



Deep Learning-Driven Orange Fruit Disease Identification using Squeezenet Feature Extraction

Seelam Devi Parvathi², Neethu Krishnan³, M. Nabeela⁴, P. Karputha Pandi⁵

²Assistant professor, Department of Computer Science and Engineering, Marri Laxman Reddy Institute Technology and Management. seelam.devi@gmail.com

³Assistant Professor, Department of Electronics and Communication Engineering, Dhanalakshmi Srinivasan College of Engineering, Coimbatore-641105, India. neethukrishnan@dsce.ac.in

⁴PG Scholar, Department of Electronics and Communication Engineering, Arunachala College of Engineering for Women, Kanyakumari, 629002, E-mail id: nabeeramee27@gmail.com

⁵Assistant professor, Department of Electrical and Electronics Engineering, Erode Senguthar Engineering College, Perundurai. karputhapandi15102@gmail.com

seelam.devi@gmail.com

ABSTRACT: Orange is a citrus fruit which is one of the agricultural crops that have a lot of added value in worldwide. Early disease diagnosis is essential for citrus fruits, as it is for agricultural products, depending on market demands and possible financial losses. In this paper, a XG Boost classification is proposed to quickly recognise orange fruit disease. Firstly, image is smoothening and removes the noise using Gaussian filter to get high quality image using orange disease dataset. Next, the processed image is given to the Fuzzy C-Means clustering for partition of data. After that orange fruit image is featured by Gray-Level Co-occurrence Matrix (GLCM) textured is designed for further analysis. Finally, a XG-Boost framework is processed to enhance the classification of orange fruit disease image. Using python software proposed framework have accuracy of 95% is accomplished when compared to other techniques.

Keywords: XG Boosting, Gray-Level Co-occurrence Matrix (GLCM), Fuzzy C–Means clustering, Gaussian filter, orange fruit image.

1. Introduction: Oranges are a popular and nutrient-dense citrus fruit that are renowned for their juicy, segmented flesh and high vitamin C content. Because they produce a variety of citrus fruits, including tangerines, oranges, lemons, and grapefruits, citrus plants are planted all over the world and have significant economic value [1]. According to the Food and Agriculture Organization (FAO) of United Nations -2022 projection, almost 50% of the 197,198 million tonnes of citrus fruit will be produced in the worldwide as orange [2]. Whereas, countries like Spain, Brazil, and US produce about 34 million tons of citrus fruits a year [3].

However, in the food sector, citrus fruits are a flexible ingredient that is utilized to make a wide range of products, including baked goods, jellies, and juices. [4-5]. Moreover, Citrus production is difficult for a number of reasons, including their extreme sensitivity to cold weather, the need for acidic soil, and their vulnerability to a number of crop diseases. Citrus Greening, Anthracnose, Citrus Black Spot (CBS), Citrus Canker (CC), and Sweet Orange Scab are five prevalent diseases that threaten citrus output [6-7].

To increase the size and diversity of dataset for orange disease data augmentation is used. Benefits like better model accuracy, less over

fitting, and more generalization. Drawbacks of data augmentation are noise being introduced, and the necessity of technique selection [8]. Conversely, the decayed areas in orange fruit are segmented using improved watershed segmentation and global thresholding. It segments the decayed areas in orange fruit absolutely with good performance. Images with complicated intensity distributions or different lighting are difficult to distribute [9]. Although, an early citrus fruit is detected using an improved watershed algorithm. Over-segmentation and noise sensitivity be reduced, ultimately producing more reliable and accurate segmentation [10].

Whereas, to improve the classification performance an orange fruit Multi-Class Support Vector Machine (MC-SVM) is used. Low performance in high-dimensional area and high computational cost [11]. For orange fruit disease classification a Hopfield Recurrent Artificial Neural Network is used. A collection of patterns is stored using Hopfield networks, which are then extracting a complete pattern from a noisy image. Computation and training be difficult, particularly for large networks [12].

However to estimate and detect the orange fruit a Multilayer Perceptron Neural Networks is used. Provide quick detection after training and well in large input data. Proper functioning of the model depends on the quality of training [13]. Moreover, using Convolutional Neural Network (CNN) the citrus or orange fruit disease is detected. By combining several layers, the CNN model extracts complimentary discriminative characteristics. CNN have high processing difficulties and problems with interpretability [14-15]. To overcome this limitation of orange fruit disease the proposed XG boost classifier is used.

2. Related work

Syed-Ab-Rahman *et al* [16] (2022) have proposed a Two stage Deep Convolutional Neural Network (DCNN) for the detection and classification of citrus disease. It maximizes

disease detection plan, accurate and effective for screening of orange diseases. The drawback of this method is it needs large amount of labelled data to calculate and it takes long time to train the input data.

Momeny *et al* [17] (2022) have proposed a Deep Convolutional Neural Network (CNN) model with ResNet-50 model for detecting black spot disease and ripeness levels in orange fruit. A Deep Convolutional Neural Network (DCNN) with ResNET-50 to classify healthy and diseased orange and precise detection of diseases in orange. DCNN with ResNET-50 have high processing difficulties, lacking a portion of labelled data.

Uğuz *et al* [18] (2023) have proposed a new Convolutional Neural Network (CNN)-based model with Citrus Net, for the classification of defective and deformed citrus fruits also a Mask R-CNN models adequately detected the citrus diseases. Using Mask R-C NN model orange image improve the diagnostic accuracy and lower the overall misdiagnosis rate. However, it is not suitable for the real time application.

Dhiman *et al* [19] (2022) have proposed a Deep Neural Network (DNN) with VGGNet to classify the citrus fruit. DNN model is used to detect specific parts of the disease with its severity degree. However, training period is computationally costly and demands a large amount of processing power.

da Silva *et al* [20] (2023) have proposed a You Only Live Once (YOLO) and Faster Region-Convolutional Neural Network (Faster R-CNN) detects the fruit on trees. Faster R-CNN reduces computational requirements, making it possible to employ low-footprint models involving both detection and classification. It is more difficult for YOLO to precisely locate things in complex surroundings or situations with overlapping objects.

The work's contributions include the following:

- Gaussian filter is utilized to smoothen the image and remove noise from the input orange images, enhancing image quality for analysis.
- Fuzzy C-means clustering is employed to segment orange fruit images, fuzzy partition for better feature extraction and analysis.
- GLCM is employed to structure and texture the features from the segmented orange fruit images for its properties.
- A XG Boost based framework is proposed, to classify the featured orange fruit image for identification.

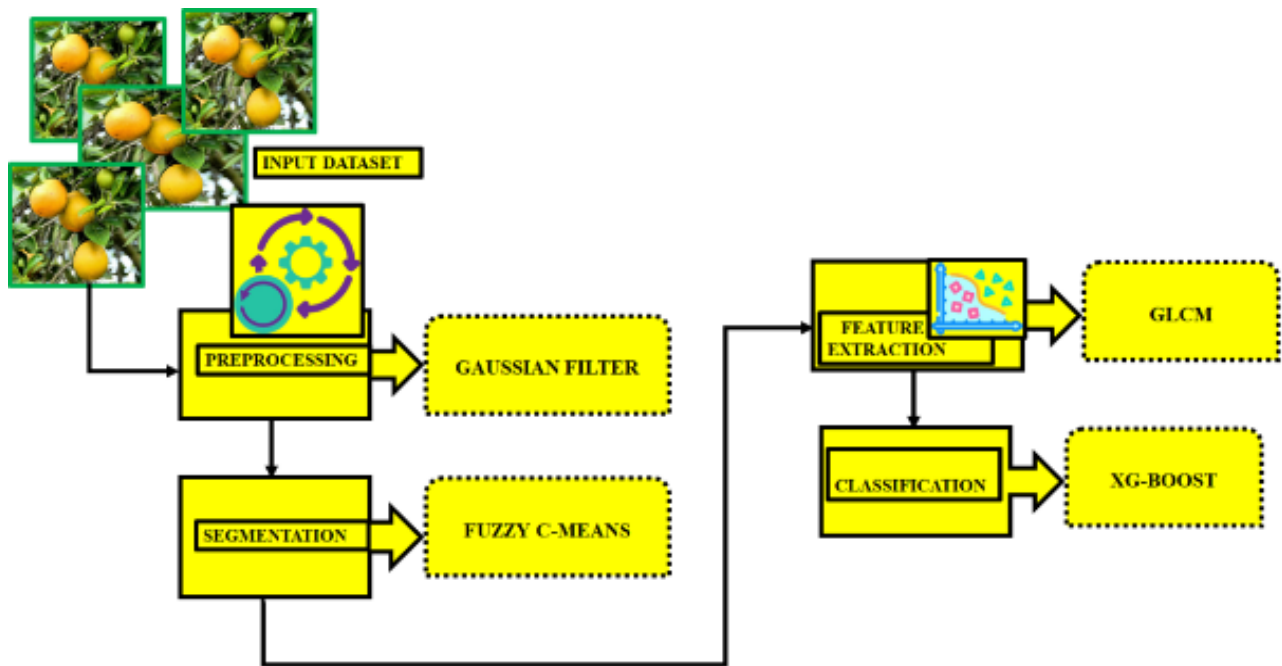


Figure 2: Block diagram of proposed work

Finally, featured image is given to the classification process where a proposed XG Boost is used for classifying the orange image. The XG Boost has boost trapping the trees which is used to classify the orange image for getting better accuracy when compared to the other techniques.

3.1 Pre-processing using Gaussian filter

The Gaussian filter is used to denoise the orange image since the noise affects it. According to Gaussian theory, the noise image is smoothed out and the filter functions as a kind of local filter with linear. A frequent procedure in computer vision and image processing is the Gaussian

3. Proposed work

The proposed block diagram in figure 2 shows the orange fruit disease classification. Firstly, the input image is given to the preprocessing step where a Gaussian filter smoothen image and reduces noise for good quality image. Secondly, the segmentation process uses the Fuzzy C-means clustering for segmenting orange fruit image. Thirdly, by using GLCM feature extraction technique the segmented image are featured by its properties.

principle filter. It pays less attention to the image's edges and features as compared to other filters. The Gaussian filters have a respectable effect on denoising that conforms to the normal distribution. It accurately forecasts the image fixation.

$$\frac{1}{2 \times 3.14 \times \sigma^2} e^{\frac{-(u^2+v^2)}{\sigma^2}} \quad (1)$$

Where, u indicates u coordinate value, v indicates v coordinate value and σ stands for standard deviation, 3.13 indicates mathematical constant PI. Following, to filtering technique segmentation is carried out using Fuzzy C-Means clustering.

3.2 Segmentation by Fuzzy C-Means clustering

Fuzzy C-means (FCM) is a clustering technique which is separated from hard k-means that employs hard partitioning. The FCM employs fuzzy partitioning such that a data point belong to all groups with different membership grades between 0 and 1. FCM is an iterative algorithm. The aim of FCM is to find cluster centers (centroids) that minimize a dissimilarity function. This algorithm is employed for distance-based analysis of different input data points. The distances between the data points are used to create the clusters and each cluster have its own cluster center. Figure 3 and 4 shows the hard and soft clustering.

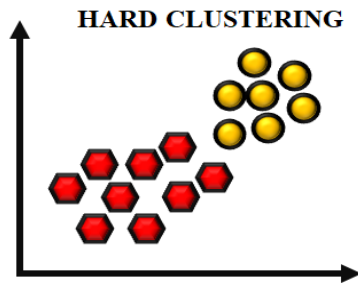


Figure 3: Hard clustering

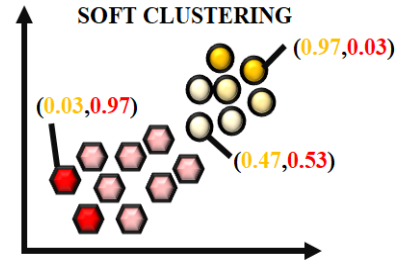


Figure 4: Soft clustering

3.3 Steps in the algorithm for fuzzy C-Means clustering

After choosing a value for parameter "m," we must initialize the partition matrix $U(0)$ and fix c , where c equals $(2 \leq c < n)$. This algorithm will designate each step as "r," where $r = 0, 1, 2, \dots$

- 1) Determine the c center vector $\{V_{ij}\}$ for every step.

$$V_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^m x_{kj}}{\sum_{k=1}^n (\mu_{ij})^m} \quad (2)$$

- 2) Evaluate the distance matrix $D_{[c,n]}$

$$D_{ij} = (\sum_{j=1}^m (x_{ij} - v_{ij})^2)^{1/2} \quad (3)$$

- 3) Modernize the partition matrix for the r^{th} step, $U^{(r)}$ as

$$\mu_{ij}^{r-1} = \left(\frac{1}{\sum_{j=1}^c (d_{ik}^r / d_{jk}^r)^{\frac{2}{m-1}}} \right) \quad (4)$$

If $\|U^{(k+1)} - U^{(k)}\| < \delta$ end; if not, must go back to step 2 and update the cluster centers and membership grades for the data points iteratively.

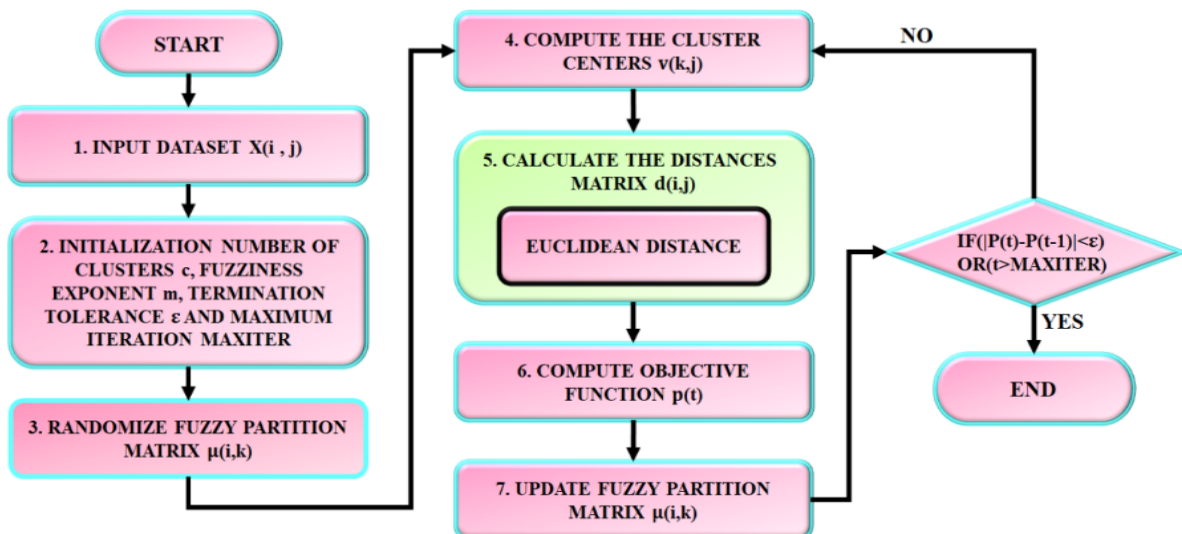


Figure 5: Flow chart of Fuzzy C-Means clustering

The cluster centers are iteratively moved to the proper location within a dataset using FCM. Following, to segmentation technique feature extraction is carried out using GLCM.

3.4 Feature extraction by GLCM

The Grey-Level Co-occurrence Matrix (GLCM) feature, which uses the distance among two pixels in an image to define its value in the orange fruit images. Whereas a coarse texture region changes more slowly, a fine texture region fluctuations more quickly. In order to extract the texture features, first choose the grayscale images that have additional image information, following

their extraction from grayscale images, the texture indices were chosen as texture characteristics based on their strong correlations with image. The GLCM intentions in dissimilar window sizes (division: pixel $\times$ pixel) were used to quantitatively describe the texture statistical attributes, such as the dissimilarity, entropy, second moment, mean, variance, homogeneity, contrast, and correlation, as well as to quantify the texture thickness, similarity, contrast, uniformity, or regularity of GLCMs in rows or columns. Figure 6 shows the basic GLCM feature.

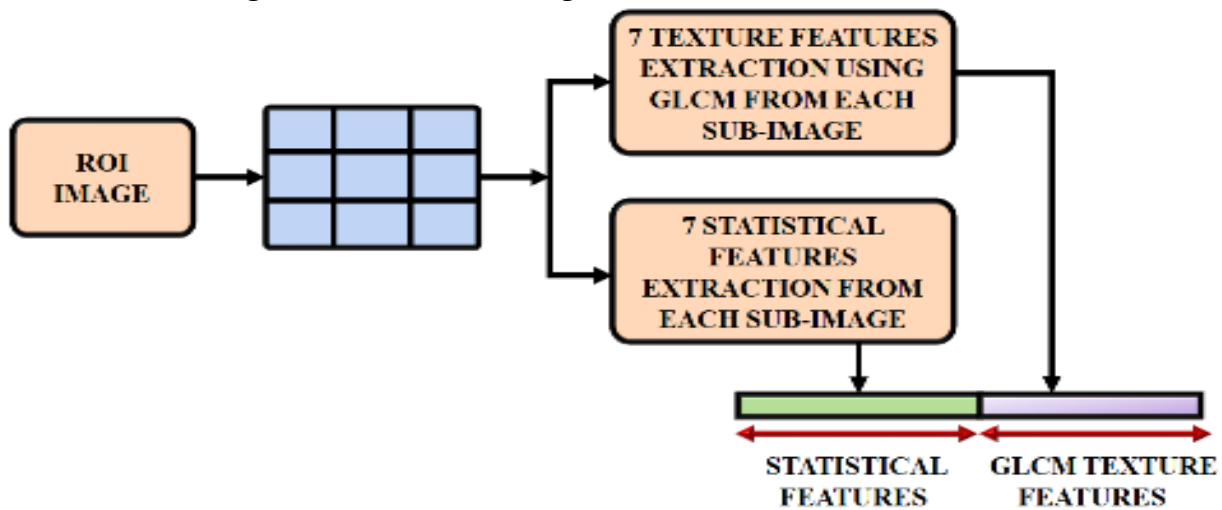


Figure 6: Basic GLCM features

3.4.1 Algorithm of GLCM feature for orange fruit images

- i. The first step is to change the color image to grayscale.
- ii. A 5x5 filter is applied to the input orange image.
- iii. The filter image is divided into 4x4 pieces.
- iv. GLCM Characteristics Energy, standard deviation, mean value, homogeneity, and contrast are defined for each block. The computation of these properties is done in four directions: horizontal (90°), vertical (0°), and diagonal (45° and 135°).
- v. Features Extracted: Followed by this a featured image is given to the classification step to classify the orange fruit disease images.

3.5 Classification by XG Boost

The proposed XGBoost (Extreme Gradient Boosting) model solve both regression and classification issues. Gradient Boosting, an ensemble learning method, is the foundation of XGBoost. Ensemble learning is the process of combining several weak predictors, like decision trees, to create a stronger model. The way gradient boosting operates is by training weak predictors one after the other, each of which attempts to fix the mistakes of the one before it. The XGBoost algorithm's default parameter settings were used in this investigation. To increase predicting accuracy of orange fruit disease, XGBoost combines several weak classifiers (Tree 1-t) into a single strong classifier. For instance, in regression, every

XGBoost tree (apart from tree 1) learns the loss (negative gradient) of the sum of the outcomes of all earlier trees. The true value can be

continuously approached by adding the loss value to the previously predicted value. Figure 7 shows the proposed XG boost classifier.

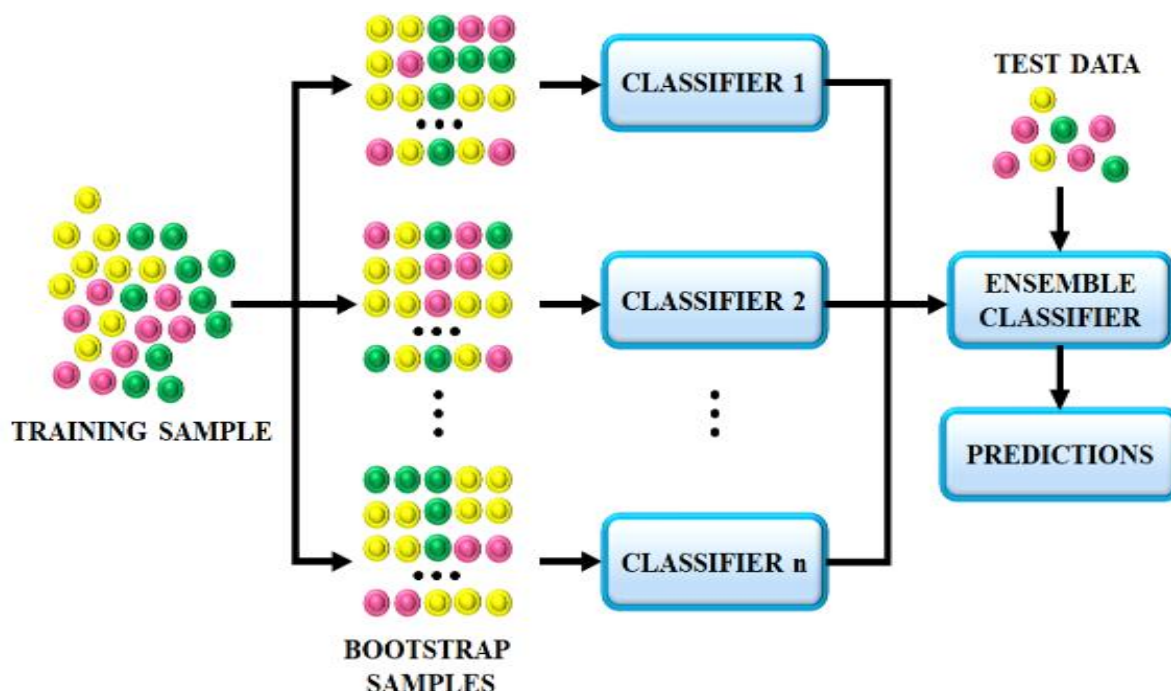


Figure 7: Proposed XG Boost

The algorithm sorts the greatest use of memory assets and is very effective in reducing computation time. It takes the distinct ability to boost on extra data that has already been contributed to the trained model, allows parallel structure in tree construction, and capable of handling missing information. Because it addressed the mainstream of the issues that arose from the previous models, XGBoost is selected. XGBoost is well-suited to identify and handle missing values. At this time, resultant to the condition at the layer, the data splits up into aggregating. To refine the parameters of XG Boost Classifier the proposed work is classified.

3.6 Result and discussion

The aim of this work is to classify the four classes of image. Dataset is taken from kaggle.com for fruit disease analysis. Deep learning models are then applied to train orange disease dataset.

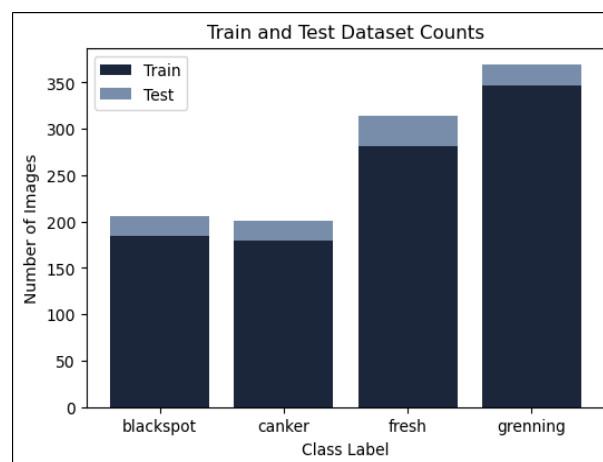


Figure 8: Train and Test dataset counts

Figure 8 shows the train and test dataset counts for orange fruit disease. Here total number of images used is 380 and four classes of labels are black spot canker, fresh, greening. The four classes' labels are trained using the orange disease dataset. The train and test image for black spot is 180 for train and 200 for test, for canker the train images is 170 and 190 images for test, for fresh 280 images for train and 310 for test

images at last for greening 340 for train and 360 for test.

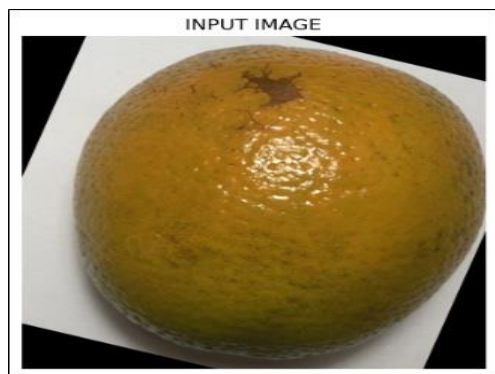


Figure 9: *Input image*

Figure 9 shows the input orange fruit image. Here the input image is taken from the orange disease dataset. Using the image processing techniques the input image is processed.

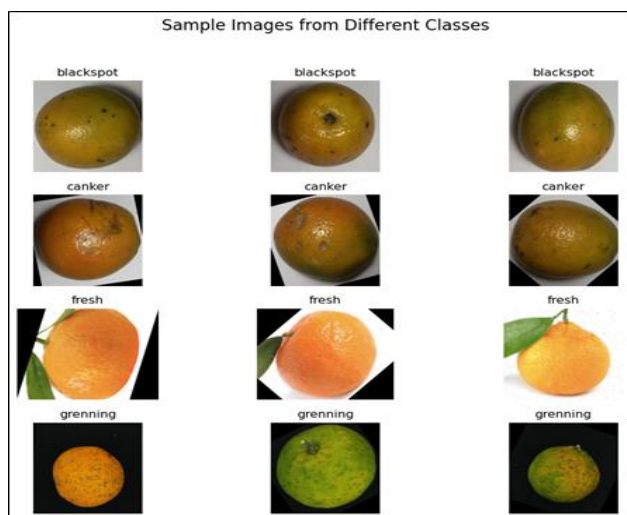


Figure 10: *Sample images from different classes*

Figure 10 shows the sample images from different classes. Here four classes of orange fruit image are blackspot, canker, greening and fresh. The samples images are taken from the orange disease dataset.



Figure 11: *Resized image*

Figure 11 shows the original image to be resized. Here the input original image is taken from the orange disease dataset which is resized using image resizing technique.

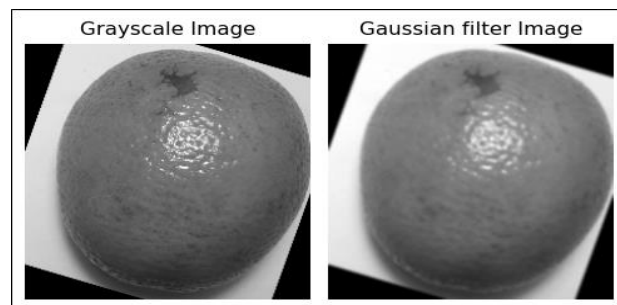


Figure 12: *Gaussian filter image*

Figure 12 shows the Grayscale image to Gaussian filter image. Here the noises are removed and smooth's the gray image using Gaussian filtering.

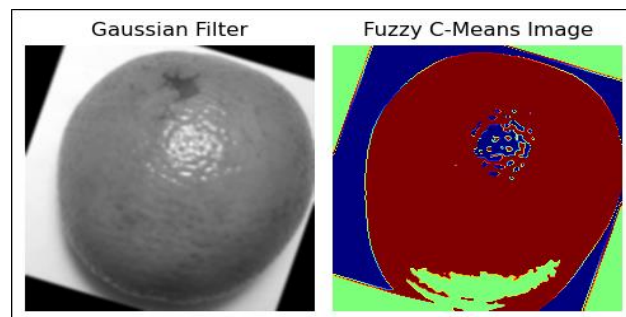


Figure 13: *Fuzzy C-Means clustering*

Figure 13 shows the Fuzzy-C means clustering applied to orange image. Here the Gaussian filter remove noise and given as an input to the fuzzy-C means clustering to cluster the orange image.

Table 1: *GLCM properties*

GLCM properties	Values
Contrast	0.0698
Dissimilarity	0.0455
Homogeneity	0.9796
Energy	0.6760
Correlation	0.9375

Table 1 shows the Grey Level Co-occurrence matrix for the orange fruit image. The GLCM features for the orange fruit are calculated with their properties that are contrast, homogeneity, energy and correlation is in table 1.

3.7 Performance metrics

By examining accuracy, precision, sensitivity, specificity, Area Under the Curve (AUC), classification with "positive" and "negative" labels, the performance of each classification model in this study explained in table 2.

Table 2: Performance metrics

Performance metrics	Formula
Accuracy	$\frac{TN + TP}{(TN + TP)}$
Precision	$\frac{TP}{(TP)}$
Recall	$\frac{TP + FP}{(TP)}$
F1 – score	$\frac{(2 * Precision * Recall)}{(Precision + Recall)}$

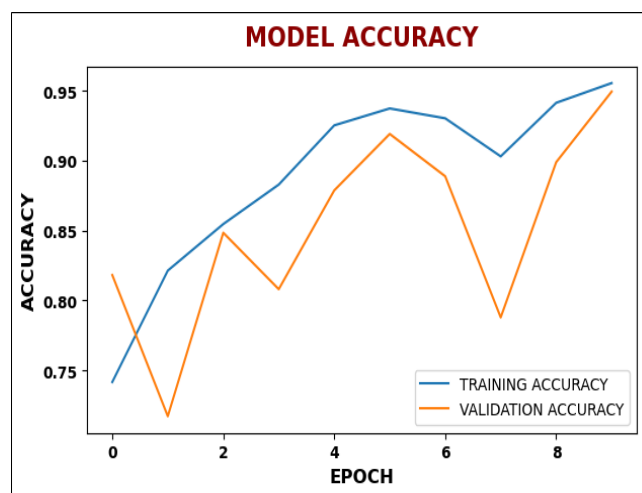


Figure 14: Model accuracy

The model Accuracy graph in figure 14 illustrates increasing accuracy of 95%, with the model achieving high performance and minimal over fitting.

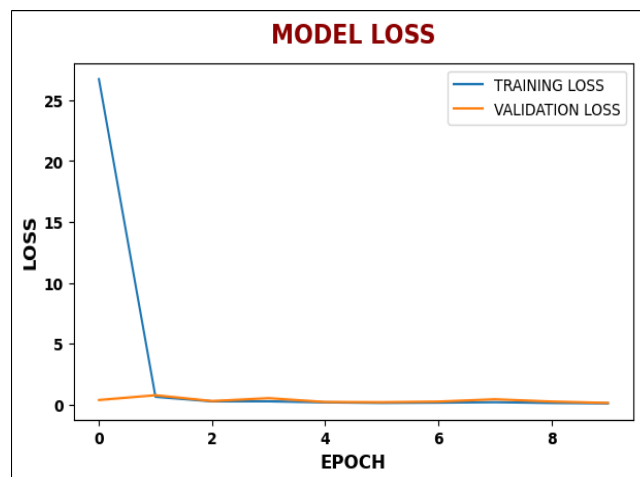


Figure 15: Model loss

The model Loss graph figure 15 shows a reliable decline in training and validation loss, representing improved learning and reduced error. The X-axis denotes the epochs and y-axis is loss. The validation loss is constant still 8 epochs, the training loss is going on decreasing and then constant.

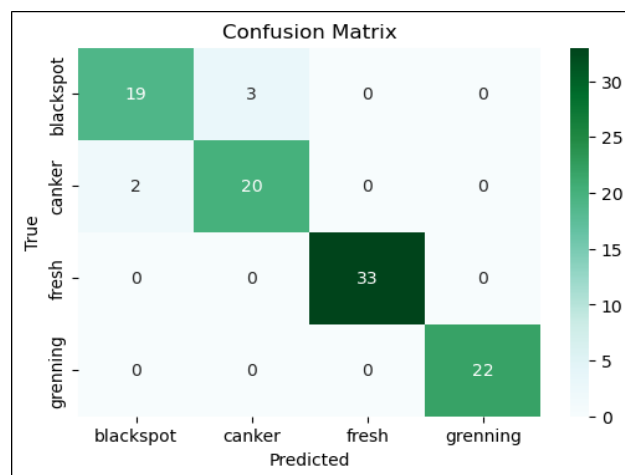


Figure 16: Confusion matrix for proposed XG Boost

Figure 16 illustrates the confusion matrix for proposed XG Boost. This matrix shows the values for true and predicted labels. Confusion matrices are commonly used to evaluate the performance of classification models, particularly those tasked with predicting categorical labels for input instances. Here the four classes of orange disease image for blackspot is 19, canker is 20, fresh is 33 and greening is 22 are matrixes.

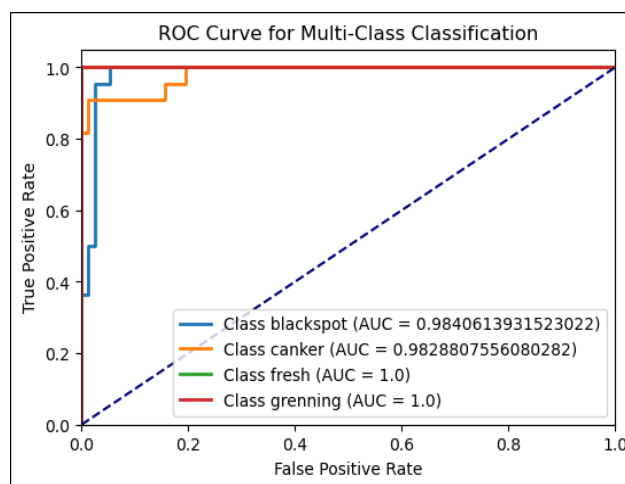


Figure 17: ROC curve for multi-class classification

According to the AUC results shown in Figure 17, the orange disease dataset significantly lowers the AUC value of XG Boost, the sub-optimal model on datasets, to 0.98 for black spot, 0.98 for canker, 1 for fresh and 1 for greening. In comparison, the suggested XG Boost continues to have the highest AUC value of 1, demonstrating the exceptional overview and flexibility of this approach.

3.8 Comparison for the proposed method

The comparison of the proposed method with existing method is presented in table 3. The MC-SVM have the classification accuracy of 93.74% [11], CNN have the classification accuracy of 93.8% [15] and CNN have the accuracy of 94.5% [14]. The proposed XG Boost have the higher accuracy of 95% compared to the existing method in table 3. The proposed XG Boost have better performance.

Table 3: Comparison for the proposed method

Study	Method	Accuracy
Khan <i>et al</i> [11]	MC-SVM	93.75%
Zeeshan <i>et al.</i> [15]	CNN	93.8%
Khattak <i>et al.</i> [14]	CNN	94.5%
Proposed approach	XG Boost	95%

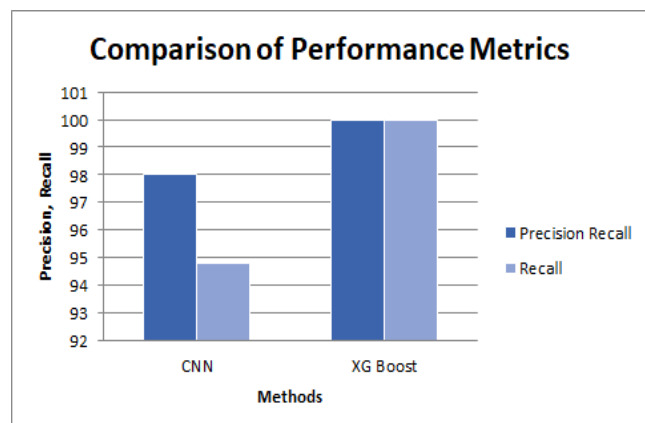


Figure 18: Comparison of performance metrics

The Precision and recall value for classifiers CNN and the proposed is presented in Figure 18. The CNN have the value of 98% and 94.8% [15] and proposed XG Boost methods have the precision value of 100% and recall of 100% respectively.

4. Conclusion

In this study, XG Boost is proposed for the classification of orange fruit disease. The integration of advanced preprocessing techniques, such as resizing, grayscale conversion, Gaussian filtering, ensures that the input images are optimized for segmentation. The use of approximate Fuzzy C-means clustering for segmentation further enhances the precision of identifying partition data in the images. Additionally, the GLCM feature analyzed the properties. The performance evaluation, conducted on the orange disease dataset, demonstrates the superior abilities of the XG Boost related to existing methods. The improved accuracy of 95% is achieved as a higher accuracy when compared to the existing techniques.

References

1. Sathian Dananjayan; Yu Tang; Jiajun Zhuang; Chaojun Hou; Shaoming Luo, Year: 2022, "Assessment of state-of-the-art deep learning based citrus disease detection techniques using annotated optical leaf images", computers and Electronics in Agriculture, Vol: 193, pp. 106658.
2. Poonam Dhiman; Amandeep Kaur; Yasir Hamid; Eatedal Alabdulkreem; Hela Elmannai; Nedat Ababneh, Year: 2023, "Smart disease detection system for citrus fruits using deep learning with edge computing", Sustainability, Vol:15, No: 5, pp. 4576.
3. Guangyan, Zhong; Elisabetta Nicolosi, year: 2020, "Citrus origin, diffusion, and economic importance", The citrus genome, pp. 5– 21.
4. Ramesh Kumar Saini; Arina Ranjit; Kavita Sharma; Parchuri Prasad; Xiaomin Shang; Karekal Girinur Mallikarjuna Gowda; Young-Soo Keum, Year: 2022, "Bioactive compounds of citrus fruits: A review of composition and health benefits of carotenoids, flavonoids, limonoids,

and terpenes”, *Antioxidants*, Vol: 11, No: 2, pp. 239.

5. Franklin Jackson Machado; Fabrício Eustáquio Lanza; Marcela Olivetti Ferretti; Régis Oliveira Fialho; Franklin Behlau; Geraldo José Silva-Junior, Year: 2022, “Citrus black spot severity related to premature fruit drop in sweet orange orchards”, *Plant Pathology*, Vol: 71, No: 2, pp. 400– 410.

6. Leng Han; Guijie Li; Xuting Wang; Bo Yu; Tenghui Zhang; Yujiao Cheng, Year: 2024, “Characterization of volatile compounds from healthy and citrus black spot-infected Valencia orange juice and essential oil by using gas chromatography–mass spectrometry”, *Food Chemistry: X*, Vol: 22, pp. 101374.

7. Diego Henrique Ferreira; Rafaele Regina Moreira; Geraldo José Silva Junior; Franklin Behlau, Year: 2022, “Copper rate and spray interval for joint management of citrus canker and citrus black spot in orange orchards”, *European Journal of Plant Pathology*, Vol: 163, No: 4, pp. 891– 906.

8. Poonam Dhiman; Poongodi Manoharan; Umesh Kumar Lilhore; Roobaea Alroobaea; Amandeep Kaur; Celestine Iwendu; Majed Alsafyani; Abdullah M. Baqasah; Kaamran Raahemifar, Year:2023, “PFDI: a precise fruit disease identification model based on context data fusion with faster-CNN in edge computing environment”, *EURASIP Journal on Advances in Signal Processing* 2023, No: 1, pp. 72.

9. Jiangbo Li; Yuzhen Lu; Renfu Lu, Year: 2024, “Identification of early decayed oranges using structured-illumination reflectance imaging

coupled with fast demodulation and improved image processing algorithms”, *Postharvest Biology and Technology*, Vol: 207, pp. 112627.

10. Xi Tian; Chi Zhang; Jiangbo Li; Shuxiang Fan; Yi Yang; Wenqian Huang, Year: 2021, “Detection of early decay on citrus using LW-NIR hyperspectral reflectance imaging coupled with two-band ratio and improved watershed segmentation algorithm”, *Food Chemistry*, Vol: 360, pp. 130077.

11. Muhammad Attique Khan; Tallha Akram; Muhammad Sharif; Majed Alhaisoni; Tanzila Saba; Nadia Nawaz, Year: 2021, “A probabilistic segmentation and entropy-rank correlation-based feature selection approach for the recognition of fruit diseases”, *EURASIP Journal on Image and Video Processing*, Vol: 2021, No: 1, pp. 14.

12. Igor R. Fermo; Thiago S. Cavali; Lucas Bonfim-Rocha; Caio L. Srutkoske; Franklin C. Flores; Cid MG Andrade, Year: 2021, “Development of a low-cost digital image processing system for oranges selection using hopfield networks”, *Food and bioproducts processing*, Vol: 125, pp.181– 192.

13. Sajad Sabzi; Hossein Javadikia; Juan Ignacio Arribas, Year: 2020, “A three - variety automatic and non-intrusive computer vision system for the estimation of orange fruit pH value”, *Measurement*, Vol: 152, pp.107298.

14. Asad Khattak; Muhammad Usama Asghar; Ulfat Batool; Muhammad Zubair Asghar; Hayat Ullah; Mabrook Al-Rakhami; Abdu Gumaei, Year: 2021, “Automatic detection of citrus fruit and leaves diseases using deep neural network

- model”, *IEEE access*, Vol: 9, pp.112942–112954.
15. Sadaf Zeeshan; Tauseef Aized; Fahid Riaz, Year: 2023, “The design and evaluation of an orange-fruit detection model in a dynamic environment using a convolutional neural network”, *Sustainability*, Vol: 15, No: 5, pp. 4329.
16. Ab-Rahman-Syed; Sharifah Farhana; Mohammad Hesam Hesamian; Mukesh Prasad, Year: 2022, “Citrus disease detection and classification using end-to-end anchor-based deep learning model”, *Applied Intelligence* Vol: 52, No: 1, pp. 927– 938.
17. Mohammad Momeny; Ahmad Jahanbakhshi; Ali Asghar Neshat; Ramazan Hadipour-Rokni; Yu-Dong Zhang; Yiannis Ampatzidis, Year:2022, “Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks”, *Ecological Informatics*, Vol: 71, pp. 101829.
18. Sinan Uğuz; Gulhan Şikaroğlu; Abdullah Yağız, Year: 2023, “Disease detection and physical disorders classification for citrus fruit images using convolutional neural network”, *Journal of Food Measurement and Characterization*, Vol: 17, No: 3 pp. 2353– 2362.
19. Poonam Dhiman, Vinay Kukreja; Poongodi Manoharan; Amandeep Kaur; M. M. Kamruzzaman; Imed Ben Dhaou; Celestine Iwendi, Year: 2022, “A novel deep learning model for detection of severity level of the disease in citrus fruits”, *Electronics*, Vol: 11, No: 3, pp. 495.
20. Jonathan CF da Silva; Mateus Coelho Silva; Eduardo JS Luz; Saul Delabrida; Ricardo AR Oliveira, Year: 2023, “Using mobile edge AI to detect and map diseases in citrus orchards”, *Sensors*, Vol: 23, No: 4, pp. 2165.