INTELIGENCIA ARTIFICIAL

An Automatic Non-Destructive External and Internal Quality Evaluation of Mango Fruits based on Color and X-ray Imaging using Computer Vision Techniques

Vani Ashok[1], Sheela N[2], Bharathi R K[3]

[1][2]Department of Computer Science and Engineering, JSS S&TU, Mysuru, Karnataka, India

[3]Department of Computer Applications, JSS S&TU, Mysuru, Karnataka, India

[1]vanisj@sjce.ac.in, [2]sheela_cse@sjce.ac.in, [3]rkbharathi@sjce.ac.in

Abstract Quality evaluation of food products, agricultural produce to be specific, has gained momentum from past few decades due to the increased awareness among consumers across the world. This has resulted in the increased emphasis on the development and use of quality assessment techniques in food industry. Moreover, there is a need to automate the quality monitoring of agricultural produce like fruits and vegetables which is otherwise done manually in developing countries hence labor intensive, time consuming and subjective in nature. This paper presents an empirical analysis to build a rapid, robust, real-time, non-destructive computer vision based quality assessment model for mango fruits. The work employs the automatic disease classification of mango fruits based on machine and deep learning models. Firstly, the dataset of colored mango fruits images with 2279 images falling into three classes and another dataset of soft X-ray images of mango fruits with 572 images belonging to two quality classes are developed for detecting external and internal defects, respectively. The multilayer perceptron neural network (MLP NN) with two hidden layers, which may be considered as the starting point for deep learning technique, is proposed as machine learning model to classify the color images of mango fruits into one of three external quality classes with 95.1% accuracy and also to classify the soft X-ray images into two internal quality classes with 97.5% accuracy. In order to step out of feature engineering, actual deep learning convolutional neural network (CNN) models, a customized CNN model and pre-trained CNN models, VGGNet (VGG16) and DenseNet121 were also explored for mango disease classification. The maximum validation accuracy of custom CNN was found to be with 91.52% and 98.7% for color and augmented X-ray images, respectively. The classification accuracy of pre-trained models were found to be reasonably good for the color images but exhibited high variability in results and made it difficult to draw a general conclusion for the proposed datasets. However, the proposed MLP NN model based on few basic intensity and geometric features and also the proposed customized CNN model were found to be the best models and they outperform the state of the art reported in the literature.

Keywords: Quality Assessment, Machine Learning, Multilayer Perceptron, Deep Learning, Convolutional Neural Network.

1 Introduction

Quality evaluation of food products, agricultural produce to be specific, has gained momentum from past few decades due to the increased awareness among consumers across the world [1]. This has resulted in the increased emphasis on the development and use of quality assessment techniques in food industry. The acceptability of fresh horticultural produce by consumers and global market depends on the quality parameters which are broadly classified into external and internal components [2]. Some of the external quality attributes that influence the appearance of fruits are size, shape, color, defects and the internal quality parameters constitute flavor, firmness, crispness, nutrition constituents and defects [3]. The traditional method of quality assessment involves manual inspection of each fruit individually or random sampling from large batches by graders or trained personnel and

classifying it into its appropriate category based on external quality parameters [4]. Also, the older method of assessing the internal quality parameters of fruits is carried out by either puncturing or destroying the fruit.

Mango (Mangifera indica Linn) is the most popular and economically important fruit of many countries including Republic India. In a defect free mango, the shape, color, size, weight, taste, aroma and texture are the major parameters that define both internal and external quality. However, its biochemical constituents such as proteins, vitamins, total soluble solids (TSS), sugar contents, soluble solids content (SSC), total acidity determines other internal quality parameter such as sweetness, fragrance and firmness [5]. The major external attributes of mangoes such as color, weight, shape predict the maturity and ripening phases of fruit and considered to be the important parameters for grading fruits into different quality classes. The mangoes are susceptible to injuries and diseases during its ripening stage and subjected to numerous physical and physiological post-harvest disorders affecting fruit quality [6]. Stem-end-rot, scab, sap burn, spongy stem-end, sooty mold, soft nose, spongy tissue, internal flesh breakdown are some inherent physiological disorders commonly found in mangoes [7]. The mangoes are also affected by numerous pests and insects, some of which are quarantine in nature, at various development stages [8]-[10]. All the internal defects or injuries are not apparently visible on the outside surface of the fruit and can be detected only after cutting open the fruit. These disorders hinder the production of fruits and subsequently results in huge loss to growers. These diseases also affect the quality of fruit and limit the growth of export market of fresh mango due to stringent quarantine regulations followed by leading import countries.

From past two decades, considerable efforts have been put in developing non-invasive systems to monitor and control the quality and safety of specialty crops, mainly fresh fruits, using wide range of sensing technologies [11]. Recent reviews on advanced sensing technologies for non-destructive fruit quality evaluation highlight the use of computer vision, near infrared (NIR) and infrared spectroscopy and various imaging techniques operating at different bands of electromagnetic spectrum like X-ray, computed tomography, magnetic resonance imaging, nuclear magnetic resonance, near infrared, ultrasound and hyper-spectral imaging [12]-[14]. The traditional visual inspection based on external quality parameters, which is time consuming, labor intensive, inconsistent and subjective in nature, is still largely practiced in India and other developing countries for grading and sorting mango fruits into different quality classes. And, majority of the quality parameters that reflect the internal quality are assessed using chemical and biochemical testing, mechanical methods or sonic principles which destroy the samples completely or partially making it unfit for consumption. Most of the recent studies predict the mango quality indicators such as firmness, color, size, weight and physico-chemical parameters separately using various non-destructive techniques like visual spectroscopy, X-ray, NIR, hyperspectral imaging and computer vision [15]-[18].

Even though, the traditional fruit quality assessment method which involves manual inspection is time consuming, subjective and inconsistent in nature, it is largely practiced in developing countries like Republic India. Republic India being the largest producer of world's mangoes has faced import ban on mangoes by leading import countries like USA due to stringent quarantine regulations. The defects mainly caused by quarantine pest infestation have limited the growth of the export market and has caused huge losses to mango growers. Therefore, the proposed work is undertaken to develop methods to non-destructively visualize and evaluate defects in mango fruits. It has two principal themes: The external defect evaluation of fruits using machine learning (ML) and deep learning (DL) techniques with color images and the assessment of internal defects using X-ray imaging techniques. Treating the MLP NN, which is a traditional ML model, with two hidden layers as the starting point of deep learning methods, other DL models such as custom CNN, pre-trained VGG16 and DenseNet121 are also proposed for quality evaluation.

2   Related Work

Computer vision system has been used by the food industry for automated external quality inspection and was found to be a faster, non-destructive, powerful and reliable quality evaluation technique for agricultural products by research community over past few decades [19]. The quality inspection of fruits and vegetables using computer vision and image processing technique include the following steps: image acquisition, pre-processing, extraction of feature descriptors and classification [20]. A detailed review of these steps carried out specifically on mango fruits from past three decades is reported in this section.

Zheng H et al. [21] have proposed least-squares SVM classifier for detecting the browning levels of mangoes based on fractal analysis and L*a*b* values. The classification accuracy was 85.19% and 88.89% for Fractal dimension and L*a*b* values, respectively. A non-destructive computer vision system was developed by Pujari et al. [22] for grading and classification of anthracnose affected mango fruits. The affected areas of fruit was segmented from the health

y region by employing different segmentation algorithms like thresholding, region growing, watershed etc. and the grading was performed by calculating the percentage of affected area. For classification, the textural features extracted from diseased and healthy fruit images using Runlength matrix were used to train the ANN which classified the healthy and defective fruits with 84.65% and 76.6% success rate, respectively. Khoje S and Bodhe S [23] also attempted grading and classification of mango fruits based on size parameters extracted using various size estimation methods like morphological, principal axis and statistical methods. The experimental results indicated that feed forward NN classifier and SVM with polynomial kernel function achieved high classification accuracy of 97% and 95% for size features extracted using statistical method. Sapan Naik [24] has proposed mango grading using four pre-trained CNN models namely Inception v4, Xception, ResNet and MobileNet as feature extractors and support vector machine as classifier to grade the mango fruits into four quality classes with MobileNet achieving highest accuracy of 83.97%. The original dataset with 400 randomly selected RGB images was enlarged to 2432 images using data augmentation. Ashok V et al. [25] have developed an android application for real-time detection of defects in mango fruits using transfer learning approach on pre-trained MobileNetV1 deep CNN model. The model was able to classify more than 95% of images correctly into their respective classes with significant accuracy. Varsha Bhole and Arun Kumar [26] have proposed a deep learning centered non-destructive mango sorting and grading system based on RGB and thermal images. The pre-trained SqueezeNet model was able to grade the mangoes into three quality classes with 93.33% and 92.27% classification accuracy for RGB and thermal images, respectively. The augmented dataset was created with nearly 10000 images for the study.

X-ray and CT imaging techniques, which are predominantly used as medical diagnostics in hospitals, for luggage checking and security purpose at airports and for inspection of industrial components at manufacturing industries, are also being used in food industry to detect defects and contaminants in packaged and processed foods. However, due to its ability to reveal the internal structure of the given sample, X-ray is gaining popularity in recent years for internal quality inspection of fresh agricultural produce [27]. Barcelon E G et al. [28] proposed X-ray absorption technique to measure the quality indicators of mangoes non-destructively by correlating the CT number with physico-chemical properties such as density, acidity, moisture content, total soluble sugars (TSS) and pH value of mango fruit using regression equation. The experimental results indicated positive correlations between the CT numbers, moisture content, density and acidity, whereas TSS and pH values were found to be inversely correlated to CT measurements. The jelly seed disorder in various mango varieties has been detected by Sharma R R et al. [29],[30] using X-ray imaging technique and the same has been proposed to be included in supply chain to enhance consumer acceptance. The resultant X-ray images exhibited dark gray areas for jelly seed disorder and complete light gray region for healthy fruits. A non- destructive method for detecting spoilage in mango seeds based on VGG16 deep CNN model has been proposed by Ansah F A et al. on an augmented dataset of 8000 green mango fruits X-ray images. The model was able to identify bad seed from good seed with an accuracy rate of 97.66% [31].

From the earlier works related to non-destructive quality evaluation of various fruits and specific to mango fruits, it is found that, compared to other popular fruits like apples, oranges, watermelons and not much research progress has been made on mangoes which are the most economically important fruit of South Asia. The potential of *computer vision* and *X-ray imaging* techniques to investigate the external and internal defects or diseases of mango fruits have not been explored to the maximum.

Moreover, the popularity of deep learning in computer vision field has given way for the design of systems that provides end-to-end solutions for image classification problems with huge feature space and considerable number of classes. Even though traditional machine learning classifiers achieve significant classification accuracy within reasonable computational time, these classification methods suffer from the following limitations:

- Great efforts go into engineering features. Given an application, the bottleneck is creating and identifying better features. The whole basis of feature engineering is to identify feature vectors in the direction of achieving a good classification accuracy which is done on trial and error basis.
- To find a robust and global feature extractor method that summarizes and represents all the defect patterns in a single vector is very difficult.
- Regardless of the type of classifier chosen, the accuracy cannot go beyond some threshold and it is impossible to determine which classifier is better than the other for different combination of features.

In this context, the present work proposes deep learning methods to automate the non-destructive quality evaluation of mango fruits. However, since MLP NN, a traditional ML technique is considered as the starting point for deep learning, the performance of the MLP NN model with 2 hidden layers with various activation functions on manually extracted basic geometric and intensity features are proposed for defects detection. In addition, a custom CNN model and off-the-shelf pre-trained models, VGG16 and DenseNet121, all of which use representation learning with deep learning to extract features for classification are also explored for defects detection. The contribution of the proposed work is threefold: First, developing datasets of coloured and soft   X-ray images of

Mango fruits. Second, conducting empirical and comparative evaluation of machine learning and deep learning defects classification models based on the external quality of mango fruits. Third, carrying out similar tasks based on the internal quality of the mango fruits using soft X-ray images.

3   Materials and Methods

The proposed model for automated fruits disease identification and classification which is carried out on *Alphonso* mango cultivar is as shown in Figure 1.
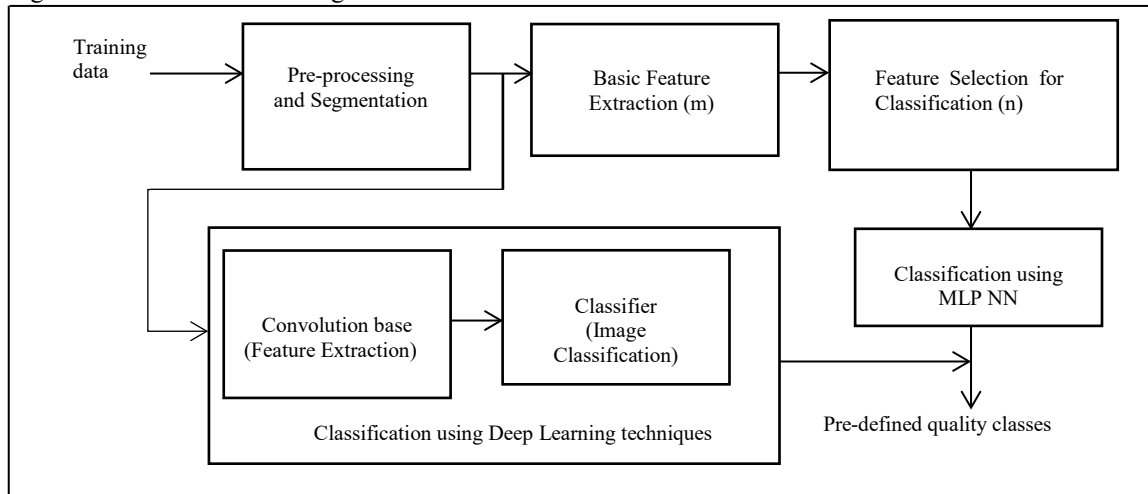

Figure 1: The framework of automated defect classification using MLP NN and DL models.

Only MLP NN is proposed as the machine learning model as it can be considered as the starting point for deep learning. Moreover, since the main objective of the work is to propose deep learning techniques, only very few basic geometric and intensity features are extracted in Gray, R, G, B, H, S and V channels of color images and only Gray channel is considered for X-ray images for the proposed MLP NN architectures.

The present work is motivated by latest works carried out in computer vision applications using the emerging deep learning and transfer learning techniques. At the abstract level, the deep learning approach using CNN consists of two parts: convolution base and classifier as shown in Figure 1. Convolution base consists of stack of convolutional (CONV) layers followed by pooling layers. The CONV layers with multiple convolution kernels act as feature extractors during training process and learn the feature representations of the input image by extracting features and generating feature maps at each layer (the initial layers might learn from low-level, generic features like edges, curves, and the next layers might learn from shape, colors and last layers learn more abstract, specific features) [32]. Convolution base reduces number of parameters by sharing patterns and it supports distributed representation. The pooling layers reduce the size of the input, time of computation and make common features more robust. So, convolution base gradually extracts higher level feature representations and learns hierarchical feature representations [33, 34]. The classifier contains stack of fully connected (FC) layers and an output layer. The FC layer performs high-level reasoning by building strong features with strong capabilities from the distributed representations of the convolution base. It generates global semantic information based on which the output layer performs classification using softmax operator or SVM classifier [35].

In transfer learning a CNN model which has already been trained on a similar task is reused for the current task. In the proposed study we consider the knowledge transfer from natural images to mango fruits images using powerful pre-trained CNN models. These off-the-shelf CNNs have already been trained on large benchmark datasets, ImageNet, and we reuse the feature extraction capabilities of these to train a new classification layer on top. Since our dataset size is moderate, we choose to freeze all the layers of the pre-trained model except the last FC layer, run the model as feature extractor and project the data into a new feature space through the propagation of the mango fruits training dataset into the CNN. The resultant feature vector set from the last layer which is the new representation of our fruit images is then used to train the classifier (output layer) with softmax function.

3.1   Datasets

The proposed model for automated fruits disease identification and classification is carried out on *Alphonso* mango cultivar as it is grown on large scale in southern part of Republic India. Besides, it is one of the popular and leading export mango cultivar. The mango dataset containing color images is prepared by collecting Alphonso variety mango fruits from an orchard in two batches of harvests. Before the experimentation, the farmers, who are the natural graders and fruit experts from Department of Horticulture were consulted to visually inspect these fruits and to grade them manually. Based on appearance, surface defects and Codex Standard [36] for mangoes, fruits were graded into two quality classes: non-defective and defective. According to Codex Standard, "Extra class", which contains all the clean, superior quality mango fruits free from latex stains, bruises, blemishes, insect damage, defects, uniform size and maturity level, are considered as non-defective class. And, the remaining fruits with visible defects due to insects, skin blemishes due to sunburn, sunscald, skin breaks, scab, cracks, latex stains, fungal infection, stem end rot, harvest wounds, overripe, bruises or punctures are considered to be in defective class. A set of healthy mangoes were selected to induce chilling injury manually, thereby generating chilling injured class. In order to induce chilling injuries manually, some of the healthy mangoes were stored at 4°C to 5°C for 3 to 4 weeks and then exposed to normal temperature for one week. Depending on the cultivar, temperature and extent of exposure, the symptoms of chilling injury include skin discoloration, uneven ripening, pitting, poor taste, reduced aroma and flavor, increased vulnerability to decay and pulp, and browning in severe cases. The sample images of three classes are shown in Figure 2. In the proposed work, the images are acquired using Sony Cyber-shot DSC-WX7 digital camera. A total of 2279 color images are acquired with a resolution of 640x480 pixels and 60.5KB size each. The image acquisition is done between 12 P.M to 3 P.M under natural light. Camera is positioned so as to avoid the formation of fruit shadow and a uniform white background is used to ease the segmentation process.

After further visual inspection, the samples are labelled and the class of each sample is determined. The quality of fruits is categorized into one of three classes: *chilling injury*, *defective* and *non-defective*. Among 2279 images, 736 images are labelled as chilling injured mangoes (Class 1), 632 images are labelled as defective mangoes (Class 2) and 911 images are included under non-defective mangoes (Class 3).

Classes	Samples of postharvest "Alphonso" mango fruits with and without diseases
Chilling injury	
Defective	
Non-defective	

Figure 2: Sample images of Mango fruits in color image dataset.

The second dataset with 572 soft X-ray images of *Alphonso* mango fruits was prepared for the internal quality evaluation. The mangoes were collected from two batches of harvests. The fruits were cleaned from dirt and were imaged using fixed digital radiography machine from a healthcare diagnostic centre. The technical details of the X-ray machine used in the proposed study are specified in Table 1.

Table 1: Technical details of X-ray machine.

X-ray Generator	Anode voltage range in kVp	0-150 with 1 kVp step
	Power input	Stabilized single phase 250VAC, 50Hz
	Exposure time range	0.02 millisecond -5 seconds
	X-ray tube current range in mA	0.5-5 with 0.1mA step
	Penetration	Suitable for agriculture produce
Image Generating System	Scan area	170 mm X 140 mm
	Image resolution in pixels	2043 X 1710
	Image capture method	Manual trigger synchronized with the X-ray generator
	Image capture software	Image suite

The X-ray can be 'hard X-ray' with shorter wavelength and energy above 12keV or longer wavelength 'soft X-ray' with low energy generated with tube voltage less than 100kV [37]. Soft X-rays which are used in food processing and packaging industries for food irradiation and also to analyse the internal density changes of fleshy fruits was generated by setting the output voltage, composite factor and exposure time of the X-ray source to 45-50kV, 25mA and 0.04ms, respectively. The images acquired at these optimum output conditions revealed internal injuries of the mangoes with superior contrast and less noise.

After obtaining the X-ray images of 572 mango fruits, fruits were cut open and in consultation with farmers and experts from Department of Horticulture, the images were classified into two classes: *Defective* and *Non-defective*. All the mangoes with internal diseases such as anthracnose, softnose, jelly seed, spongy tissue etc. are labelled as defective fruits and the remaining healthy mangoes are labelled as non-defective fruits. Accordingly, of 572 to X-ray images, 277 images were labelled as "defective" and 295 images as "non-defective".

3.2   Pre-processing and Image Segmentation

In order to reduce the computational complexity, in the pre-processing stage, the color images were cropped and resized to a dimension of 200x150 pixels with 5.58KB size. And, the initial X-ray images were rather dark, low in contrast and noisy. Image enhancement in terms of increase in brightness, contrast and reduction of noise is must for X-ray images in order to extract useful textural information and perform best interpretation. In the proposed study, Contrast limited adaptive histogram equalization (CLAHE) enhancement technique was examined to increase the quality of the X-ray image. Unlike other contrast enhancement methods, CLAHE works on small regions of the image instead of the entire image. The input image is divided into smaller tiles, the contrast of which is enhanced in such a way that the histogram of every output region closely matches with the uniform distribution by default. Since the contrast enhancement can be limited in this method, it avoids amplification of noise that is normally present in input X-ray images [38].

The sample X-ray images for defective and non-defective mango fruits are shown in Figure 3. The images of original fruits from which X-ray images were acquired are shown in the first row. The first three images are defective fruits which seem to be healthy and free from defects and the next three images are non-defective fruits. The original X-ray images of the corresponding fruits are shown in next row. After acquiring X-ray images, the fruits were cut open and images were labelled. In the next step, the quality of these images was enhanced using CLAHE technique. The sample images of these steps are shown in subsequent rows of the figure.

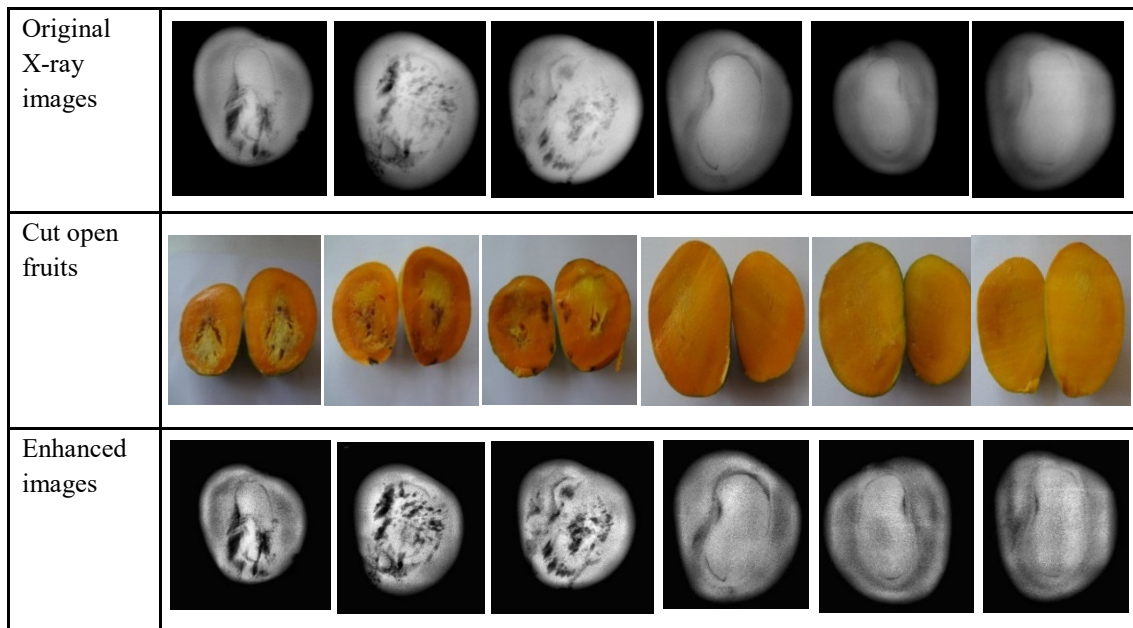|	Sample Images of Defective and Non-defective Fruits
Whole fruits	

Figure 3: Sample defective and non-defective mango fruits X-ray images.

Segmentation using histogram based global thresholding [39] was performed on the pre-processed color and X-ray images. Without segmentation, the learning models treat the input as rectangular image, extract the information from the insignificant background along with the features from foreground i.e., fruit part, thereby increasing the computational time of the model.

An initial image is computed by thresholding *Cb* channel of *YCbCr* information of the color image (Figure 4a) and Gray channel of X-ray image, with the estimation of a global threshold using statistical approach. A high contrast gray value image, *I*, with a bimodal histogram, as shown in Figure 4b, where the left distribution corresponds to the background pixels and the right to the pixels of fruit image is obtained. From this image, a binary image *J* (Figure 4c) is obtained by creating a separation between foreground and background estimating global threshold T as given in Equation 1.

$$J(i,j) = \begin{cases} 1 & \text{if } I(i,j) > T \\ 0 & \text{else} \end{cases} \qquad (1)$$

where '1' (value 255) is assigned for all pixels in foreground i.e., fruit part and '0' (value 0) is assigned to all pixels of background. Finally, the binary image is mapped to RGB color space to obtain the segmented image as shown in Figure 4d.

(a)   (b)   (c)   (d)
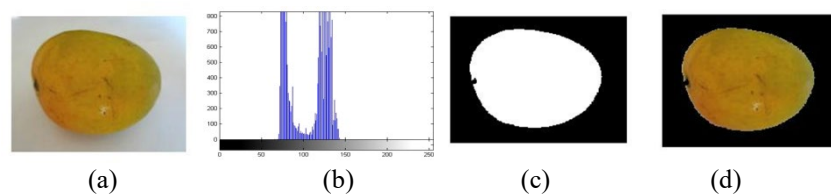
Figure 4: Segmentation of fruit image (a) Original image (b) Histogram (c) Binary image (d) Segmented image.

Thus, by setting the background intensity levels to '0' and retaining the original intensity levels of the foreground, the model is allowed to interpret the details of high-level visual information of the fruit part. Moreover, without segmentation the extraction of geometrical information has no sense, since all shapes will be treated as rectangles of the entire images.

3.3   Classifiers

In the proposed framework while using traditional machine learning algorithm, since the idea is to obtain the highest accuracy defined as the proportion of true results and also to choose the model similar to DL model, artificial neural network with various activation functions and different architectures of Multilayer Perceptron are designed for the mangoes disease classification.

The experimentations were also carried out with CNNs trained from scratch (customized CNN) and off-the-shelf CNNs (pre-trained) using them as fruit imaging feature extractor. In case of customized CNN, the convolutional base which consists of stack of convolutional and pooling layers generates features adapted to the original dataset during CNN training and in case of pre-trained models the patterns that have been learnt on a large benchmark dataset for a particular application is used for defect detection and classification in mango fruits. In particular we explore the following three CNN architectures: Custom CNN model which is trained from scratch; two different pre-trained architectures, VGGNet (VGG16) and DenseNet (DenseNet121) which have been trained on large benchmark ImageNet dataset with 1000 object classes.

1. *Multilayer Perceptron*: It is a simple, commonly used, feed forward, supervised learning predictive model having structural flexibility and good representational capabilities [40]. It may have one or two fully connected hidden layers where every node in a layer is connected to all the nodes of the next layer. Each node in a given layer computes biased weighted sum of their inputs and pass this activation level via transfer function to generate their outputs. In perceptron learning, learning the classifier is learning weights. Here, features from samples are fed to the model to find output. Initially, each input is assigned with some random weights and a bias. The model takes input one-by-one and performs feed forward pass. At the output, non-linearity is introduced by the activation function and the difference between the actual and the desired outputs (cost/loss function) is found. Then the weights are changed such that the cost function is minimized. This is done by using a back-propagation algorithm that computes the gradient for each sample.

2. *Custom CNN model*: A custom CNN model is proposed for defect classification in mango fruits. The sequence is formed by three sets of two convolutional layers and a pooling layer, followed by two FC layers. The input image size is 64x64x3. The first FC layer with 128 units is added to the last layer with softmax function for three pre-defined object classes (chilling injury, defective and non-defective) of mangoes. The outputs of all convolutional layers are fed to Rectified linear unit (ReLU) rectifier function which adds non-linearity.

3. *VGG16*: It is a variant of VGGNet model which consists of 16 CONV layers stacked on top of each other with very small 3x3 filters and five max pooling layers that reduce the number of connection between convolutional layers. The two FC layers with 4096 units are followed by softmax classifier with 1,000 target classes corresponding to ILSVRC classification. The size of input image is 224x224 and all the layers are followed by ReLU activation function to introduce non-linearity to the model and prevent overfitting [41].

4. *DenseNet121*: In case of deeper CNNs like VGGNet with 19 layers, AlexNet and ResNets surpassing 100 to 1000 layers, increasing the depth of the networks results in problem of large forward information and vanishing gradient [42]. These issues are addressed in Densely connected CNNs, DenseNets, supporting narrow layers which add small set of new features keeping the remaining feature maps unaffected [43]. Every layer has access to its preceding feature maps, the gradients and the original image. DenseNets concatenate the output feature maps of a current layer to the incoming feature maps instead of adding their value. This efficient use of parameters also makes DenseNet less prone to overfit. Among other variants of DenseNets designed over ImageNet dataset, DenseNet121 is the simplest model with 4 DenseBlocks and 3 transitional layers [44]. The input image size is 224x224 and the size of the output vector of the linear layer is 1024x1.

3.4   Experimental Setup

3.4.1   Multilayer Perceptron NN

Here, the performance of MLP architecture is evaluated for two hidden layers topology, using different activation functions in hidden and output layer and using two optimization algorithms, Scaled conjugate descent (SCG) and Gradient descent (GD) to minimize the training errors during back-propagation. Performances of these algorithms are evaluated separately for different architectures. The activation functions applied at the hidden layer are

hyperbolic tangent, sigmoid and those at the output layer are identity, softmax, hyperbolic tangent and sigmoid. These functions introduce non-linearity without which multiple layers of neural network are equivalent to a single layer.

GD is an important optimization technique which computes the gradient, the derivative of the error function with respect to the weight matrix W [45]. Gradient or change in weight W is a vector in weight space and when the training moves in that direction loss function increases. Since gradient indicates the direction where loss increases, updating weights is to be done in the reverse direction i.e., every iteration, the weights are updated by subtracting the gradient from weight by a small step called as 'learning rate'. Finally, the algorithm halts in a low point, which may be a local or global minimum. And, SCG is a variant of linear search based conjugate gradient algorithms and is a combination of model-trust region based Levenberg-Marquardt algorithm and conjugate gradient approach [46]. The performance of SCG is independent of hand woven parameters. It converges quickly to the solution as it uses stepwise optimization instead of time consuming line search in every iteration.

Batch-wise training which is normally used if the size of training dataset is small, is used to train the model with SCG optimization algorithm. The network updates the weights only after passing through all training samples. The weights are updated many times until there is minimum or no decrease in error between two successive steps. The MLP model with GD optimization algorithm is trained for both batch and mini-batch. If batch type training updates the network parameters after evaluating all training samples per epoch, mini-batch updates the weights after evaluating a subset of training samples per epoch [47]. As error functions, sum-of-squares is used if the activation functions of the output layer are identity, sigmoid or hyperbolic tangent and cross entropy is used for softmax function.

The original dataset of color images and X-ray images are divided into train and test datasets in the ratio of 80:20 and then batch and mini-batch training is performed to train the network on training or subset of training dataset in every epoch. After training, the performance accuracy of the model is evaluated on unseen samples of test dataset.

Feature Extraction: Basic intensity features from Gray (g), red, green, blue (RGB) and hue, saturation and intensity (HSV) color channels and basic geometric features were extracted from color images. But, for X-ray images basic intensity features from only Gray channel and also the basic geometric features were extracted. The descriptions of the extracted features are given below:

- Basic intensity features: The 6 standard intensity features extracted from each color space of the segmented color image and Gray color space in X-ray image, give the overall image intensity information. They include mean (μ), standard deviation (σ), kurtosis (k), skewness of the intensity in the image (s), mean first order derivative in the boundary of the region ($f'(x_{ij})$) and mean Laplacian second order derivative in the region ($f''(x_{ij})$) [48]. They are given in Equation 2 through Equation 7.

$$Mean, \mu = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{2}$$

$$Standard\ deviation, \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij-}x')^2} \tag{3}$$

$$Kurtosis, k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{ij-}x')^2}{(\frac{1}{n}\sum_{i=1}^{n}(x_{ij-}x')^2)^2} \tag{4}$$

$$Skewness, s = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{ij-}x')^3}{(\frac{1}{n}\sum_{i=1}^{n}(x_{ij-}x')^2)^3} \tag{5}$$

$$Mean\ boundary\ gradient, \quad f'(x_{ij}) = \frac{x_{ij+1} - x_{ij}}{(ij+1) - ij} \tag{6}$$

$$Mean\ Laplacian, \quad f''(x_{ij}) = \frac{f'(x_{ij}) - f'(x_{ij-1})}{(ij+1) - ij} \tag{7}$$

where pixel x_{ij} is evaluated against n number of pixels in the segmented image.

- Geometric features: 18 basic features consist of center of gravity, height, width, area, perimeter, roundness, Danielsson factor, Euler number, equivalent diameter, major and minor axis lengths, orientation, solidity,

extent, eccentricity, convex area, filled area were extracted from each image sample. These features on the location, shape and size have been extracted from the segmented fruit image. The location and size features are given in terms of pixels whereas shape features are coefficients without units.

The summary of the extracted features for color and X-ray images with elapsed time (in seconds) is specified in Table 2. This feature exploration is done offline and it does not affect the computational time of the classifier since at disease detection time only feature selection and classification processes constitute for the computational time.

Table 2: Summary of extracted features.

SI. No.	Type of Dataset	Geometric Features	Intensity Features	Color Channels	No. Extracted features	Elapsed time in seconds
1	Color images	center of gravity i [px], center of gravity j [px], Height [px], Width [px], Area [px], Perimeter [px], Equivalent Diameter [px], Major Axis Length [px], Minor Axis Length [px], Convex Area [px], Filled Area [px], Orientation, Roundness, Danielsson factor, Euler Number, Solidity, Extent, Eccentricity	Intensity Mean, Standard deviation, Kurtosis, Skewness, Mean Laplacian, Mean boundary gradient	g,R,G,B,H,S,V	60	130
2	X-ray images			Gray	24	60

Feature Selection: In this study, the sequential forward selection algorithm with Fisher score objective function (SFS-Fisher) is used for selecting the best features [49]. Feature selection algorithms normally involve increasing or diminishing the objective function for a feature which will be used as feedback to select the next best feature. The objective of SFS is to find the best single feature that has maximum interclass variation and minimum intra-class variance and then to augment one feature at a time which maximizes the objective function of classification in combination with the previously chosen features. This iterative process continues until there is considerable improvement in the objective function by the addition of a new feature. By doing so, all the irrelevant features, such as highly correlated features and features with constant values, are eliminated. The objective function of SFS-Fisher is given in Equation 8.

$$J = \max\left(\frac{S_b}{S_w}\right) \tag{8}$$

where *max* returns the feature with the inclusion of which into the feature subset (Y_n) maximizes the objective function $J(Y_n)$. S_b is the interclass scatter and S_w depends on the intra-class covariance i.e., the correlation between a feature and class labels. The Fisher discriminant simultaneously ensures a larger S_b and a smaller S_w.

The extracted features from each dataset mentioned in Table 2 were reduced to maximum of 20 features, in order to avoid overfitting of the classifier models. The details of the best selected features for color and X-ray images which act as predictor variables for MLP NN classifier forming the input layer are given in Table 3. In the present work, original dataset is divided into training and test datasets in the ratio of 80:20 and then batch and mini-batch training is performed to train the network on training or subset of training dataset in every epoch. After training, the performance accuracy of the model is evaluated on unseen samples of test dataset.

Table 3: Details of the selected features for MLP NN classifier.

Features for Color image Dataset		Features for X-ray image Dataset	
Intensity Features	g-Mean Boundary Gradient	Intensity Features	g-Intensity StdDev
	S-Intensity StdDev		g-Intensity Skewness
	B-Intensity StdDev		g-Intensity Mean
	G-Intensity Mean		g-Mean Laplacian
	G-Intensity StdDev	Geometric Features	Eccentricity
	G-Intensity Skewness		center of grav j [px]
	H-Intensity Mean		Height [px]
	R-Intensity StdDev		Perimeter [px]
	g-Intensity StdDev		Roundness
	S-Intensity Kurtosis		Width [px]
	H-Mean Boundary Gradient		Euler Number
	H-Intensity Skewness		Solidity
	g-Intensity Mean		Danielsson factor
	H-Intensity Kurtosis		
	H-Intensity StdDev		
	S-Intensity Skewness		
	B-Intensity Kurtosis		
	S-Intensity Mean		
Geometric Features	Eccentricity		

3.4.2 Custom CNN Model

The proposed architecture of custom CNN model is given in Table 4. It contains three sets of layers, each having two CONV layers followed by max pool layer. The number of filters is 64, 128 and 256 in first, second and third set of CONV layers respectively. The filter size is 3x3, padding is 1 and stride is 1 in all CONV layers. Maxpool layers decrease the input size by a factor of 2. Finally, a FC layer with 128 units and another FC layer with softmax classifier for three object classes are added to the model. For every epoch the model learns 1,670,211 parameters. Initially, the performance of the proposed custom CNN was explored without adding dropout layers for 50 epochs. The train accuracy obtained was 96.3% and the validation accuracy was 90.12%. The test loss increased at the end of training (between 40-50 epochs) and there was clear deviation in the plots of train and test accuracy and the model loss. So we added dropout layers after every set of layers and also after first FC layer as given in Table 4. Dropout layers with dropout rate of 0.2, 0.3, 0.4 and 0.4 are added respectively and then the model was trained for 50 and 100 epochs. With the input image dimension of 64x64x3, the samples were divided into train, test and validation set in the ratio of 70:15:15.

Table 4: Architecture model of Custom CNN.

Layer (Type)	Description	Output shape	# parameters
Input layer	64x64	(64, 64, 3)	0
Conv	64;3x3;p=1,st=1	(62, 62, 64)	1792
ReLU		(62, 62, 64)	0
Conv	64;3x3;p=1,st=1	(60, 60, 64)	36928
ReLU		(60, 60, 64)	0
Maxpool	2x2,st=1	(30, 30, 64)	0
Dropout	0.2	(30, 30, 64)	0
Conv	128;3x3;p=1,st=1	(28, 28, 128)	73856
ReLU		(28, 28, 128)	0
Conv	128;3x3;p=1,st=1	(26, 26, 128)	147584
ReLU		(26, 26, 128)	0

Maxpool	2x2,st=1	(13, 13, 128)	0
Dropout	0.3	(13, 13, 128)	0
Conv	256;3x3;p=1,st=1	(11, 11, 256)	295168
ReLU		(11, 11, 256)	0
Conv	256;3x3;p=1,st=1	(9, 9, 256)	590080
ReLU		(9, 9, 256)	0
Maxpool	2x2,st=1	(4, 4, 256)	0
Dropout	0.4	(4, 4, 256)	0
Flatten		4096	0
FC	(4096+1) x 128	128	524416
ReLU		128	0
Dropout	0.4	128	0
FC	(128+1) x 3	3	387
Softmax classifier		3	0
Total trainable parameters			1,670,211

3.4.3   Pre-trained CNNs

In the next experimentation, VGG16 CNN which has 16 weighted layers stacked on one above the other with increasing depth and pre-trained on ImageNet dataset for 1000 classes is considered for defect classification of mango fruits. The pre-trained weights from the ImageNet were used for training the proposed model. All the layers were frozen except the last fully connected layer which is fine tuned to our application with three classes and softmax activation function i.e., we are using the model as starting point and re-training the classification backend of the model for our specific problem. A new input layer is also added as the model should fit on new data of our specific problem. Hence with frozen weights from the convolution base, pre-trained weights were used to get important feature activations as bottleneck features into fine-tuned FC network, and now this new FC network, trained on mango dataset, gives us the required inference as per training. The architecture model of the fine- tuned VGG16 model for color image dataset with 3 classes is given in Table 5. The images in dataset are resized to 224x224 and normalized so that the pixel intensities are within [0, 1] range.

Table 5: Architecture model of VGG16 model.

Layer (Type)	Output shape	# parameters
Input layer	(224,224, 3)	0
Block1_CONV1	(224,224, 64)	1792
Block1_CONV2	(224,224, 64)	36928
Block1_pool (Maxpool)	(112,112, 64)	0
Block2_CONV1	(112,112, 128)	73856
Block2_CONV2	(112,112, 128)	147584
Block2_pool (Maxpool)	(56,56, 128)	0
Block3_CONV1	(56,56, 256)	295168
Block3_CONV2	(56,56, 256)	590080
Block3_CONV3	(56,56, 256)	590080
Block3_pool (Maxpool)	(28,28, 256)	0
Block4_CONV1	(28,28, 512)	1180160
Block4_CONV2	(28,28, 512)	2359808
Block4_CONV3	(28,28, 512)	2359808
Block4_pool (Maxpool)	(14,14, 512)	0
Block5_CONV1	(14,14, 512)	2359808
Block5_CONV2	(14,14, 512)	2359808
Block5_CONV3	(14,14, 512)	2359808
Block5_pool (Maxpool)	(7,7,512)	0
Flatten	(25088)	0
FC1	(4096)	102764544
FC2	(4096)	16781312
FC3	(3)	12291

Softmax classifier	(3)	0
Total parameters		134,272,835
Trainable parameters		134,272,835

Finally, the DenseNet121 architecture, a next step on the way to keep increasing the depth of deep CNN, having blocks of CONV and pooling layers, DenseBlocks, is explored for our application. This model is pre-trained on ImageNet dataset and keeping all the layers in the convolutional base frozen we only learn a classifier on top of it. The DenseNet121 architecture for color image dataset with 3 classes is shown in Table 6. While fine tuning, the last layer of the original architecture is chopped off. A global average pooling layer followed by a dropout layer with 0.3 dropout rate and finally a FC layer with softmax classifier for the proposed three object classes are added. The input layer is replaced with samples from mango fruits dataset which are resized to 221x221 image size and normalized such that the pixel intensities are in [0, 1] range instead of [0, 255].

Table 6: Architecture model of DenseNet121.

Layer (Type)	Output shape	# parameters
Input layer	(221, 221, 3)	0
Densenet121 (Model)	(7, 7,1024)	7037504
Global_average_pooling2D	(1024)	0
Dropout	(1024)	0
FC	(3)	3075
Softmax classifier	(3)	0
Total parameters		7,040,579
Trainable parameters		6,956,931
Non-trainable parameters		83,648

The performance of the CNN models is analyzed in terms of classification accuracy and model loss. The model accuracy graph, a plot of train and test accuracy over the training epochs and model loss graph which is the plot of train and test loss over training epochs are obtained. The learning curves of these graphs are used as diagnostic tools to analyze if the model has over learned, under learned, and whether the train and test datasets are suitable representatives. The learning curves obtained from the training dataset indicate how well the model is learning whereas the learning curves calculated from test dataset determines how well the model works on unseen images i.e., they depict the generalization behavior of the network. All the models were trained on Intel Xeon® CPU with NVIDIA Quadro K600 GPU. The computational complexity of training deep CNN models was estimated by measuring training time which indicates time required to learn the class [50].

4   Results and Discussion

4.1   Multilayer Perceptron NN

The performance details of MLP NN classifier using SCG and GD optimization algorithms with two hidden layers for the selected best features, 19 and 13 for color and X-ray images, respectively are enumerated in Table 7. The combination of activation functions at the hidden and output layers that result in terms of maximum performance of the MLP NN model was also identified. From the results it can be deduced that for color images, maximum disease classification accuracy of 95.1% was achieved by the MLP topology with SCG optimization algorithm and two hidden layers, using tangent hyperbolic and sigmoid functions at hidden layers and output layer, respectively. For SCG, the optimum number of nodes at the hidden layers were set on trial basis in order to get maximum classification accuracy. The training parameters, initial λ (lambda) which indicates the change in weight is set to 0.0000005, initial σ (sigma) which is a regulating parameter is set to 0.00005 and minimum relative change in training error is set to 0.0001 for batch-wise training. The training of network stops when there is there is no decrease in error in one consecutive step(s).

Table 7: Performance details of MLP NN classifier with SCG and GD optimization algorithms.

Activation fn. (hidden layers, output layer)	Test Accuracy in % for MLP with SCG (batch-wise training)		Test Accuracy in % for MLP with GD (batch-wise training)		Test Accuracy in % for MLP with GD (mini-batch)	
	Color image Dataset	X-ray image Dataset	Color image Dataset	X-ray image Dataset	Color image Dataset	X-ray image Dataset
Sigmoid, Identity	94.5	92.5	91.7	91.2	91	92
Sigmoid, Softmax	93.5	94.8	89.7	93.3	93.2	**97.5**
Sigmoid, Hyperbolic Tangent	92.5	90.8	91.4	92.6	92.1	94.3
Sigmoid, Sigmoid	91.9	93.4	**93.1**	89	**93.9**	88.6
Hyperbolic Tangent, Identity	90.8	92.3	92.9	91.4	92.4	94.2
Hyperbolic Tangent, Softmax	93.3	94	92.8	**95.8**	89.4	91
Hyperbolic Tangent, Hyperbolic Tangent	91.5	**95.3**	92.3	93.3	93.3	93.7
Hyperbolic Tangent, Sigmoid	**95.1**	93.5%	92.1	90.6	92.1	94.9

The classification results of the MLP NN model which give number and percentage of total samples that were classified correctly are given in Table 8. Sum-of-squares error function is chosen as the activation function at the output layer is sigmoid. Since the output layer is categorical in nature, relative error is given in terms of average percentage of incorrect predictions and it is 7% for training set and 4.9% for testing set. It is observed that the test accuracy of chilling injury and non-defective classes are significantly high, 98.6% and 95.9%, respectively and is relatively low for defective class. The overall training accuracy is 93.0% and testing accuracy is 95.1%. The training time of the model is 0.44 seconds.

Table 8: Classification results of MLP NN for color images with SCG optimization algorithm and maximum accuracy of 95.1%.

Sample	Observed	Predicted			
		Chilling injury	Defective	Non-defective	Percent Correct
Training	Chilling injury	570	22	4	95.6%
	Defective	23	445	56	84.9%
	Non-defective	3	20	691	96.8%
	Overall Percent	32.5%	26.6%	40.9%	93.0%
Testing	Chilling injury	138	2	0	98.6%
	Defective	3	96	9	88.9%
	Non-defective	3	5	189	95.9%
	Overall Percent	32.4%	23.1%	44.5%	95.1%

Similarly, for X-ray images the maximum classification accuracy achieved by the MLP NN model with both SCG and GD optimization algorithms was found to be above 95%. However, GD optimization algorithm with mini-batch training achieved the highest accuracy of 97.5% for 13 selected features forming the input layer as predictor variables. The activation functions are sigmoid for hidden layers and softmax with cross entropy error function at the output layers. In mini-batch training, the training dataset is divided into batches of approximately equal sizes and the model updates the weights after passing through each batch. The updating of weights continues until the stopping rule, negligent or zero error improvement between two successive epochs, is met. For GD optimization algorithm with mini-batch training, initial learning rate is set to 0.4 and this learning rate gets smaller and smaller as the learning progresses. It is reduced for every 10 epochs until a lower boundary of 0.001 is reached. The learning in a fixed direction is encouraged by including momentum parameter set to 0.9.

The disease classification results of MLP NN for X-ray images are given in Table 9. It is observed that test accuracy for defective class and non-defective class is 100% is 95.4% respectively with 6.4% incorrect predictions for training set and even lower 2.5% incorrect predictions for testing set. The overall training accuracy is 93.6% and testing accuracy is 97.5%. The cross entropy error function is used as the activation function of the output layer is softmax. The training time of the model is 0.28 seconds.

Table 9: Classification results of MLP NN with for X-ray images with GD (mini-batch) optimization algorithm and maximum accuracy of 97.5%.

Sample	Observed	Predicted		
		Defective	Non-defective	Percent Correct
Training	Defective	206	17	92.4%
	Non-defective	12	218	94.8%
	Overall Percent	48.1%	51.9%	93.6%
Testing	Defective	54	0	100.0%
	Non-defective	3	62	95.4%
	Overall Percent	47.9%	52.1%	97.5%

The network model using GD mini-batch training optimization algorithm for classification of X-ray images is shown in Figure 5. It has 13 selected features as predictor values at the input layer, two hidden layers with 8 and 6 neurons in first and second layer, respectively and two categorical classes at the output layer.
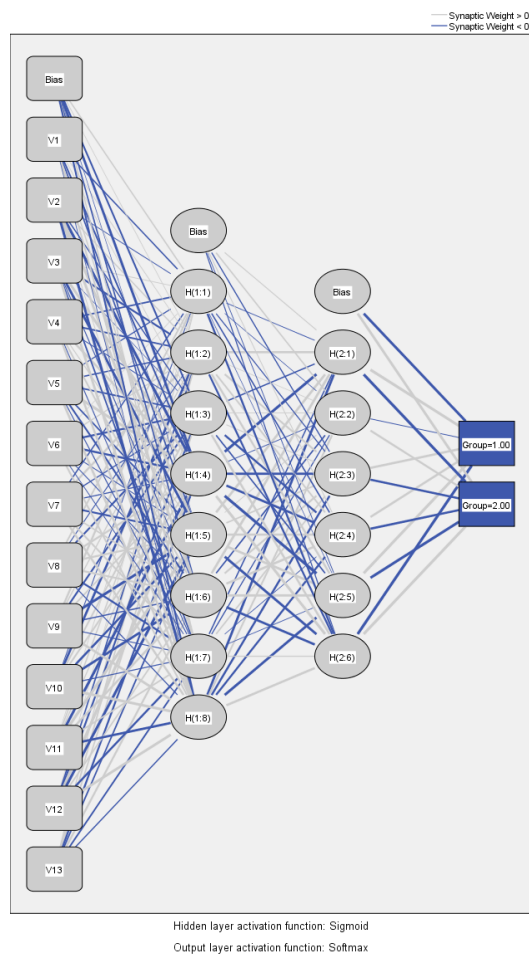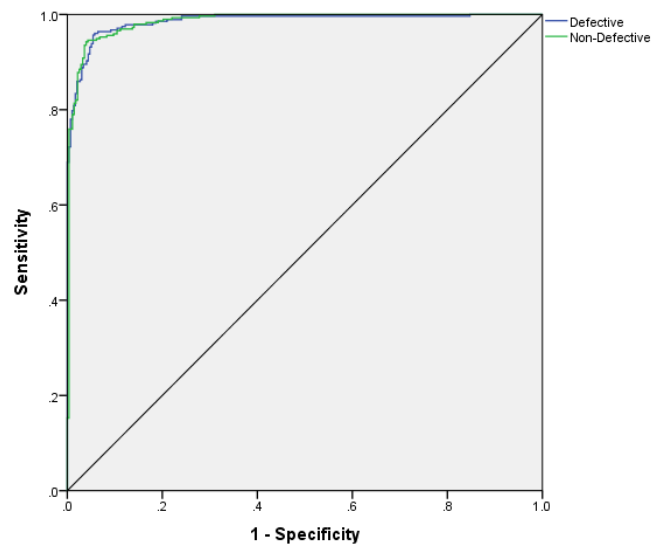

Figure 5: Network Model.

Figure 6: ROC Curve.

The classification performance is also assessed by computing the area A_z for each categorical class under Receiver operating characteristic (ROC) curve as shown in Figure 6. ROC is a plot of 'sensitivity' (S_n) against '1-specificity' ($1-S_p$). Sensitivity is the ratio of true positives (TP) to the sum of true positives (TP) and false negatives (FN) and 1-specificity is the ratio of false positives (FP) to sum of true negatives (TN) and false positives (FP). For, defective class, TP is correctly classified samples of defective class, TN is correctly classified samples of other class (non-defective), FP is the samples of non-defective classes which are classified as defective and FN is the samples

of defective class which are classified as other class. Similarly, S_n and $1\text{-}S_p$ are computed for non-defective class. Since the dependent variable has two categories in X-ray images based study, ROC chart displays one curve for each category and each curve treats the category at issue as the positive state versus the aggregate of all other categories. Ideally, for each curve, $S_n = 1$ and $1\text{-}S_p = 0$ indicates that all samples belonging to that class are classified correctly without flagging false alarms.

The area under each curve (A_z), a performance metric which indicates how reliably the classification can be performed is also calculated. As the value of A_z approaches 1, it indicates perfect classification and the results indicate that the samples of 'defective' and 'non-defective' classes are almost perfectly classified with $A_z = 0.985$ for both the classes.

4.2   Custom CNN model

In the proposed custom CNN model for color images, the model was trained and tested on batch of 64 images for 50 and 100 epochs. AdaBound optimization algorithm, with initial and final learning rate initialized to 0.0001 and 0.1 respectively, was used to minimize the cross entropy error loss. AdaBound is a variant of Adam optimizer that covers both faster convergence and better generalization [51]. The train and test accuracy plot for custom CNN model trained on color image dataset for 50 epochs is depicted in Figure 7(a). The trend in test accuracy and test loss is similar to the one resulting from fully connected layers solution. So the model does not overfit. At the end of training, the train accuracy is 93.98%, train loss is 0.14174 and validation accuracy is 91.52%, validation loss is 0.280. Time taken for training is 3s2ms per epoch.

The same custom CNN model is trained and evaluated on original X-ray image dataset with 572 X-ray images. And later the performance of the same model is evaluated on huge dataset of 5,000 X-ray images prepared by using data augmentation. The AdaBound optimization algorithm is used to minimize the cross entropy error loss on batch size of 32 for original dataset and batch size of 400 for augmented dataset. The train and test accuracy plot for custom CNN model trained on original and augmented dataset for 50 epochs is depicted in Figure 7(b) and Figure 7(c), respectively. In Figure 7(a), the learning curve for train accuracy continues to increase even after 50 epochs indicating that the model is not fully trained for 50 epochs. But the test loss curve starts to gradually increase after 40 epochs indicating that the model is uncertain about the predictions made after 40 epochs. So, there is no point in continuing the training further. However for augmented X-ray dataset the custom CNN model also gave significant classification accuracy of 98.7%. The training and testing learning curves show a good fit with an expected 'generalization gap'.
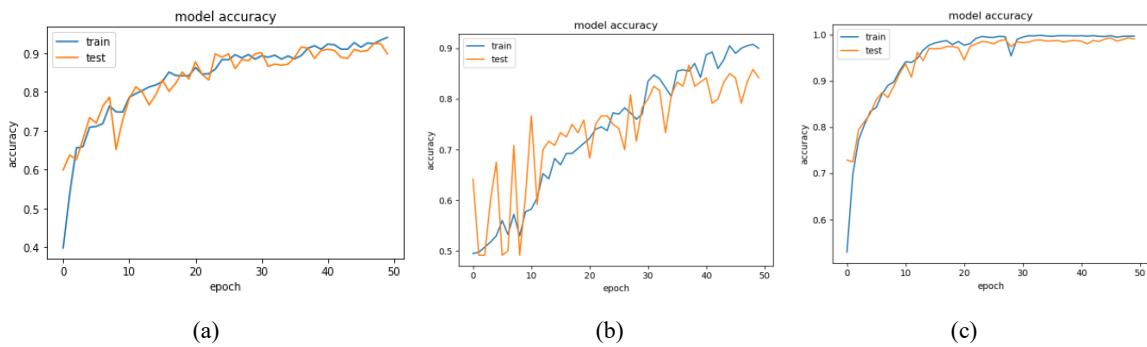

| (a) | (b) | (c) |

Figure 7: Accuracy plot of custom CNN model for (a) Color image dataset (b) Original X-ray image Dataset (c) Data augmented X-ray image dataset.

The performance of the custom CNN model for different number of epochs with and without dropout regularization for color image dataset and with and without data augmentation for X-ray image dataset are summarized in Table 10. Maximum validation accuracy of 91.52% and 98.7% was achieved by CNN model trained for 50 epochs with dropouts for color and X-ray image dataset, respectively.

Table 10. Performance summary of custom CNN model.

Accuracy and Loss	Color image Dataset			X-ray image Dataset	
	Without Dropouts	With Dropouts		Without data augmentation	With data augmentation
	50 Epochs	50 Epochs	100 Epochs		
Train Accuracy in %	96.3	93.98	97.99	90	99.8
Train Loss	0.128	0.1474	0.0575	0.233	0.0068
Validation accuracy in %	90.12	**91.52**	90.35	84.1	**98.7**
Validation Loss	0.398	0.280	0.260	0.427	0.054
Train time per epoch	3s4ms	3s2ms	3s2ms	1s2ms	13s2ms

4.3   Pre-trained CNN models

In the proposed VGG16 model for color image dataset, the Adam optimizer with 0.001 learning rate and softmax classifier with categorical cross entropy loss function was used for training the model for 50 and 100 epochs. When the model was trained for 50 epochs, the classification accuracy on validation set was 87.7% which is very low when compared with CNN model that was trained from scratch. The training was painfully slow with 26 seconds per epoch and the model was found to overfit during initial training with training loss more than test loss. Training the model for 100 epochs slightly improved the performance to 88.3% accuracy, but at the cost of computing time and resource.

The same VGG16 model when trained for 100 epochs on original X-ray dataset with a batch size of 16, resulted in 79.8% validation accuracy. However, the learning curves of both test accuracy and test loss exhibit noisy transitions indicating the unstable learning pattern of the model. Elapsed time for one epoch is 7s16ms.

In the proposed DenseNet121 model, the Adam optimizer with learning rate 0.001, categorical-cross entropy loss function and batch size of 12 was used for training the model for 50 and 100 epochs on color images. Even though the maximum validation accuracy was approximately 93% and 93.27% for 50 and 100 epochs, respectively, the model accuracy and loss plots exhibited unusual spike patterns in both test accuracy and test loss indicating sudden dip in accuracy and shoot in model loss at certain epochs due to noisy values. This clearly indicates that, even if the classification accuracy is optimal the model is underfit, may not suitably fit for the training and testing dataset and there is scope for further learning and further improvements. The performance of the DenseNet121 model can be explored further either by increasing the size of the original dataset or through data augmentation but at the cost of computationally intensive evaluation.

For original X-ray images, the DenseNet121 model was trained for 25 epochs with AdaBound optimizer on batch of 16 images. The learning curves of test accuracy and train loss clearly indicate that the model is underfit. The performance summary of VGG16 and DenseNet121 CNN architectures at the end of 100 and 25 epochs respectively for X-ray images and at the end of 100 epochs for color images is given in Table 11. The classification accuracy of both models for original dataset is low compared to CNN models which are trained from scratch. The pre-trained models performance can further be improved by increasing the size of the dataset. However, the network architecture weights of these two models themselves are quite large in terms of storage and bandwidth and elapsed training time is quite high. Since the proposed custom CNN architecture gave best classification accuracy on augmented dataset, pre-trained models were not explored for the same.

Table 11. Performance details of VGG16 and DenseNet121 models for color and X-ray images.

Accuracy and Loss	Color image Dataset		X-ray image Dataset	
	VGG16	DenseNet121	VGG16	DenseNet121
Train Accuracy in %	93.17	98.3	84.9	91.68
Train Loss	0.1658	0.0402	0.353	0.164
Validation accuracy in %	88.3	93.27	79.8	78.9
Validation Loss	0.298	0.2982	0.402	0.510
Train time per epoch	35s16ms	64s40ms	7s 15ms	8s18ms

The results of internal and external disease classification of mango fruits using color and X-ray images with existing machine and deep learning approaches for different datasets are presented in Table 12. It can be observed that for both color image and soft X-ray image datasets the classification accuracy of the proposed MLP NN on the

original datasets with just 2 hidden layers based on very few selected features performs better than other existing deep learning based methods on augmented datasets. And, with augmented dataset of X-ray image, the classification accuracy of proposed DL method is slightly higher than the existing method.

Table 12. Results comparison of similar methods on different mango fruits datasets.

Method	Color image Dataset		X-ray image Dataset	
	Accuracy (%) using ML approach	Accuracy (%) using DL approach	Accuracy (%) using ML approach	Accuracy (%) using DL approach
Sapan Naik, 2019 [24]	-----	83.97	-----	-----
Varsha Bhole, 2020 [26]	-----	93.33	-----	-----
Ansah FA, 2023 [31]	-----	-----	-----	97.66
Proposed	**95.1**	**91.52**	**97.5**	**98.7**

5   Conclusion

In this work, an empirical analysis has been employed to build a rapid, robust, real-time, non-destructive computer vision based quality assessment model for mango fruits. Since there was lack of benchmark dataset availability two datasets of mango fruits were developed for external and internal quality evaluation using color and soft   X-ray imaging techniques, respectively. Even though same machine learning and deep learning models were used, the results remain almost similar for deep learning models but deviate slightly for machine learning model. For external quality evaluation, the best selected features for which MLP NN gave the maximum disease classification accuracy of 95.1% were found to be intensity features. However for internal quality evaluation using X-ray images, contribution of geometric features is significant and resulted in 97.5% classification accuracy, which is significant considering the fact that unlike color images, X-ray images contain more noise and with the availability of just single intensity channel, gray, the number of features is also limited. From this it is also clear that exploring huge feature vector space with various intensity and geometric features for training machine learning models disease classification is not necessary as significant classification accuracy have been achieved in the proposed work using basic and fewer features.

From the deep learning CNN techniques employed on the color images of mango fruits, it is observed that the custom CNN architecture with limited depth size exhibited promising defects classification results inferring 'generality' from few samples and its ability to deal with unseen image samples was quite commendable. However, the pre-trained CNN models with deeper architectures, VGG16 and DenseNet121, exhibited high variability in results and made it difficult to draw a general conclusion for the same dataset. The performance of the proposed three deep CNN models on original X-ray image dataset was reasonable with values ranging from 78.2% to 84.1%. However, the analysis of the model accuracy plot and model loss plot indicate that custom CNN model overfit and VGG16 and DenseNet121 highly underfit the training dataset. However, custom CNN performed extremely well on augmented X-ray image dataset with classification accuracy of 98.1% with minimal binary cross entropy loss. From disease classification using deep learning CNN models it is concluded that, pre-trained deep CNN models exhibits high variability in results and makes it difficult to draw a general conclusion, while classic CNN models with fewer number of layers show promising results for the defect classification of mango fruits. Also, from the results it is evident that the simple statistical learning method like MLP NN with less feature engineering, fewer layers and reasonable computational time (0.44 and 0.28 seconds for color and X-ray image datasets, respectively) is appropriate for automating disease detection and classification in the medium sized datasets than the costlier (in terms of storage, bandwidth and elapsed training time) deep learning techniques. Further, the accuracy of quality prediction of mango fruits using non-destructive technique could be improved by correlating external and internal quality attributes.

References

[1] Jha SN, Matsuoka T, "Non-destructive techniques for quality evaluation of intact fruits and vegetables- Review", Food Science Technology 6 (4), pp. 248–251, 2000.

[2] Kader AA, "Quality assurance of harvested horticultural perishables", *IV International Conference on Postharvest Science*, pp. 51-56, 2000.

[3] Judith A. Abbot, "Quality measurement of fruits and vegetables", Postharvest Biology and Technology 15, pp. 207–225, 1999. Cakmak H, "Assessment of fresh fruit and vegetable quality with non-destructive methods", In Food quality and shelf life, Academic Press, pp. 303-331, 2019.

[4] Adedeji AA, Ekramirad N, Rady A, Hamidisepehr A, Donohue KD, Villanueva RT, Parrish CA, Li M, "Non-destructive technologies for detecting insect infestation in fruits and vegetables under postharvest conditions: a critical review", Foods, 9(7):927, 2020.

[5] Jha SN, Kingsly AR, Chopra S "Physical and mechanical properties of mango during growth and storage for determination of maturity", Journal of Food engineering, 72(1), pp. 73-6, 2006.

[6] Jantra C, Slaughter DC, Roach J, Pathaveerat S, "Development of a handheld precision penetrometer system for fruit firmness measurement", Postharvest Biology and Technology, 144, pp. 1-8, 2018.

[7] Jha SN, Narsaiah K, Sharma AD, Singh M, Bansal S, Kumar R, "Quality parameters of mango and potential of non-destructive techniques for their measurement-a review", Journal of Food Science Technology, 47(1), pp. 1-14, 2010.

[8] Vargas-Ortiz MA, De la Cruz-Medina J, de los Monteros JE, Oliart-Ros RM, Rebolledo-Martinez A, Ramírez JA, García HS, "Effect of high hydrostatic pressure on the physiology of Manila mango", Plant foods for human nutrition, 68(2), pp. 137-44, 2013.

[9] Sandeep S Musale, Pradeep M Patil, "Database development of defective and healthy Alphonso mangoes", International Journal of Advances & Environmental Engg, (IJAEE) Vol. Issue 1, ISSN 2349-1523, 2014.

[10] Adel A. Kader, "Mango quality attributes and grade standards: A review of available information and identification of future research needs", pp. 1-28, 2008.

[11] Yildiz, Fikret, Selman Uluisik, Ahmet Turan Özdemir, and Hakan İmamoğlu, "Non-destructive Testing (NDT): Development of a Custom Designed Ultrasonic System for Fruit Quality Evaluation", Nondestructive Quality Assessment Techniques for Fresh Fruits and Vegetables, pp. 281-300, Singapore: Springer Nature Singapore, 2022.

[12] Saldaña E, Siche R, Luján M, Quevedo R, "Review: Computer vision applied to the inspection and quality control of fruits and vegetables", Brazilian journal of food technology, 16(4), pp. 254–272, 2013.

[13] Alander JT, Bochko V, Martinkauppi B, Saranwong S, Mantere T, "A review of optical nondestructive visual and near-infrared methods for food quality and safety", International. Journal of Spectroscopy, 2013.

[14] El-Mesery HS, Mao H, Abomohra AE, "Applications of non-destructive technologies for agricultural and food products quality inspection", Sensors, 19(4), pp. 846-54, 2019.

[15] Kaur G, Kapoor S, Gandhi N, Sharma S, "Techniques for Quality Estimation of Fruits", Emerging Technologies for Shelf-Life Enhancement of Fruits, Apple Academic Press, pp. 345-377, 2020.

[16] Siedliska A, Baranowski P, Zubik M, Mazurek W, Sosnowska B, "Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging", 139, Postharvest Biology and Technology, pp. 115-26, 2018.

[17] Pujitha, N.,Swathi, C.,Kanchana, V, "Detection of external defects on mango", International Journal of Applied Engineering Research, 11(7), pp. 4763-4769, 2016.

[18] Haff RP, Saranwong S, Thanapase W, Janhiran A, Kasemsumran S, Kawano S, "Automatic image analysis and spot classification for detection of fruit fly infestation in hyperspectral images of mangoes", Postharvest Biology and Technology, 86, pp. 23-8, 2013.

[19] Sapan Naik, Patel B, "Machine vision based fruit classification and grading-a review", International Journal of Computer Applications, 170(9), pp. 22-34, 2017.

[20] Bhargava A, Bansal A, "Fruits and vegetables quality evaluation using computer vision: A review", Journal of King Saud University-Computer and Information Sciences, 2018.

[21] Zheng H, Lu H, "A least-squares support vector machine (LS-SVM) based on fractal analysis and CIELab parameters for the detection of browning degree on mango (Mangifera indica L.)", Computers and Electronics in Agriculture, 83, pp. 47-51, 2012.

[22] Pujari JD, Rajesh Y, Abdulmunaf SB, "Grading and classification of anthracnose fungal disease of fruits based on statistical texture features", International Journal of Advanced Science and Technology 52(1), pp. 121-132, 2013.

[23] Khoje S, Bodhe S, "Comparative performance evaluation of size metrics and classifiers in computer vision based automatic mango grading", International Journal of Computer Applications, 61(9), pp. 1-7, 2013.

[24] Sapan Naik. "Non-Destructive Mango (Mangifera Indica L., CV. Kesar) Grading Using Convolutional Neural Network and Support Vector Machine", Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), 26-28 February, Elsevier, 670-678, 2019.

[25] Vani Ashok and Vinod, DS, "A novel fusion of deep learning and android application for real-time mango fruits disease detection". Intelligent System Design, Springer Singapore, pp. 781-791, 2021.

[26] Varsha Bhole, and Arun Kumar. "Mango quality grading using deep learning technique: Perspectives from agriculture and food industry", In *Proceedings of the 21st annual conference on information technology education*, pp. 180-186, 2020.

[27] Cubero S, Aleixos N, Moltó E, Gómez-Sanchis J, Blasco J, "Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables", Food and bioprocess technology, 4(4), pp. 487-504, 2011.

[28] Barcelon EG, Tojo S, Watanabe K, "Relating X-ray absorption and some quality characteristics of mango fruit (Mangifera indica L.)", Journal of agricultural and food chemistry, 47(9), pp. 3822-5, 1999.

[29] Sharma RR, Krishna KR, "Non-destructive evaluation of Jelly Seed Disorder in Mango", Indian Journal of Agricultural Sciences, pp. 1-7, 2017.

[30] Krishna KR, Sharma RR, Srivastav M, "Non-destructive detection of jelly seed disorder in Amrapali and Pusa Surya varieties of mango (Mangifera indica)", Indian Journal of Agricultural Sciences, 89(1), pp. 95-8, 2019.

[31] Ansah FA, Boateng MA, Siabi EK, & Bordoh PK, "Location of Seed Spoilage in Mango Fruit using X-ray Imaging and Convolutional Neural Networks", Scientific African, e01649, 2023.

[32] Ketkar N, Santana E, "Deep Learning with Python", Berkeley, CA: Apress; 2017 April.

[33] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T, "Recent advances in convolutional neural networks", Pattern Recognition, 77, pp. 354-77, 2018.

[34] Canziani A, Paszke A, Culurciello E, "An analysis of deep neural network models for practical applications", arXiv preprint arXiv:1605.07678. 2016.

[35] Tang Y, "Deep learning using linear support vector machines", arXiv preprint arXiv:1306.0239, 2013.

[36] "Preface-mango, Post-Harvest Profile of Mango", Government of India, Ministry of Agriculture, (Department of Agriculture & Cooperation), Directorate of Marketing & Inspection, Branch head office, Nagpur, 2013.

[37] Veena T, Chidanand DV, Alagusundaram K, "Soft X-ray imaging system for improving quality classification of fruits and vegetables: a review", International Journal of Agricultural Science and Research (IJASR), 5(3), pp. 25-35, 2015.

[38] Pizer SM, Eugene JR, James P, Ericksen, Bonnie CY, Keith EM, "Contrast-limited adaptive histogram equalization: speed and effectiveness", In *Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337-345, 1990.

[39] Vani Ashok, and D. S. Vinod, "A comparative study of feature extraction methods in defect classification of mangoes using neural network", *In Proceedings of Second International Conference on Cognitive Computing and Information Processing (CCIP)*, IEEE, pp. 1-6, 2016.

[40] Rumelhart D E, Hinton G E, Williams R J, "Learning internal representation by error propagation", Parallel Distributed Processing I, MIT Press, pp. 3 18-362, 1986.

[41] Pardede J, Sitohang B, Akbar S, Khodra M L, "Implementation of transfer learning using VGG16 on fruit ripeness detection", I. J. Intelligent Systems and Applications, pp. 52-61, 2021.

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

[43] Huang G, Liu Z, Weinberger KQ, van der Maaten L, "Densely connected convolutional networks", arXiv 2016. arXiv preprint arXiv:1608.06993, 2018,.

[44] Larsson G, Maire M, Shakhnarovich G, "Fractalnet: Ultra-deep neural networks without residuals", arXiv preprint arXiv:1605.07648, 2016.

[45] Fine TL, "Feedforward neural network methodology", Springer Science & Business Media, 2006 April 6.

[46] Møller MF, "A scaled conjugate gradient algorithm for fast supervised learning", Neural networks, 6(4), pp. 525-33, 1993.

[47] Ioffe S, Szegedy C, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167, 2015.

[48] Mery D, Lillo I, Loebel H, Riffo V, Soto A, Cipriano A, Aguilera JM, "Automated fish bone detection using X-ray imaging", Journal of Food Engineering, 105(3), , pp. 485-92, 2011.

[49] Mery D, Soto A, "Features: the more the better", In *Proceedings of the 8th conference on signal processing, computational geometry and artificial vision*, pp. 46-51, 2008.

[50] Canziani A, Paszke A, Culurciello E, "An analysis of deep neural network models for practical applications", arXiv preprint arXiv:1605.07678, 2016.

[51] Kingma DP, Ba J, "Adam: A method for stochastic optimization", arXiv preprint arXiv:1412.6980, 2014.