International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Detection and Classification of Disease from Mango fruit using Convolutional Recurrent Neural Network with Metaheruistic Optimizer

[1]Vigneswara Reddy K., [2]Prof. Dr. A. Suhasini, [3]Dr. V. V. S. S. S. Balaram

Submitted: 19/10/2023 **Revised**: 11/12/2023 **Accepted**: 16/12/2023

Abstract: Fruits play a vital role in providing essential nutrients to sustain goodhuman health. The significant reduction in crop output is mostly attributed to the substantial impact of fruit diseases, which arise as a consequence of inadequate maintenance practices and the proliferation of fungal pathogens. The mango fruit has significant global consumption and is vulnerable to illnesses that may impact both its quality and quantity. The process of manual inspection is characterized by its arduous nature, extensive time requirements, heavy reliance on human labor, and lack of efficiency. This study employs an image classification methodology to discern several illnesses present in mangoes and distinguish them from the unaffected specimens. The preprocessing phase consists of two primary stages, namely background removal and contrast enhancement. The method of histogram equalization involves enhancing the contrast of animage. Following the preprocessing stage, the subsequent step involves the use of instance segmentation, which serves as a significant operation. The radiomic characteristics that have been retrieved are inputted into the CNN_FOA, a Convolutional Recurrent Neural Network with a classifier based on the FireFly Optimizer. This CNN_FOA is used to classify the mango images into different categories. The suggested model has undergone experimental verification and validation, yielding optimal outcomes with a precision rate of 97%.

***Keywords*:** *Mango fruit, optimization, background removal, instance segmentation, CNN, disease classification*

1. Introduction

Mangoes are a very profitable fruit that is extensively cultivated in tropical and sub-tropical locations around the globe. The alluring fragrance, delectable flesh, and substantial nutritious content of mangoes have garnered a significant following among enthusiasts around the globe, resulting in substantial economic advantages for both mango cultivators and exporting nations. It is important to highlight that the economic valuation of a mango fruit is significantly influenced by its visual aesthetics. The mangoes with the most visually appealing characteristics are often designated for exportation, while those with inferior visual appeal are allocated for local consumption. The mangoes with the least desirable appearance are typically used for further processing, such as the production of canned fruit or jam. Nevertheless, the assessment of mango quality is a time-consuming procedure that has mostly depended on manual examination so far. The process of fruit preservation, which requires a significant amount of time, not only reduces the duration during which fresh fruits may be sold at a profit, but also carries the risk of human mistakes that may result in financial losses. In order to assist mango growers in efficiently and precisely carrying out the grading process, the current investigation, in partnership with the Taiwan AI CUP 2020 rivals, intends to use widely recognised deep learning techniques in the context of computer vision, specifically CNN [1]–[4]. The application of machine learning to various domains presents challenges that primarily revolve around two key aspects: the assurance of data quality and the selection of appropriate existing tools with task-specific modifications. In this particular case, our experience aligns with this common trend, as we do not focus on developing novel learning algorithms or network architectures. The dataset used in this study comprises 6,400 individual images of mangoes, with each image being assigned a quality rating of either A, B, or C. Nevertheless, the images are captured informally by individuals inside mango processing facilities, resulting in challenges such as disruptive ambient noise, inconsistent proximity and orientation of the mangoes being photographed, and a wide range of lighting conditions. In order to address these challenges, a set of data preparation approaches are employed to enhance the quality of the data. One notable approach involves the utilization of Mask R-CNN [5], which is customized based on our human annotations of the border of the target mangoes in the images, to effectively exclude a significant portion of the irrelevant background. The DL models used for our classification challenge include of renowned architectures that have had success in the past. Specifically, these models include Alex Net [2], VGGs [3], and ResNets [4], all of which have been recognized as winners in the ImageNet Large Scale Recognition

[1]*Ph.D.Research Scholar, Department of CSE, Annamalai University.* [2]*Professor, Department of CSE, Annamalai University.* [3]*Professor, Department of CSE, Anurag University.* *Email:*[1]*vignesh2friend@gmail.com,*[2]*suha_babu@yahoo.com,*[3]*vbalaram 23@gmail.com.*

Challenge (ILSVRC) [6]. The ILSVRC is a notable competition that involves classifying a vast collection of over a million images into 1000 distinct categories. Additionally, the effectiveness of transfer learning [7] is also discussed in transferring information acquired from large datasets in broad domains to specialized domains with restricted data availability. To use this approach, we utilize the pretrained weights from ImageNet, which are made accessible via the torch vision package1. In addition to the aforementioned well-known models, our motivation for incorporating multi-task learning into classification problems, as shown in reference [8], leads us to explore the integration of a convolutional auto encoder into the CNN classifier. This integration is achieved by joint optimization throughout the training process. The justification for examining such networks is as follows: There are two key observations to be made: Firstly, the inclusion of an auto encoder in the network compels it to retain crucial information necessary for reconstruction while extracting features for classification. This has the effect of regularization. Second, the network's latent features include compressed reconstruction information, which may benefit other tasks. Thus, the auto encoder component may be left alone and a new classifier component added for defect type categorization. The two types of networks are called "single-task convolutional neural networks" and "convolutional auto encoder-classifiers". In the context of complex applications, it is often essential to not only achieve high accuracy but also to provide a comprehensive explanation for the choices made by the model. Utilizing the model's "explainability" not only facilitates the acquisition of concise insights into the sophisticated decision-making process of the model, but also cultivates users' confidence in the opaque nature of deep learning [9]. Within the context of our research, we will examine the performance of our suggested solutions in addition to conducting tests and engaging in debates. This will allow us to get a comprehensive understanding of the functioning of the model. These methods include the use of saliency maps to analyze the model's attention during prediction [10], as well as the application of PCA to discern the differentiation of mangoes of varying quality classes in the latent feature space [11]. Through the use of these strategies, the automated grading system has the capacity to provide human supervisors with additional information in conjunction with the predictions, so enhancing the evaluation process of grading outcomes. In light of the aforementioned, this study's contributions may be summarized as follows:

- An enhanced version of UNet++ is presented for the purpose of intelligent background removal. This improved model has spatial attention and channel attention gates, which are designed to learn the significant representations of the preprocessed pictures by identifying their spatial and channel-wise importance.
- A deep neural classifier using a scalable metaheuristic optimizer is proposed to effectively address the complex attributes associated with illness detection and fruit grading.
- The radiomics-based feature extraction method is designed to analyse many characteristics of mangoes, including colour, volume, size, shape, flaws, and particularly density.

The paper has been organized in the following manner: Section 1 comprises the project's introduction. Section 2 provides a concise summary of the literature survey that was conducted. Section 3 elucidates the operational procedures of the system that we have put forward, as well as its execution. Section 4 presents the derived conclusions and achieved outcomes. In conclusion, Section 5 of this study provides a summary of our findings and discusses potential avenues for further research.

2. Related Works

This section reviews image-based studies on image categorization, fruit identification, classification, and illness diagnosis. The majority of fruit identification and disease detection research has utilisedcolour and texture to categorise. The vast majority of fruit identification research have focused on tree-based fruits. This research only covers fruit variety classification methods. The present literature largely focuses on fruit disease detection, restricting study in this field. This section discusses research methods to learn about the latest research on the topics covered in this work.

In [12] utilises the CNN method to extract features, while using the HOG approach to capture shape and texture information. The characteristics that have been retrieved are then used to input into a disease classification model in order to identify diseases. The efficacy of the suggested approach is substantiated by empirical evidence, exhibiting exceptional precision in both illness identification and categorization endeavours. The performance of the CNN-HOG hybrid model surpasses that of either the CNN or HOG techniques in the context of mango illness detection and classification, therefore illustrating the mutually reinforcing characteristics of these two approaches. In [13] approach utilises a low-cost Vector Network Analyzer (VNA) device, which is enhanced by including both the K-nearest neighbour (KNN) algorithm and a Neural Network model. In [14], a novel approach is suggested for predicting artificially ripened mango fruits using a CNN. The proposed method incorporates the use of binary cross entropy as a means to minimize the loss

during the prediction process. In [15], uses CNN to classify four types of fruits: Banana, Papaya, Mango, and Guava. These fruits are categorized into three phases based on their ripeness: raw, ripe, and over-ripe. The model uses a dataset of local fruits to examine and analyse their life cycle across several developmental phases. The study examined the use of a single and two-stage deep learning framework, namely YOLO and R2CNN, in the context of both upright and rotating bounding boxes [16]. The weighted F1 scores obtained by the models MangoYOLO(-upright), MangoYOLO-rotated, YOLOv3-rotated, R2CNN(-rotated), and R2CNN-upright for a validation picture set and total panicle count were 76.5, 76.1, 74.9, 74.0, and 82.0, respectively. In [17], introduced a novel approach for picture categorization, using lightweight CNNs, with the objective of enhancing the efficiency of the checkout procedure in retail establishments. This study presents a novel dataset of images that include three distinct categories of fruits, namely those contained inside plastic bags, those without plastic bags, and those that fall into neither category. To enhance the accuracy of classification, the CNN design incorporates many input features. The inputs include three elements: a single RGB colour, histogram, and centroid obtained by K-means clustering. In the study conducted by [18], three feature vectors were generated. The color_moment feature extraction technique encompasses the calculation of statistical properties, namely the mean and standard deviation, for the three-color channels (RGB). The binarized representations of the images of the fruits were used in order to extract attributes that are mostly dependent on their physical structure. Subsequently, a vector including many features, including colour moments and shape characteristics, was constructed. In [19] focuses on the development of an automated system for classifying mangoes. This system aims to monitor and assess the quality of mangoes prior to their packing and shipment to the market. The present study focuses on conducting research, designing, and fabricating a mango classification model. Additionally, an autonomous mango classification system will be developed by integrating image processing technology with artificial intelligence.

When examining the prevailing pattern in DL, it is evident that researchers in the domain of fruit recognition and classification mostly use the supervised learning approach. Consequently, this necessitates the utilisation of extensive datasets for training purposes. The process of manually annotating datasets is a time-consuming and labor-intensive task. As a result, researchers are advised to explore the use of unsupervised learning methods. The deployment of deep learning models on mobile devices is deemed impractical owing to the substantial memory requirements during testing and the time-consuming nature of the models. The investigation of methods to decrease complexity and examine efficient models that do not compromise outcomes is of utmost importance for researchers. Hyperparameter selection has always been a significant challenge in the field of deep learning, particularly when applied to novel tasks. The impact of this factor on the model's outcome may be both positive and negative, hence exerting a significant influence on the ultimate training outcomes.

3. Proposed Methodology

The dataset is first populated with images of mango fruits, as seen in Figure 1. The datasets underwent preprocessing using a median blur filter to eliminate noise, followed by the use of the UNET++ algorithm. This algorithm is capable of properly identifying and extracting the topic from images that are complicated and provide challenges.

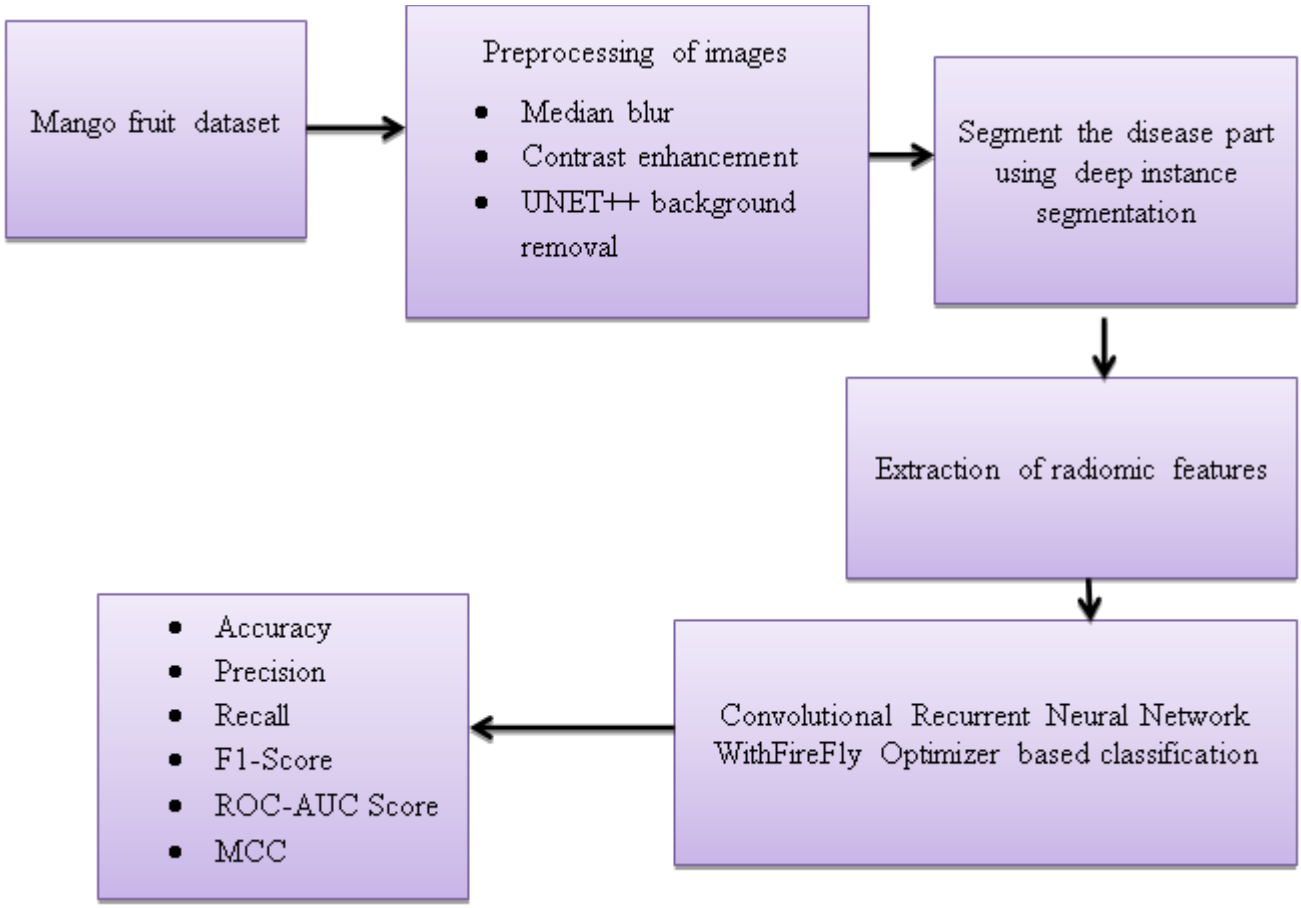

Fig-1 system architecture for fruit classification

Following the elimination of the noise, the overall contrast is improved. By changing the image to a colour space with image luminosity included as one of its components, such as the L*a*b* colour space, is often used to accomplish this improvement. The luminosity layer L* is the only one to which the contrast adjustment procedure is particularly done, and only then is the image returned to RGB colour space. The preprocessed picture is inputted into the deep instance segmentation method, which performs segmentation of the disease regions, followed by the extraction of randomic features. The collected characteristics are categorisedusing a Convolutional Recurrent Neural Network (CRNN) with a classifier based on the FireFly Optimizer.

Preprocessing of images

In order to enhance the original image's quality and streamline subsequent processing stages, it is necessary to do preprocessing. The preprocessing phase of this investigation encompasses many processes, including noise reduction, background removal, and contrast enhancement. Let (i, j) represent the block of an image's beginning index. Let the grayscale pixel value at the index (p, q)be $x_{p,q}$. The median operator is $median(.)$. Let $r \times s$ represent a window's size. Let $y_{i,j}$ be the median filter's result. Then

$$y_{i,j} = median([x_{i,j}\Delta x_{i,j+s-1}\Delta x_{i,r-1,j}\Delta x_{i+r-1,j+s-1}]) \quad (1)$$

It is evident that the process of median filtering involves the arrangement of pixels inside a designated window, followed by the selection of the pixel value situated in the centre of the sorted pixel arrangement. This selected pixel value is then assigned as the output of the filtering operation. It is noteworthy to mention that median filtering is a common nonlinear operator used for the purpose of attenuating impulsive noise. In contrast to standard linear filtering approaches, this particular nonlinear filtering technique exhibits reduced blurring effects and superior edge preservation capabilities when applied to images. Following the filtering process, the U-Net++ architecture was used to eliminate the background from the pictures. This was achieved by theutilisation of skip connection. Furthermore, the encoder and decoder components also allow the incorporation of long-range connections. Consequently, the decoder is able to effectively integrate hierarchical feature maps from the encoder. This integration enhances the precision and scalability of the network.Let $x^{i,j}$ stand for the node's output, where i stands for the ith down-sampling layer along the decoder route and j for the j th convolution layer along the skip pathway. The maps with the backdrop removed $x^{i,j}$are denoted by:

$$x^{i,j} = \begin{cases} \delta^{cov}(x^{i-1}, j) & j = 0 \\ \delta^{cov}\left(\delta^{cat}\left(\delta^{cat}(x^{i,0}, x^{i,1}), \delta^{up}(x^{i+1,j-1})\right)\right) & j > 0 \end{cases} \quad (2)$$

Where δ^{cov} stands for a convolution operation with an activation function, δ^{cat} stands for concatenation, and δ^{up} stands for an up-sampling layer. Nodes in the encoder sub-network are represented by $X^{i,j}$ if $j = 0$. When $j > 0$, the concatenation results of all nodes in the same level are represented by $X^{i,j}$, which also includes the up-sampled result of $X^{i,j}$with deeper, coarser, and sematic information. The backdrop picture is created from D recent background pixels, and it depends on the optical flow image's characteristics. The backdrop pixel for pixel (m, n)at time t is determined by the following formula:

$$B_t(m,n) = \begin{cases} \frac{\sum_{i=p_n-D}^{p_n} L_{bag}(m,n,i)}{D}, if\ P_n > P_0 \\ \frac{A_1+A_2}{D} \qquad otherwise \end{cases} \quad (3)$$

Where, $A_1 = \sum_{i=0}^{p_n} L_{bag}(m,n,i), A_2 = \sum_{i=70-(D-p_n)}^{70} L_{bag}(m,n,i). D$ is the total number of pixels used to create the background picture, $L_{bag}(m,n,i)$is the i^{th} item in the pixel bag corresponding to pixel position (m, n) and $B_t\,(m,n)$is the value of the pixel at location (m, n) at time t.

The variable D is adjusted according on the attributes of the picture sequence. This is necessary since maintaining a constant value for D leads to either a fuzzy background image or an outdated background image in complicated circumstances, such as camera jitter. Therefore, it is necessary to modify the value of D in accordance with the dynamic qualities that are inherent in the picture. The use of optical flow imagery allows for the extraction of both spatial and temporal mobility inside a given frame, relative to the preceding frame. The presence of spatial or temporal mobility is indicated by the non-zero pixel values in the optical flow picture. Contrast enhancement is achieved by using contrast-limited adaptive histogram equalisation after the removal of the background. An innovative technique for automated contrast enhancement is proposed, showcasing remarkable effectiveness when applied to images with predominant high-amplitude histogram components located toward the left portion of the non-zero histogram component (NZHC) region. This proposed method is a combination of Histogram Equalisation (HE) and Fast Gray-Level Grouping (FGLG). The suggested approach is implemented through a series of distinct steps. The discrete function $h(r_k) = n_k$, where r_k is the kth intensity level and k is the number of pixels in the picture with intensity r_k, creates the histogram for an image with intensity levels in the range n$[0, L-1]$. M and N represent the image's row and column dimensions, as is common. Each component of a histogram is divided by the total number of pixels in the picture, represented by the product MN, in order to standardise the data. This means that for k0,1,2 …. $L-1$, $p(r_k) = \frac{n_k}{MN}$, yields a normalised histogram. Assuming that a histogram was computed for an input picture I with intensity levels between $[0, L-1]$, the fundamental steps for histogram separation are provided below. Determine where on the grey scale P_{hist}, the histogram component with the largest amplitude, is located.If P_{hist} is in the left segment but not in the first component of the NZHC, the histogram may be split into two sub-histograms: the first starting from 0 to (P_{hist}-1) intensity and the second from $P_{hist}$to maximum intensity level (L-1).

Segmentation of disease from preprocessed image

We divide the instance segmentation issue into the labeling/segmentation and the instance recognition challenges. We write $D = \{(x_n, y_n, z_n), n = 1,2, \ldots N$ represents the training dataset's first input, where $N$the number of images is. As we now examine each picture individually, we remove the subscript n for simplicity. The instance label is represented by $Z = \{R_k, k = 0,1,2, \ldots K\}$, where $R_k = \{(p, q)\}$ signifies the coordinates set of pixels within of region R_k and $X = \{x_j, j = 1,2, \ldots |X|\}$is the raw input picture. $Y = \{y_j, j = 1,2, \ldots |X|\}, y_j \in \{0,1\}$is the matching segmentation label. The background region is shown when k is equal to 0, and the matching instance is indicated when k has other values. The overall instance count is K. Areas of the picture that meet the following relations include:

$$R_k \cap R_t = \emptyset, \forall k \neq t \quad (4)$$

$$\cup R_k = \beta \quad (5)$$

βstands for the whole image region. The goal is to discriminate between each incidence of the illness while segmenting the disease region. In the subproblem of segmentation and labelling. The labeling/segmentation outcome is represented by Y'. As for the cost function:

$$dist\ (Y, Y') = \frac{1}{|Y|}\sum_{j=1}^{|Y|} \delta(y_j \neq y'_j) \quad (6)$$

$$y'_j = arg\ max_y P(y|X) \quad (7)$$

Z' refers to the instance prediction in the instance recognition subproblem. As for the cost function:

$$dist(z, z') = 1 - \frac{1}{K}\sum_{K'=0}^{K'} L(R_{k'}, Z) \quad (8)$$

$$\text{Where, } L(R_{k'}, Z) = \begin{cases} 1, & \exists k \neq 0, \frac{R_{k'} \cap R_k}{R_{k'} \cup R_k} \geq thre \\ 0, & otherwise \end{cases} \quad (9)$$

Where the instance label area is designated by $R_k \in Z$, while the instance segmentation prediction region is designated by $R_{k'} \epsilon z'$. The entire projected region count is represented by K'. $thre$is a threshold in this method that is set at 0.5. The algorithm considers an area to be an instance prediction when the ratio of the gland instance's overlap with labels in a given prediction region exceeds a specific threshold. SGD cannot be used to train instance recognition because the cost function is non-differentiable. By using object and edge detection, we herein approximate instance identification. When the four nearby pixels (above, below, right, and left) are part of the same instance, e_j equals 0, and we create edge labels $E = \{e_j, j = 1,2, \ldots |X|\}, e_j \in \{0,1\}$and object labels O via Y and Z to train edge detector and object detector. The fruit instance's smallest bounding box is indicated by the letter O.

Extraction of randomic features

The instance segmentations underwent further analysis via the extraction of radiomics features using a Python programme called PyRadiomics (version 3.0.1), which adheres to the definition-based standardisation approach for imaging biomarkers. The radiomic characteristics from the GTV were extracted using PyRadiomics [20] inside the 3D Slicer software. The GTV was reconfigured to a 1×1×1 mm3 voxel size. The initial statistical characteristic was utilized in creating a histogram that illustrates the spread of pixel values, enabling the extraction of relevant features from this histogram. The texture characteristic was used to transform the pixel value relationships into a matrix, enabling the quantification of picture uniformity and heterogeneity. The radiomic characteristics used in PyRadiomics were derived from the Image Biomarker Standardisation Initiative. This initiative has developed verified definitions and standards for these features, with the exception of four specific features. The variables in question are shape_Maximum2DDiameterSlice, shape_Maximum2DDiameterColumn, first-order_TotalEnergyand shape_Maximum2DDiameterRow. Three separate feature selection approaches were used to decrease the 107 radiomic characteristics that were derived from the GTV. The full training dataset was subjected to the three aforementioned procedures. This strategy was used in the current investigation to identify the most efficient feature selection strategies in terms of their ability to forecast. Following that, the subgroup analysis involved the utilization of the selection technique that demonstrated the most effective performance.FS1 uses a methodical selection process to choose characteristics that exhibit resilience via test-retest reliability and a variety of segmentation strategies.The test-retest technique makes use of a dataset that Zhao et al. created in order to assess the illness's variability taking into account its volumetric, bidimensional, and unidimensional characteristics. The test-retest technique uses a radiomic analysis approach to evaluate illness by comparing two pictures. It identifies characteristics that exhibit large changes during this little time period and removes them from consideration, since they are deemed less reliable. The two variables' measurements were compared using the Concordance Correlation Coefficient (CCC), which was used to guide feature selection. The CCC threshold was set at >0.85. Feature Selection 2 (FS2) involves the exclusion of one of the correlated characteristics from the study, deeming it redundant. A correlation coefficient with an absolute value of 0.8 or above was considered the threshold for indicating a substantial link between two characteristics. The combination of FS1 and FS2 results in FS3. Following the use of test-retest and multiple segmentation techniques, the selection of robust features is carried out. After that, the threshold of 0.8 is used to Pearson's correlation analysis in order to identify non-redundant features.

LASSO Cox regression model

This research included the development of two unique models. The initial model, primarily employing the chosen features from FS1, FS2, and FS3, was only based on radiomic properties. The second model included clinical and radiomic data, as well as the chosen characteristics with clinical forecasts. A prognostic model for predicting survival outcomes was built using the LASSO Cox regression model. Radiomic analysis has often been done using the regression model under examination. The LASSO process is influenced by the weight of the constraint term, denoted as λ, which is applied to the likelihood function. This weight determines the extent to which the regression coefficients are shrunk towards zero, resulting in the elimination of coefficients associated with unnecessary characteristics. The effectiveness of learning models is heavily influenced by the parameter λ, where a higher value of λ leads to a more simplified model, while a lower value of λ diminishes the significance of weights and may result in overfitting [21]. To address concerns related to model simplification and overfitting, a fivefold cross-validation technique was employed, aiding in the identification of the optimal value of λ for the dataset.

During the process, the dataset used for estimating model parameters was randomly divided into five subsets.Among these subsets, four were assigned as training data, while the remaining group was designated for validation purposes. The training data was used to optimise a model for each λ, then implementing it on the validation data. The mentioned method was iteratively performed, with the therapy being provided on five occasions. Subsequently, the square errors obtained from each of the five repetitions were computed for every λ value. These square errors were then averaged to ascertain the ideal λ value that would yield the least mean square error. Rad scores were computed through the utilization of linear combinations of features, featuring non-zero coefficients identified at the optimal λ. The $rand$ score may be calculated by summing the features with nonzero coefficients, each multiplied by their respective coefficients (β), as seen in Equation 10.

$$rad\ score = \sum_{i=1}^{n} \beta_1 . feature\ (i) \quad (10)$$

The threshold for partitioning the training dataset into high- and low-risk groups was established by calculating the median rad score using Equation (10).

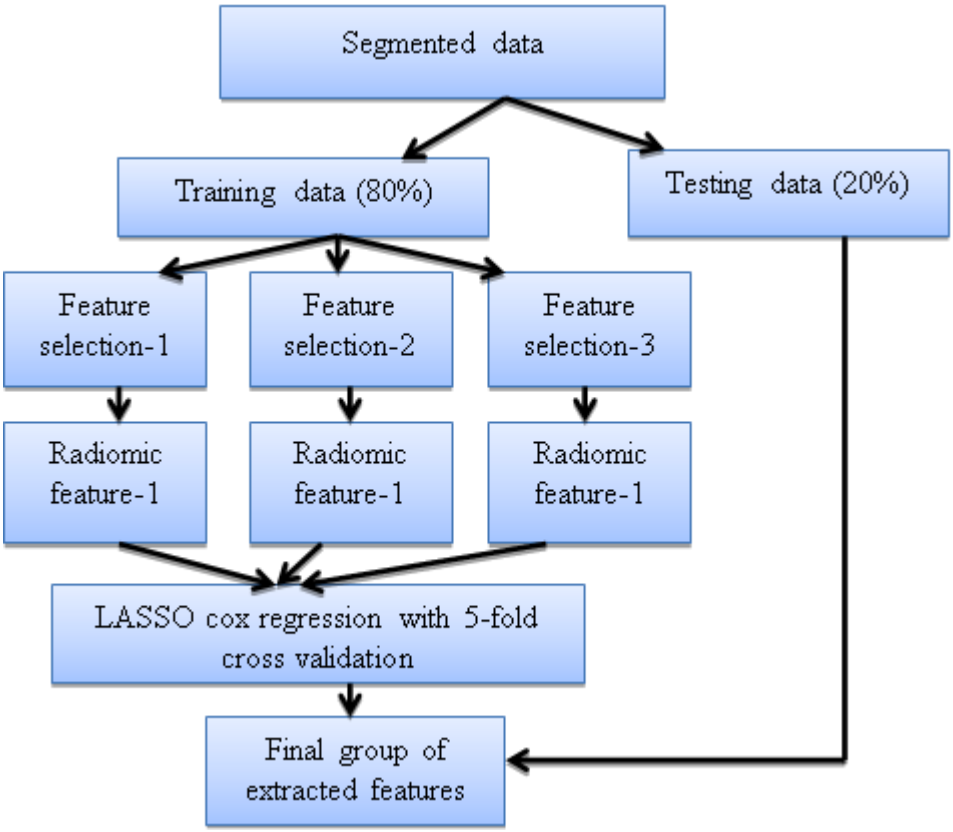

Fig-2 feature extraction using radiomics

The developed model in subgroup analysis was validated using a fivefold cross-validation, as seen in Figure 1. To maintain a steady ratio between patient fatalities and survivors, the information relevant to the subgroups was separated into five parts using stratified sampling. After that, one segment was chosen as the test dataset and four segments were designated as the training dataset. Moreover, the use of cross-validation may successfully mitigate duplication in the constructed model. This is due to the fact that, unlike the bootstrap approach, cross-validation partitions the dataset in a manner that prevents any kind of repetition. During the analysis of the entire dataset, the cumulative C-index values of both the radiomic and mixed models were calculated for each feature selection approach. The feature selection technique that produced the greatest C-index was then applied to the training dataset. The radiomic models and combination models were built using the LASSO Cox regression model, which was also employed for the analysis of the whole dataset. Next, the test dataset was used to assess each model's performance. The predictive value of the created model was evaluated using the C-index and Kaplan-Meier survival analysis. The evaluation makes use of the C-index, which is determined by averaging the C-index values obtained from the five models developed by fivefold cross-validation. The identical validation process used for the subgroup analysis was also employed in the analysis of the entire dataset, ensuring a comprehensive and

equitable examination of the results obtained from both the complete dataset and the subset dataset.

Classifier for fruit classification

The classification of the retrieved features is performed using a CRNN using a classification approach based on the FireFly Optimizer. The present discussion provides a first explanation of the firefly method activation process. The low-level heuristics refer to a collection of issue-specific rules that are created to provide solutions for each individual problem occurrence. A novel set of solutions is generated by the use of various search procedures, which include the transformation or combination of one or more existing solutions. This work use the FA-based search process as a method for generating novel solutions. Following the establishment of the novel solutions, the overarching approach emulates the process of selection. The high-level approach implements an automated process for heuristic selection, whereby the heuristics are sequentially chosen and applied to the solutions. The heuristics used in this study are derived from a collection of rules established by a low-level approach. The selection of these heuristics is performed via an online process designed specifically for heuristic selection. The empirical reward and confidence level variables are the major metrics used to evaluate the efficacy of the heuristics. The incentives acquired from previous performance are referred to as empirical rewards, whilst the frequency of heuristic use indicates the amount of confidence. Based on the evaluation of these two factors, the appropriateness of the heuristics for the present operational condition is determined. Therefore, the aforementioned heuristics are used in the firefly foraging process to optimise the solutions. The heuristics are initialised as the population of fireflies, denoted as $x_i(i = 1,2, \ldots . z)$. In the context where the firefly with the highest brightness at position x represents the optimal solution, it is possible to formulate this scenario as a maximisation problem denoted by $I(x) \sim f(x)$. Here, $I(x)$ represents the intensity of light emitted by the firefly, and $f(x)$ represents the fitness function associated with the solution. In the present research, the error rate will serve as the fitness metric, so transforming the issue into a minimization objective,

$$I(x) = \begin{cases} \frac{1}{f(x)}, if\ f(x) > 0 \\ 1 + |f(x)|, otherwise \end{cases} \quad (11)$$

The method is based upon the light intensity function, denoted as $I(x)$, and the attractiveness function, denoted as β. As the light intensity and attractiveness of the source firefly rise, there is a corresponding reduction in the distance (r) between the originand the target firefly,

$$I(r) = \frac{I_0}{1+\gamma r^2} \quad (12)$$

In this context, I_0represents the initial intensity of the light source, whereas γ denotes the absorption coefficient. The approximation of this may be represented by the Gaussian form.

$$I(r) = I_0 \exp(-\gamma r^2) \quad (13)$$

The degree of attraction, represented as β, has a positive correlation with the magnitude of light seen by the next firefly. Hence, the parameter β, denoting the degree of attraction, has a direct correlation with the intensity of light present in the solutions.

$$\beta(r) = \beta_0 \exp(-\gamma r^m) \quad (14)$$

Where, β_0 represents the attractiveness value at $r = 0$, while m denotes the total quantity of rounds. In the context of CNN optimisation, it is essential to decrease the computational complexity in order to minimise resource use. The expression of attraction has been adjusted for practical use as shown below.

$$\beta(r) = \frac{\beta_0}{1+\gamma r^2} \quad (15)$$

The distance between two fireflies (nodes) i and j, located at positions $x_i$and x_j, is represented as $r_{i,j}$ and is calculated using the Cartesian distance measure.

$$r_{i,j} = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2} \quad (16)$$

The Cartesian coordinates of x_i and x_j are denoted as $x_{i,k}$and $x_{j,k}$, respectively, where $d$denotes the dimension of the space. The firefly exhibits movement towards the firefly with the highest fitness value, and the location of this firefly is updated iteratively using the subsequent equation.

$$x_i = x_i + \beta_0 e^{-\gamma r^2}(x_j - x_i) + \alpha(rand - \tfrac{1}{2}) \quad (17)$$

In this context, the variable "rand" represents a pseudo-random number that falls within the range of [0, 1]. The parameter "α" is used as a regulating factor for the step-size. To optimise results, it is recommended to fine-tune the values of α and β0 within the range of α ∈ [0,1] and β_0 = 0.2. The values of α and β0 are changed so that α ∈ [0,1] and β0 = 0.2.The heuristic is implemented on each of the solutions acquired by the firefly, taking into consideration the light intensity and attraction of the firefly method. The option that has been identified as the optimal global solution includes the solution that is intended for implementation. The heuristic is used in tandem with the selected solution to develop a new set of

solutions. In this particular phase, the use of serial scheduling and twofold justification procedures is implemented. The use of serial scheduling entails the selection of solutions in a manner that avoids the interleaving of possible solutions. Similarly, the twofold justification method is a straightforward local search approach that systematically explores potential solutions by iteratively adjusting certain parameters to optimise the quality of the search process. The novel solutions are subjected to a comparative analysis, whereby their respective qualities are examined. This study, conducted in terms of configuration, aims to ascertain whether to include the identified whiteners into the current set of solutions or discontinue their use in order to make room for newer solutions in future iterations. After using low-level heuristics to produce original answers and then selecting them using a high-level strategy, these solutions are preserved inside the archive's non-dominated set of solutions which is used to classify the archive into many tiers, so enabling the retention of more current solutions. The first level is assigned to the solution with a significant degree of importance, while the subsequent level is allocated to the second-best priority, and this pattern is reciprocated. The firefly optimizer algorithm employs a selection process that considers the Pareto-front to choose the solutions from the archive. The optimal configuration, identified as the best option, is then returned as the final outcome.

Combination of CNN-FOA

The structure of the CNN is referredas $h = \{h_{cl}, h_{pl}, h_{fcl}\}$, where $h$denotes the collection of structural parameters, convolutional layers h_{cl}, pooling layers h_{pl}, and fully-connected layers h_{fcl}. The convolutional parameter set, denoted as h_{fcl}, is described in the problem as $h_{cl} = \{C_1, C_2, \dots C_{a-1}\}$. where$a$shows the number of convolutional layers. Each element $C_i = (C_{count}, C_{size})$ in h_{cl}shows the configuration tuple of the i-th layer in the CL.

C_{count} represents the count of kernels per i-th layer, whereas C_{size} represents the kernel size of the i-th layer in the context of CL. Similarly, let $h_{pl} = \{p_1, p_2, \dots p_{b-1}\}$represent the set of PL, and $h_{fcl} = \{s_1, s_2, \dots s_{n-1}\}$represent the set of FCL. Here, b and $n$signify the number of layers in PL and FCL, respectively.The size of each layer in the PL is represented by the variable p_i, whereas the size of each layer in the FCL is signified by the variable s_i. Let H be the set of potential configurations for CNNs, and the goal is to identify the configuration $h \in H$ that yields the lowest classification error.Due to the NP-hard nature of this optimisation issue, the search space of the FOA is constrained by the upper and lower borders. The process of optimising the hyper-parameters of a CNN involves selecting the most suitable combination of problem parameters. This selection is achieved by the firefly search process, which consists of two main components: exploitation and exploration. The regulation of exploitation in FOA is determined by the given values of $\beta$and γ. When the value of $\beta = \beta_0$and $\gamma = 0$, the solutions tend to migrate towards those with the biggest step size due to the presence of maximum exploitation and little exploration.When the value of $\beta = 0$and $\gamma = \infty$, the solutions exhibit random movements due to the greatest level of exploration and the complete absence of exploitation.The determination of the trade-off between the values of β and γmay be achieved by the adjustment of their respective values, which can be informed by practical experimentation. The best value of γ is often regarded as being within the range of 0.01 to 100.Similarly, the values of $\beta = 0.2$ and $\propto \epsilon[0,1]$ are taken into consideration in order to achieve optimum performance. The issue of hyper-parameters is represented as $h = \{h_{cl}, h_{pl}, h_{fcl}\}$ inside the framework of FOA, where each solution corresponds to a potential configuration of a CNN.The population of the FOA is represented as P at each time point t, and it comprises n firefly (solutions). The population at time t is indicated as $P^t\{h_0, h_1, h_{n-1}\}$. During the initialization step, a population of solutions is formed using a uniform distribution,

$$x_{i,j} = lb_{i,j} + \varphi.(ub_j - lb_j)$$

In this context, the notation $x_{i,j}$ is used to represent the j-thposition of the i-th solution. The symbol φ denotes a pseudo-random number within the interval [0, 1].In order to optimise computational efficiency and minimise the search space, lb_j and ub_jvalues for the count of kernels have been established as 1 and 128, respectively.

The hyper-parameters, such as the dropout rate and learning rate, were not optimised in this work because to their mostly continuous nature. The optimisation of hyperparameters that include integer values is exclusively performed using the FOA. The equations (11) to (17) shown in the FOA may be effectively used for the optimization issue of CNN, since the upper and lower limits for each parameter are set to values larger than 1. The fitness function used is the classification error rate. The primary goal is to reduce the rate of error while evaluating the fitness of the i-th solution, which may be mathematically represented as

$$x_i = \frac{1}{1 + |error_i|}$$

Where the variable $error_i$reflects the test dataset's classification error for the i-th solution.

4. Experimental Analysis

The classification performance was assessed using AUC, accuracy, precision, f1-score, recall,ROC-AUC score and Matthews Correlation Coefficient (MCC). These parametes are analyzed for existing CNN-HOG [12], CNN [14], lightweight Convolutional Neural Networks (L_CNN) [17] and proposed CNN_FOA.

Dataset description- [22] The Fruit-360 dataset has 81 distinct classes, with each class representing a particular kind of fruit. The dataset has been segmented into three separate subgroups: the training set, including 41,322 images; the validation set, consisting of 9,744 images; and the testing set, encompassing 4,133 images. The collection consists of a total of 55,244 pictures. The fruits underwent rotational motion through the use of a low-speed motor running at a rate of 3 rpm, and a film lasting 20 seconds was shot to document this process. The same white backdrop color was applied to all images in order to address the issue of inconsistent lighting situations, hence ensuring a consistent visual appearance. Among the 81 distinct fruit categories, it was possible to get just one image for each respective category. All images were standardized to size of 100 × 100 pixels, and a white backdrop was uniformly applied.

Powdery mildew, a major mango disease, affects practically all cultivars in all mango-growing countries (Figure 2). The fungus Oidiummangiferae causes severe yield losses. A fungus infects inflorescences, leaves, and early fruits. Mango stems endophytically host cytosphaeramangiferae. Endophytic mycelium may cause fruit infection and stem cankers. Monotypic genus.

	Powdery mildew, attributed to Oidiummangiferae, results in a whitish-gray film on the panicles, leading to the browning and graying of affected flowers.
	The bacterium responsible for causing black spot is known as Xanthomonascampestrispv. mangiferae-indicae. It has the capability to infect leaves, twigs, and fruit.
	Mangoes at 6 days old are depicted here. The skin color of these mangoes remained consistent from day 6 to 22 following harvest.
	Mango skin contains polyphenol oxidase.During transportation, when the mango skin is scratched, rubbed, or otherwise exposed to pressure, the polyphenol oxidase will undergo a chemical reaction, and under the action of oxygen, it will turn into a black substance, that is, the black spots on the surface of the mango skin.

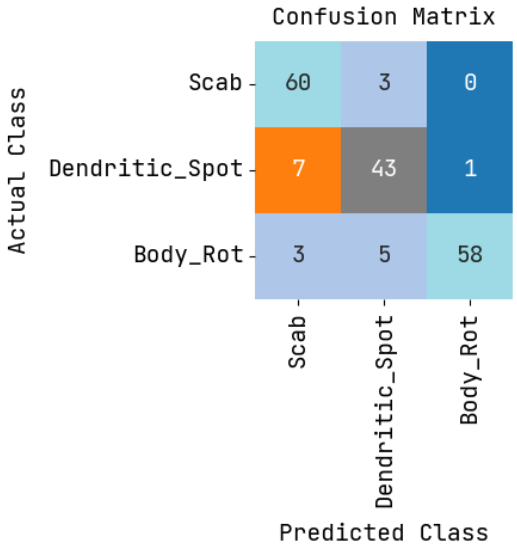

Fig-3 confusion matrix of CNN_FOAfor testing

The confusion matrix for features obtained during testing for the CNN_FOA model is shown in Figure 3. The rows and columns of the matrix represent the expected and actual class of data pertaining to the prediction of mango fruit. The networks that have been subjected to testing and have been accurately or inaccurately identified are visually represented by the contrasting colours in a crosswise pattern. The table below displays the implementation of each actual class, whereas the rightmost column represents each projected class.

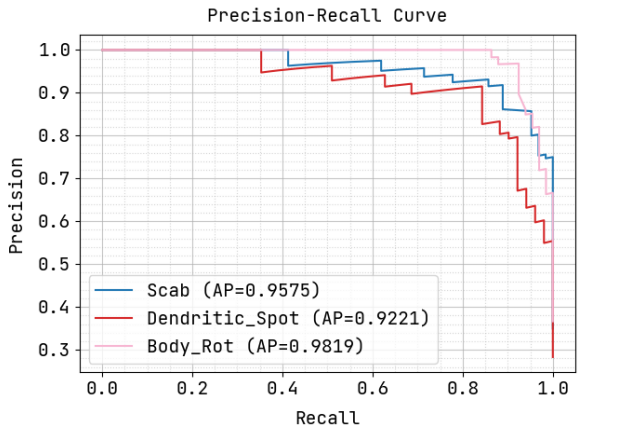

Fig-4 precision-recall curve of CNN_FOA for testing

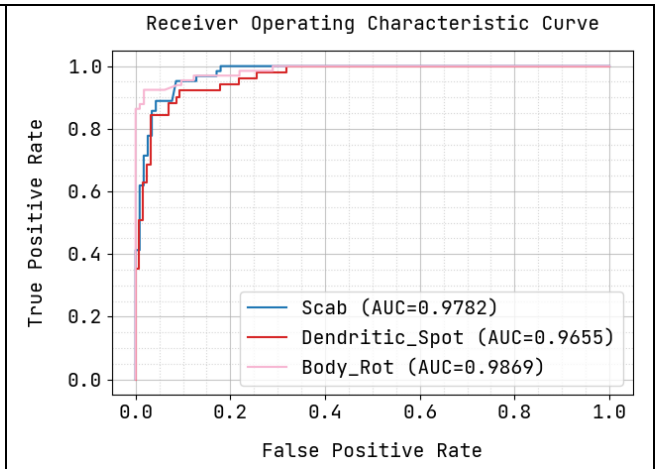

Fig-5 ROC curve of CNN_FOA for testing

The precision-recall curve for testing CNN_FOA is depicted in Figure 4. The x-axis and y-axis labelled the precision and recall value. During the experiment, the AP values obtained for the scab, dendritic spot, and body rot categories were 0.9575,0.9221 and 0.9819. Figure 5 is the ROC curve for the evaluation of the CNN_FOA. The x-axis and y-axis labelled the true positive rate and false positive rate. During the experiment, the AUC attains its highest values of 0.9782,0.9655 and 0.9869 for scab, dendritic spot, and body rot, respectively.

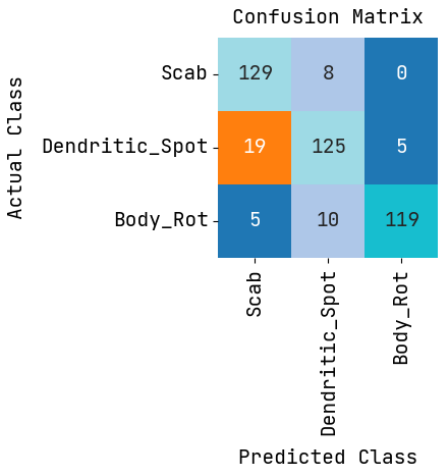

Fig-6 confusion matrix of CNN_FOAfor training

The confusion matrix for features obtained from training the CNN_FOA model is shown in Figure 6. The rows and columns of the matrix represent the predicted and actual class labels, respectively, pertaining to the prediction of mango fruit. The networks that have been subjected to testing and have been properly or incorrectly classed are visually represented by the contrasting colours in a crosswise pattern. The table below illustrates the implementation of each actual class, whereas the rightmost column represents each projected class.

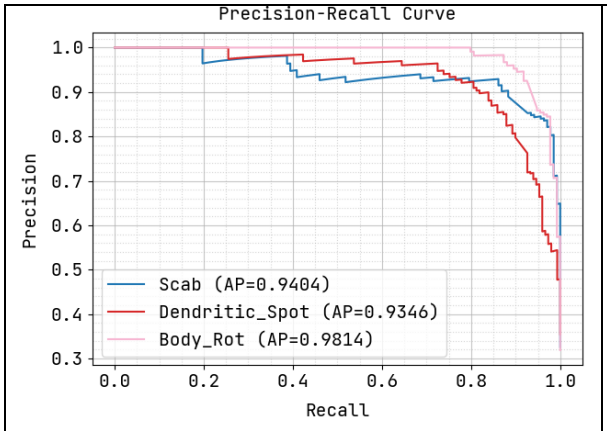

Fig 7 precision-recall curve of CNN_FOA for training

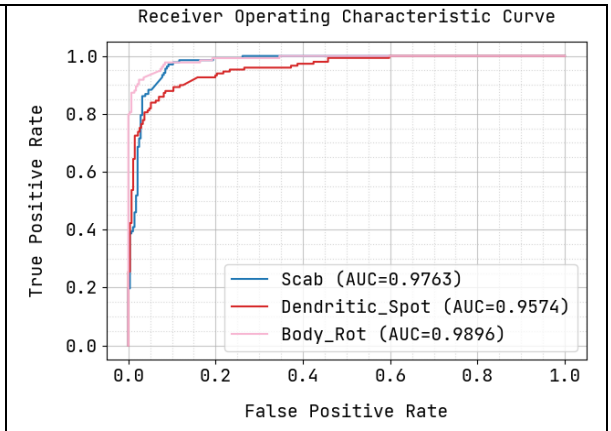

Fig-8 ROC curve of CNN_FOA for training

The precision-recall curve for training CNN_FOA is seen in Figure 7 above. The x-axis represents the recall value, while the y-axis is the precision value. In this, AP values attained by the AP model were seen to reach their highest levels at 0.9404 for the scab, 0.9346 for the dendritic spot category, and 0.9814 for the body rot category.Figure 8 depicts the ROC curve for the training of the CNN_FOA. The y-axis marked as the true positive ratio, while the x-axis marked as false positive ratio. The Area AUC attains its values, namely 0.9763 for the scab, 0.9574 for the dentritic spot condition, and 0.9896 for the body rot condition.

Table-1 Performance ofCNN_FOAfor testing and training

Metrics	Testing	Training
Accuracy (%)	0.9296	0.9254
Precision (%)	0.8944	0.8923
Recall (%)	0.8914	0.8895

F-Score (%)	0.8911	0.8894
ROC AUC Score (%)	0.9769	0.9744
MCC (%)	0.8408	0.8347

Table -2 Comparison of the Existing and Proposed Techniques

parameters	**CNN-HOG**	**CNN**	**L_CNN**	**CNN_FOA**
Accuracy (%)	95	82	86	97
Precision (%)	92	93	78	95
Recall (%)	82	79	69	94.3
F-Score (%)	68	63	57	94.2
ROC AUC Score (%)	73	84	83	98
MCC (%)	81	79	79	90

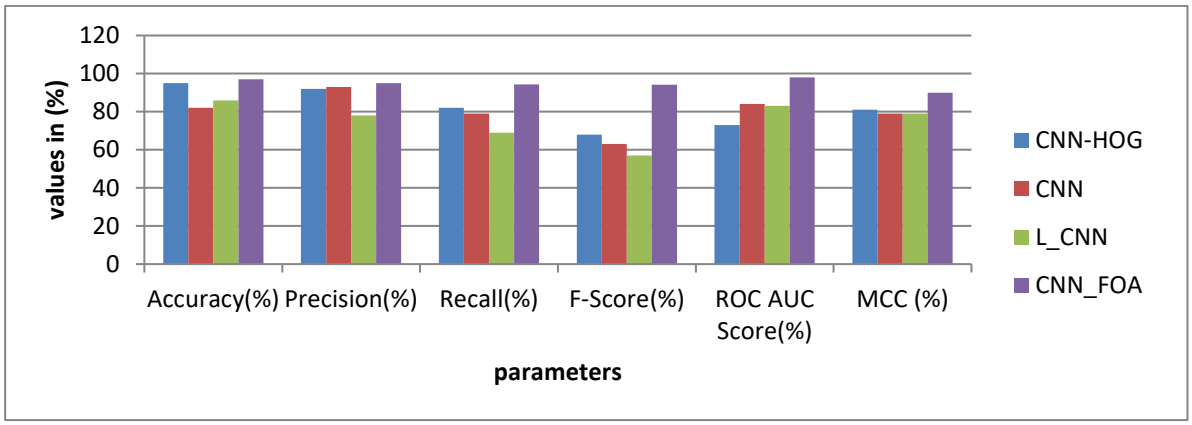

Fig-8 overall comparative analysis

The figure shown above, labelled as Figure 8, showcases the comprehensive performance of several existing techniques, namely CNN-HOG, CNN, and L_CNN, which is compared to the proffered, CNN_FOA. The developed technique demonstrates a much higher degree of accuracy, measuring at 97%. Additionally, it exhibits 95% of precision, recall of 94.3%, F-score of 94.2%, 98% of ROC AUC and 90% of MCC.

5. Conclusion

The mango is a much sought-after agricultural commodity that is extensively traded on a global scale. The evaluation of mango fruit quality has traditionally been conducted using manual methods, resulting in a significant investment of time and labour. Moreover, those responsible for inspecting the quality of mangoes must possess a high level of expertise in this domain. The evaluation of mangoes requires a manual assessment process, which unfortunately results in the destruction of the sampled fruit and therefore reduces the overall output. In order to address these concerns, many nondestructive methodologies have been developed, such as the inside inspection of the fruit. The experimental investigation will demonstrate that the suggested multi-disease categorization of Mango fruits has the potential to provide substantial and efficacious results. There is a need for improvement in the deployment of a comprehensive disease monitoring system across different plant species, as well as enhancing the applicability of the suggested methodologies. The next study involves the exploration of fusing strategies necessary for extracting critical characteristics and determining additional leaf samples in datasets.

Reference

[1] LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document

recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[5] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in ´ Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, no. 3, pp. 211–252, 2015.

[7] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1285–1298, 2016.

[8] Y. Zhang, K. Lee, and H. Lee, "Augmenting supervised neural networks with unsupervised objectives for large-scale image classification," in International conference on machine learning, 2016, pp. 612–621.

[9] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," IEEE Access, vol. 6, pp. 52 138–52 160, 2018.

[10] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," arXiv preprint arXiv:1312.6034, 2014.

[11] H. Hotelling, "Analysis of a complex of statistical variables into principal components." Journal of educational psychology, vol. 24, no. 6, p. 417, 1933.

[12] Sema, W., Yayeh, Y., &Andualem, G. (2023). Automatic Detection and Classification of Mango Disease Using Convolutional Neural Network and Histogram Oriented Gradients.

[13] Tran, V. L., Doan, T. N. C., Ferrero, F., Huy, T. L., & Le-Thanh, N. (2023). The Novel Combination of Nano Vector Network Analyzer and Machine Learning for Fruit Identification and Ripeness Grading. *Sensors*, *23*(2), 952.

[14] Laxmi, V., &Roopalakshmi, R. (2022). Artificially Ripened Mango Fruit Prediction System Using Convolutional Neural Network. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021* (pp. 345-356). Singapore: Springer Nature Singapore.

[15] Dandavate, R., &Patodkar, V. (2020, October). CNN and data augmentation based fruit classification model. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 784-787). IEEE.

[16] Koirala, A., Walsh, K. B., Wang, Z., & Anderson, N. (2020). Deep learning for mango (Mangiferaindica) panicle stage classification. *Agronomy*, *10*(1), 143.

[17] Rojas-Aranda, J. L., Nunez-Varela, J. I., Cuevas-Tello, J. C., & Rangel-Ramirez, G. (2020). Fruit classification for retail stores using deep learning. In *Pattern Recognition: 12th Mexican Conference, MCPR 2020, Morelia, Mexico, June 24–27, 2020, Proceedings 12* (pp. 3-13). Springer International Publishing.

[18] Ummapure, S. B., &Hanchinal, S. M. (2020). Multi Features based Fruit Classification Using different Classifiers. *Journal of University of Shanghai for Science and Technology*, *22*(12), 1344-1356.

[19] Thinh, N. T., Thong, N. D., & Cong, H. T. (2020). Sorting and Classification of Mangoes based on Artificial Intelligence. *International Journal of Machine Learning and Computing*, *10*(2).

[20] vanGriethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77(21):e104–7.

[21] Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, Ashrafnia S, Bakas S, Beukinga RJ, Boellaard R, et al. The image biomarker standardization initiative: standardized quantitative radiomics for highthroughput image-based phenotyping. Radiology. 2020;295(2):328–38

[22] W. Cai and Z. Wei, "Pii GAN: generative adversarial networks for pluralistic image inpainting," *IEEE Access*, vol. 8, pp. 48451–48463, 2019.