Networks," 2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, India, 2023, pp. 01-06, doi: 10.1109/ICAECC59324.2023.10560218.

[15] M. D. Nirmal, P. Jadhav and S. Pawar, "Pomegranate Leaf Disease Classification using Feature Extraction and Machine Learning," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 619-626, doi: 10.1109/ICOSEC54921.2022.9951907.

[16] Manisha Bhange, H.A. Hingoliwala,Smart Farming: Pomegranate Disease Detection Using Image Processing,Procedia Computer Science,Volume 58,2015,Pages 280-288,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.08.022.

[17] Pakruddin, B. and Hemavathy, R. (2024). Performance Analysis of Various Deep Transfer Learning Models for Bacterial Blight Disease Detection and Classification in Pomegranate Fruits. Indian Journal of Agricultural Research. 1-9.doi: 10.18805/IJARe.A-6245.