# Forecasting of Fruits Stock Life using CNN-based Deep Learning Techniques: A Comprehensive Study [*]

Neha Gautam[1,*,†], Nisha Chaurasia[2,*,†]

[1]*Department of Computer Science and Engineering, Dr. B.R Ambedkar National Institute of Technology, Jalandhar, India*
[2]*Department of Computer Science and Engineering, Dr. B.R Ambedkar National Institute of Technology, Jalandhar, India*

### Abstract

Fruits have lavish fibre and nutrients such as proteins, Vitamin A, C, and E, Folic acid, Magnesium, Zinc, Phosphorous and others which needs to be pampered for high gain. Fruits freshness have a short life, especially during the time of supply. Suppliers due to a lack of accurate knowledge about fruits freshness forecasting during the sorting and packaging process, supply such fruits that are unfit for consumption because the fruit's freshness gradually decays over days. For detecting fruits spoilage at the initial production stage of consumption is necessary to reduce fruits being rotten/spoiled. Automatic fruit grading on the basis of quality and characteristics is a commercially important process to obtain high fruits production in the food industry. This can be done traditionally but it would be time consuming, costly and required more labour and human being can be exhausted and bored after doing the same work which is not the case with machines. Fruit businesses completely rely on the quality of fruits based on colour, texture, physical appearance, shape and size incurring fast and effective methods to know the grading and worth of fruits. The exact evaluation of fruits products assumes a significant part in the rural and food industry to expand the benefit and to upgrade intensity. In this way, the nature of organic products plays a fundamental job as it is utilized in assortments of utilizations like the product, creating food things like organic product juice, jams, and so on that are healthy for human beings. The unfit fruits production affects the economy of any country indirectly and the level of emission of carbon dioxide ($CO_2$). In this paper, several methods have been implemented by reviewed authors to predict the quality of the fruits. Also, this study puts emphasis on the need to timely sell out the stock which reduced loss to the sellers.

### Keywords

CNN, SVM, K-means, Image enhancement, Fruits Classification, Recognition, Segmentation

## 1. Introduction

The fruits industry is the reinforcement of the Indian economy where the quality of fruits creation assumes a significant part. In this era, each food industry throughout the world wide wants to use such a technology that is automatic by which time and money could be saved to analyse and recognise the grading of fruits quality. India positions second in the world in the production of fruits. It offers India gigantic open doors for the export of fruits [6]. A lot of expert experience and knowledge is required in traditional Methods for fruits quality analysis. Fruits

✉ nehagautam796@gmail.com (N. Gautam); chaurasian@nitj.ac.in (N. Chaurasia)
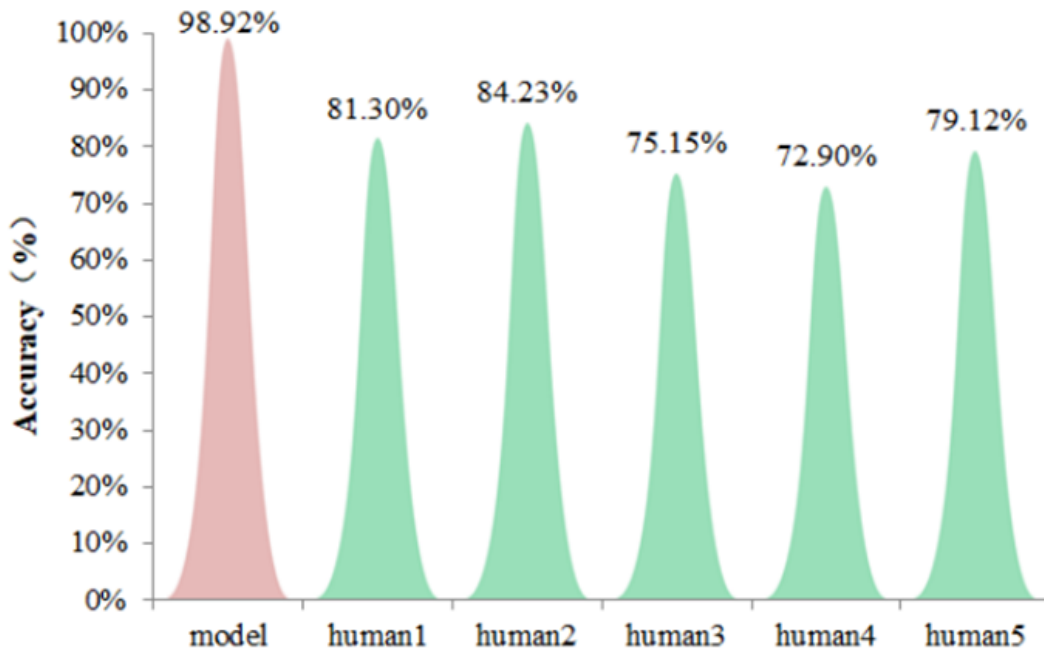
**Figure 1:** Comparison of deep learning model with human experts [7].

quality supervision can be inconsistent due to human to human based on their skill and physical factors [19]. In paper [7] author compared deep model to the human expert and found the result given in Figure 1 where deep model is more accurate than the expert. Hence to decrease human labour and effort with accurate and fast result, we are required an automatic identifier system by using this fruits production can be increased [10]. In this paper, various efforts are made by authors to develop such a system using CNN-based deep learning techniques such as ResNet50, MobileNet V2, DenseNet-121, NASNet-A, SVM and EfficientNet B0-B2 and transfer learning.

In this paper, we are going to depict CNN-based deep learning techniques to classify the fruits on the basis of fruits quality. This paper is divided into six segments. The first section is a brief introduction and in the second section, we will review the already published paper in the field of fruits quality prediction using CNN-based deep learning techniques in the literature review. In the third section, the methodology used in experiments is explained. In the fourth section, CNN architecture is briefly explained. CNN-based deep learning techniques are summarized. In the last section, the paper is finished after coming to a conclusion.

## 2. Related Work

Several researchers went through the various experiments and ideas about fruit quality analysis using deep learning techniques. The precise review of various related papers is discussed here. N. Ismail et al. [1] proposed a machine vision system using stacking ensembled deep learning

techniques that automatically inspect the quality of the fruits, providing the real-time visual inspection facility for interaction. To remove the noise, Gaussian filter with a value of 0.01 having a 3X3 kernel is used and for smoothing, cross-correlation function is applied on images, Histogram equalization, contrast limiting threshold has been used using OpenCV in image pre-processing. For image segmentation, Mean shift clustering, Otsu thresholding, and watershed segmentation techniques. ResNet50, DenseNet121, NASNet and EfficientNet B0-B2 architectures are used. The learning rate range test and Bayesian Optimization techniques were used to find the optimal value of learning rate and hyperparameters respectively. Specificity, Area under ROC Curve and sensitivity have been selected to evaluate the proposed model. EfficientNet-B2 gave a high performance that provide a 99.2% recognition rate. To improve models, multiple features of EfficientNetB0+ B1+ B2 were stacked. Sovon et al. [2] designed a model using deep CNN that will prevent fresh fruits to be contaminated by other rotten fruits. MobileNetV2 architecture having 19 layer was used to classify and recognize the rotten fruits. Max pooling and average pooling were compared the base of accuracy and resulted max Pooling give higher accuracy that is 99.46 % for training and 99.61% for validation. Lili Zhu et al. [3] presented a system based on mobile visual using two layers image processing that will grade banana consisting label as unripen, ripened, overripened further, well ripened and mid ripened are two class of ripened banana. In first layer, SVM (Support Vector Machine) is used to classify the banana that give 98.5% and in second layer, YOLOv3 locates defected area of peel which accuracy is 87.5%. CycleGAN for augmentation, K-means for segmentation, SVM for classification, YOLOv3 for grading, edge clouding to get less network communication as well as to reduce computational resource were used. Recall, Precision, F1-Score are used for first classification evaluation while for YOLOv3 mean Average Precision recall and IoU (intersection over union) were applied. V. Bhole et al. [4] worked using 4560 thermal images with resolution 720x1280 as well as RGB 2322x4128 pixel of same mangoes images having 19 classes and proposed such a predictor system that forecast the remaining time of mangoes using transfer learning concept based on lighted weighted CNN architecture like MobileNetv2, ShuffleNet and SqueezeNet. After experimentation found that ShuffleNet was faster than MobileNetv2 and SqueezNet. Precision, recall, F1-score, false discover rate (FDR) and false positive rate (FPR) was used to measure the performance and resulted thermal imaging outstrip RGB with accuracy 98.15O. M. Lawal et al.[5] proposed a robust model, can be used in robots to harvest the fruits(muskmelon), named YOLOMuskmelon for fast and accurate muskmelon detection using RELU activated ResNet43 as backbone with SPP(spatial pyramid pooling) for optimization, CIoU(complete intersection over union) loss for better performance and fast convergence, residual block arrangement to prevent Vanishing gradient, FPN(feature pyramid arrangement) to generalize the models and DIoUNMS(distance intersection over union non maximum Suppression) for overlap area consideration. YOLO Muskmelon detect the muskmelon with the speed of 96.3 frame per second and 89.6 % precision value and 56.1% faster than YOLOv4. K. Zhang et al. [6] proposed an automatic harvesting robot for strawberry using R-YOLO (rotational you only look once) that was designed using YOLOv3 with MobileNet V-1 as backbone to extract features. In R-YOLO, rotational angle of Fruits axis was calculated for precise localization of real-time video images of strawberry detection. In R-YOLO (x, y, w, h, alpha) were the bounding box parameter where alpha is angle Between long side of bounding box and y-axis, (x, y) are center coordinate, w is width and h as height of bounding box. Feature extraction, recall and recognition accuracy is

adversely affected because of using MobileNet but R-YOLO is 3.6 time faster than YOLOv3 that was 18 frame per second. N. Stasenko et al. [7] trained a model for detection and prediction of the decaying surface of apples using U-Net and DeepLab CNN architecture to improve the storage process of apples. There is done investigation for the performance of U-Net and DeepLab based on mIoU(mean intersection over union) then found U-Net and DeepLab yield 99.71% and 99.99% mIoU respectively. A testbed was used to capture RGB images of apples having 12000 images and four class Malus Domestica Borkh, Fuji, MiroLeto, and Golden and annotated into corresponding JSON files. To extract the feature ResNet was used as a backbone, ImageNet as an encoder weight, loss, and mIoU as a performance metric, learning rate 0.001, batch size was 4 for U-Net. In DeepLab Atrius convolution to control the size of the feature, ASPP (atrous Spatial Pyramid Pooling). Chai C. Foong et al. [8] proposed a model to classify the rotten fruits having six class fresh and rotten apples, oranges, and bananas,350 images each using ResNet50. Segmentation using color threshold function, feature extraction, and HSV color technique is used to detect the background of the image. The author runs the model with segmentation and without segmentation on the same parameters batch size 10, learning rate 0.0001 and epoch 6, found without segmentation yield the same accuracy that is 98.9% to segmentation in less time. Jiangong N. et al. [9] proposed such a model that will forecast the banana's storage time and freshness using transfer learning and GoogleNet not involving any destructive detection. The images of banana were captured in 11 days to create the dataset and apply augmentation to adopt generalization and avoid overfitting of the network, gradient classification activation mapping (Grad-CAM) for feature extraction were used. To determine the model generalization strawberry images were used for training and testing and yielded a 92.47% accuracy rate. S. Bulla et al. [10] proposed a model that will help to prevent spreading the of rottenness by classifying the rotten and fresh fruits, applying transfer learning. There is used max-pooling layer to get rid of overfitting, reduce the amount of memory and time required for computation, Batch Normalization for feature map normalization, and dropout that was 0.5 for fast computational speed at each stage of convolution layer and regularization is used for adding penalties used in loss function on the layer while optimizing. There are used Random_uniform to initialize the kernel, bias, and weight that can be updated according to the output value. Categorical cross-entropy as a loss function and Adam as an optimizer with 0.0001, learning rate 16 batch size and 225 epoch was used achieved 97.8% accuracy. The Proposed model used fewer filters and parameters that decrease computational time, memory usage, which makes a feasible model to predict fresh and rotten fruits. K. Roy et al. [11] proposed a deep learning model using real-time semantic segmentation for rotten parts of apple to detect and categories fresh or rotten apples on the base of peel presented on the surface of apple's RGB image using a deep learning model and dataset downloaded from Kaggle. There were 3102 images of apple in the dataset and the batch size was 97 and each batch has 32 images. En-UNet, in which U-Net was used as a backbone, was used to segment the images and yield 97.46% training and 97.54% validation accuracy, While U-Net accuracy was 95.36% and 0.066 training and 0.062 validation loss. the performance of the model evaluated on the basis of accuracy, loss, and mean IoU using 0.95 threshold value. During image pre-processing RGB images are converted into gray images, threshold and binarization were used to get output. S. Bulla et al. [12] proposed a CNN model that will automatically recognize and classifies fruits dataset consisting of two types of images one is public having 758 images with a simple background consisting of 5

classes and another is himself creating dataset having 1152 images with a complex background consisting 7 classes. In pre-processing multi-channel images are converted into mono(blue) images, the global threshold value is used for segmentation and Halcon software was used for the identification and classification of fruits. The author experimented on various batch values in the CNN model and found 56 and 64 were the best balance between memory capacity and efficiency. To improve the classification accuracy enhancement techniques were combined three methods of image enhancement such as random flip, random crop, and enhanced brightness after the experiment combination of random flip and enhanced brightness yielded 98.1% accuracy. L. Wu et al. [14] designed a model using data augmentation techniques and YOLOv4 to enable such a robot that will pick the apples quickly and accurately in orchards having complex backgrounds and used crawler technology for image labeling. EfficientNet replaced CSPDarknet53(Cross Stage Partial Darknet53) and was used as the backbone in addition Convolution layer was also added to adjust and extract the feature that will reduce the computational complexity making model lighter. The result shows that YOLOv4 with EfficientNet-B0 gave better performance comparable to YOLOv4, YOLOv3, and Faster R-CNN with ResNet for apple detection on the base of precision, recall value, and F1 score. M. O. Lawal et al. [15] used modified YOLOv3 to propose YOLO-tomato model which detects ripe and unripe tomatoes having complex environments. The author conducted an experiment on trained YOLO-tomato, YOLOv3, and YOLOv4 to verify the effectiveness of models on the basis of precision, recall, F1 score, and average precision (AP), in which YOLO-tomato using mish and SPP to reduce missed detection and inaccuracies, showed the best performance with 99.4% AP, better generalization and real-time detection. This model can be used in harvesting robots in agriculture industries. V. Bhole et al. [16] created a texture-based RGB and thermal images dataset of 11 varieties of fruits keeping them on the revolving tray using the digital and thermal capturing camera in the experiment. There was done experiments using KNN (k-nearest neighbor) and RF (random forest) algorithm to classify the images of fruits and evaluated the performance of the classifier on the basis of accuracy and Kappa value. KNN with RGB images showed more accuracy than RF. M. Oltean et al. [17] dataset having 90380 images was created and built a deep learning model that will identify fruits from an image having single and multiple fruits. The consumption of computational resources is improved by adjusting the depth and width of the network having constant computational power. The model was trained on RGB, HSV, and grayscale images and RGB images show outperform comparable to others and yield 99.86% testing accuracy and 100% training accuracy. F. Valentino et al. [18] proposed a CNN model based on computer vision for fruits freshness detection and dataset was downloaded from Kaggle having 6 classes of fruits named fresh apple, banana and oranges and rotten apple, banana, and orange. The value of dropouts was taken 0.25 to avoid overfitting and reduce the size of data. There is a used web application using Python Flask for testing and this web can be accessed through mobile and PC using the browser. Further, Table 1 describes related work briefly.

## 3. Methodology

The artificial brain is trained and modeled through design and artificial neuron that mimics the human brain neuron. These review papers confer fruits quality and grading prediction
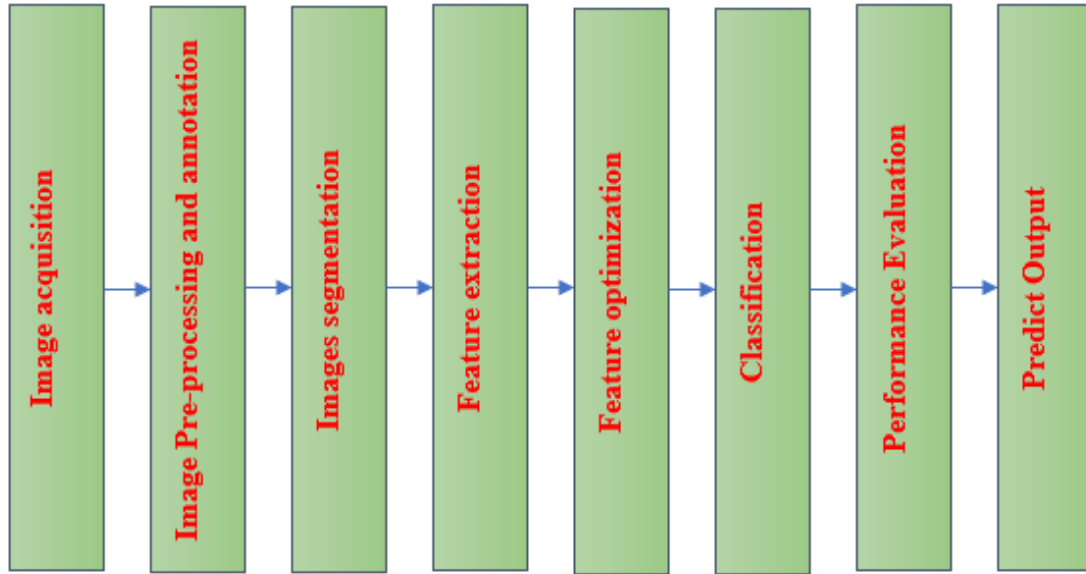
**Figure 2:** Block diagram of fruit grading system

based on classification, computer vision technology, and CNN model such as AlexNet, VGG16, VGG19, ResNet, GoogleNet, MobileNet e.t.c on the base of colour, texture, and shape of fruits. Firstly, fruits are detected then fruits texture behaviour is analysed, then grading to the fruits is assigned. On the basis of grading, time can be determined in which this particular type of fruit will completely rotten. Figure 2 shows the block diagram of the CNN base fruits quality grading system. The fruits images are separated into multiple classes by labeling them fruits type [2]. Image pre-processing, segmentation, feature extraction, and classification is used in the training phase.

## 3.1. Image Acquisition

In computer vision and image processing, image acquisition is the first step in which images are retrieved from sources using various hardware systems like Ultrasound, Tomographic imaging, stereo system, magnetic resonance image (MRI), X-ray, Thermal imaging [3] give significant result as compared to RGB imaging [4] because thermal works on the internal feature of image [4]. Machine vision system (MVS) can be used in which, images can be acquired from real-time video and photographs [3]. Drive Webcam can be used to capture a real-time image and these images are saved in google drive. Webcam can be used to capture real-time image and these images are saved in google drive.

## 3.2. Image pre-processing

The source images that are often corrupted due to poor illumination and undesirable high frequency signal, are processed to increase image information and improve the quality of raw data using various filters such as Gaussian filter [1] to remove the noise, median filter [6],

rank filter log transformation [3] can be used to reduce and improve the contrast of the image. Images can be resized according to the aspect ratio of the original image to maintain the aspect ratio of the original image [2]. The ratio of width and height is known as the aspect ratio [4]. image Annotation [5][6] [8][15] and Labeling (LWYS approach) [15] are used for annotation and labeling the image. Let consider I (x, y) is an original illuminated source image and r (x, y) is a reflected image or filter with pixel (x, y). Then the image is processed using the following formula shown in equation 1 which is known as processed image denoted by f (x, y).

$$f(x, y) = I(x, y) \star r(x, y) \tag{1}$$

### 3.3. Data Augmentation and Enhancement

It is a technique to amplify images without label changing to enlarge dataset by applying cropping, brightness, dropout [1][4][14] rotation, zooming, contrast changing, mirroring, translation [9] [14], horizontally and vertically flipping, shearing [2] and shifting methods. The size of dataset volume can be increased using CycleGAN [3]. In paper [12], there were randomly combined image enhancement methods, such as the first is a combination of random cropping and random flip, the second is a combination of enhanced brightness and random flip, the third is a combination of random crop and enhanced brightness, and the fourth is a combination of random crop, random flip and enhanced brightness. Image enhancement method has a significant improvement in the classification and recognition of self-made data sets [12].

### 3.4. Image segmentation

Image segmentation is a mechanism to partitioned the target from background to extract the region of interest and understand and analysis the important segment of an image on the base of color, texture, shape, brightness, contrast, and gray level characteristics of fruits [3] and assign a label to every pixel in the image. There are various techniques for segmentation such as clustering, threshold method, edge-based on segmentation, partial differential equation base techniques and ANN-based segmentation, k-means for segmentation [3], Grad-CAM (Gradient weighted class activation method) [9][1].

### 3.5. Feature Extraction

To decrease the number of resources and reduce the dimensionality needed to describe and analyze large data feature extraction is performed. Color features (color coherence vectors, color moments, color sets, and color histograms), RGB, and HSV (for color segmentation HSV can be used) for statistical analysis. HSV is more suitable comparable to RGB [3], texture features [LBP local binary pattern], shape features, spatial features, PCA [3] techniques, and GLCM (grayscale co-occurrence matrix) can be used to extract texture base feature of image [9][16].

### 3.6. Classification

Image classification is a process in which an image is taken as an input and gives the output in form of the probability of each class. To whom class input image belongs, would have the

highest probability. On the basis of this probability, value images are classified having labeled. KNN, Random-forest (consisting of many decision trees with no correlation to each other [9], Naïve Bayes, SVM (is capable to classify both linear and non-linear data in consisting high dimensional space with high accuracy [21]). For simple classification, the performance of KNN is better than SVM [3]. Deep learning is more accurate than the machine learning model for fruits classification [1]. The selection of Kernel plays an important role in classification accuracy improvement and the same dataset give different accuracy on different kernel [21].

## 3.7. Evaluation and Prediction

After training and testing, next phase is to measure the model performance using Sensitivity or recall, specificity, precision, f1 score, accuracy that can be calculated using following equation 2, 3, 4, 5 and 6 respectively [2][6].

$$Sensitivity or recall = \frac{TP}{(TP + FN)} \qquad (2)$$

$$Specificity = \frac{TN}{(TN + FP)} \qquad (3)$$

$$Precision = \frac{TP}{(TP + FP)} \qquad (4)$$

$$f1 - score = \frac{(Recall * Precision)}{(Recall + Precision)} \qquad (5)$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (6)$$

# 4. CNN Architecture

In the field of image pre-processing and computer vision, CNN is the most powerful method of deep learning for image classification, digital character recognition, and object recognition [21]. CNN has a stack of convolution layers to transform input image into output as a probability of classes and yield the highest probability for that class, from which the input image belongs to. That is the reason why authors have mainly studied the papers on CNN. CNN consists of the input layer, convolution layer, normalization, Pooling, and fully connected layer.

In Figure 3, the input image has 224 height and width with 3 channel is applied to the convolution neural network layer by layer. There is taken three convolution layers that use (3X3) pool size, 64, 128, and 256 filters. RELU activation function is used to reduce the exponential growth in computation in operating CNN and batch normalization is taken to increase the computation rate. To reduce the size of the dimension of data, max pooling is passed that has a size of (2X2) filter and s indicates stride to slide the max pool filter over input image.
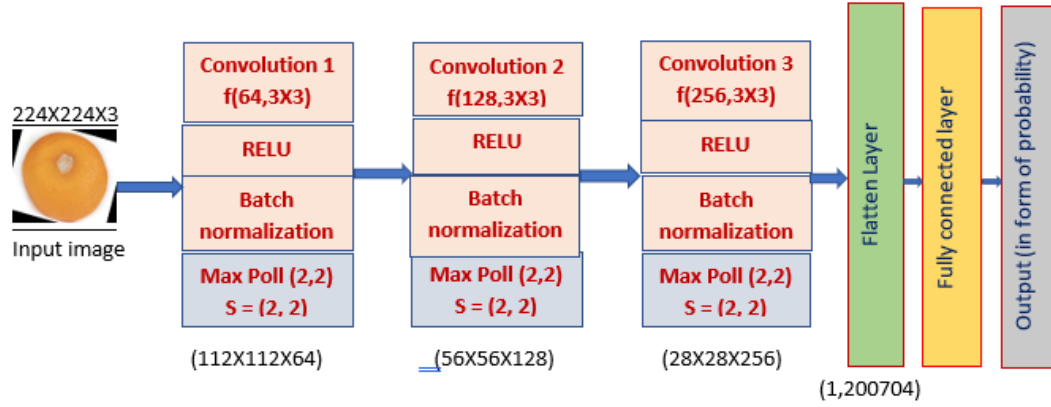
**Figure 3:** CNN architecture

## 4.1. Convolution layer

In image processing, Convolution is a basic building block that is applied for smoothing, sharpening, and edge detection. There is performed corresponding element-wise multiplication between the matrix of the input image and convolution kernel, then to represent a grid cell like a pixel in output feature map all multiplied elements are added. The convolution kernel having a square matrix of integers is applied on a subset of the input pixel value of image from the top left corner of the input image and kernel is stridden from left to right and top to bottom to apply convolution at every pixel of image [20] to get feature map. If multiple convolution kernels are applied within a convolution layer then multiple features maps are created as an output [10].

## 4.2. RELU

RELU (rectified linear unit) is used as a non-linear activation function between the convolution layer and the Pooling layer [10], which will eliminate the negative value of pixel into zero. If there is taken real-world data for the CNN model, a lot of real-world data is non-linear in nature so RELU is implemented for non-linear data on the CNN model [8]. A separate study on various activation functions such as RELU, Swish, Mish, and Leaky is done by O. M. Lawal et al. [5] to determine the most effective on the model and on the base of the P-R curve, AP (average precision) RELU perform remarkable.

## 4.3. Batch normalization

Batch normalization helps in providing stability in model prediction, overfitting reduction by using regularization and increasing the speed of training by order of magnitude. It is the process of normalization within the activation layer of the current batch subtracting the mean of the batch's activation function and dividing the standard deviation batch's activation function. SGD undo the normalization minimizing loss function [23].

### 4.4. Pooling

To reduce the dimensionality of the feature map retaining important features to avoid the overfitting pooling layer is used [8]. There are various polling methods such as min pooling, max pooling, average pooling. In max pooling, the net value is replaced by the max value of nearby elements of feature maps or channel in the window eliminating by the largest element [9][10].

### 4.5. Flatten and Fully connected layers

The image matrix finding from the final convolution layer is transformed into 1-D vector to flat the input then used as input for a fully connected layer. In a fully connected layer that is the same as ANN, all the neurons of input are connected to every neuron of the next layer and perform the mathematical operation as shown in equation [7]. Where x is input having dimension [n,1], w is weight having dimension [n, m] where n and m are number of neurons in previous and current layer respectively and b is bias having size [n, 1]. In the last, fully connected layer Softmax activation is used to predict probabilities of input being classified [4][7].

$$Y = Activation function(x * w + b) \tag{7}$$

### 4.6. Hyperparameter Tunning

Optimizers used to optimize the accuracy of the model are based on hyperparameters that are fine tuning optimization, including parameters such as the number of epochs, batch size, optimizer, iterations per epoch, and validation per step [1][18]. There is a need to set hyperparameters variables before applying learning algorithms [24]. Effects of various hyperparameters such as-

• **Effect of Batch Size :** To reduce memory usage, the number of input samples is applied to the layers of the network [10]. Batch size largely impacts on the result of the experiment if the batch size is too small then there is fear of underfitting and on taking large value there may be fear of overfitting [12].

• **Effect of Number of epochs :** One pass or iteration over the entire dataset is known as an epoch. Overfitting and underfitting are two problems that might be incurred during epochs optimization which model learns even noise that impacts negative effect on model accuracy. If the model is trained using the small number of epochs, might incur underfitting and the model is trained using large epochs that may incur overfitting. If validation error is increased then the trained model is said overfitted [24]. What could the correct number of epochs be taken is depend upon training and validation loss.

• **Effect of optimizer :** Performance of model by updating weight parameters to reduce the loss function where the loss function is differences between actual and predicted output. Feature optimization is a process that reduces the overfitting risk using mathematical functions. To converge the cost function to the global minima and for minimum misclassification

hyperparameter should be selected carefully [18].

• **Effect of learning rate :** Learning rate is known as step size, consisting of weights and biases that are initialized randomly before training and low learning rate yield overfit data while high learning rate yield underfits data and divergent nature [24]. During the training of the model, some current values of weights are updated taking place next epoch, these weights are known by learning rates consisting of a range from 0 to 1 [10]. There was seen, on reducing the value of learning rate from 0.1 to 0.0001 accuracy of the model improved from 17.36% to 97.82%. CLR is used to find appropriate learning rate in paper [1].

## 5. Deep Learning Techniques under CNN for Fruits Life Prediction

A deep neural network comprises the interconnection of neurons that perform complex tasks that are challenging for people. The input and weight are assigned to neurons then they will transform input into output. The output of the previous layer behaves like an input for neurons of the next layer [19]. Deep learning models are trained on a large amount of labeled datasets using strong computing power such as GPU [20].

### 5.1. ResNet

To solve the vanishing and exploding gradient problem, ResNet is used which is a major breakthrough in image processing. It uses skip connection technique to skip training from few layers and directly connects to output layer. ResNet consists of deep neural network. To reduce computational time, ResNet50 is used because it consists of 50 deep layers [8].

### 5.2. DenseNet

DenseNet use fewer parameters, promote the reusability of the features, lighten the problem of gradient disappearance and strengthen features flow. Connectivity, DenseBlocks, growth rate and bottleneck layers are the components of DenseNet. DenseNet121 comprise 121 layers having pretrained weights [1].

### 5.3. MobileNetv2

MobileNet is light weight deep learning network because it uses depth-wise seperable convolution [2], Mainly it is used for reducing memory consumption by reducing the number of parameters and using inverted residual which are employed between the layers of bottleneck. For this reason, it can be used in which system that has less computing power [1][2].

### 5.4. EfficientNet

It is a CNN and scaling method that scales dimension of resolution of image, depth and width of network using a set of fixed scaling coefficients. It is capable to optimize accuracy as well as

better efficiency by performing neural architectural search. The base EfficientNet-B0 is based on the inverted bottleneck residual block of MobileNetv2 [1].

### 5.5. ShuffleNet

ShuffleNet uses pointwise group convolution to reduce computation complexity and channel shuffle to flow information across feature channels. It is designed for mobile devices having efficient computation power and cost maintaining accuracy [4].

### 5.6. SqueezNet

SqueezNet is used as replacement of AlexNet having smaller network. It has 50x less parameters than AlexNet and three time faster than AlexNet. It has Sqeeze consisting 1x1 filters and expand layers consisting 3x3 filters [4].

### 5.7. GoogLeNet

GoogLeNet has 22 deep layers, is excellent CNN model which is achieve through the ImageNet pretrained data. The purpose of GoogLeNet is to reduce the networks parameters, prevent overfitting and make the network faster [9].

On the base of the parameters discussed above, the result of various authors is compared in Table 2.

## 6. CONCLUSION AND FUTURE SCOPE

Several researches have been done using state-of-art deep learning techniques for fruits quality forecasting. In this paper, several existing papers are reviewed then compared. Among all reviewed papers, existing methods yield remarkable accuracy with many challenges such as high loss rate and more detection and learning time-consuming. In these methods, only a single view of the image was used. Various image enhancements techniques were combined and applied filters to improve classification accuracy and yield higher validation accuracy. However, there is a lack of a robust and generative system that can be used for sorting, counting, rotten fruits detection, and grading in multiple fruits automatically because all researchers have used such a dataset which have only one fruit of a particular type in the images. This review will surely aid for further research work having the new design of CNN model for automatic fruits grading and quality prediction for smart agriculture industry areas. In the future, an enhanced deep learning system can be developed that can predict the time span, in how many days this fruit will rotten so that fruits could be sold before spoil. It would provide profit to the seller of the fruits industry, the customer as well as the economy of the country.

## References

[1] N. Ismail and O. A. Malik, "Real-time visual inspection system forgrading fruits using computer vision and deep learning techniques,"Information Processing in Agriculture, 2021.

[2] S. Chakraborty, F. J. M. Shamrat, M. M. Billah, M. Al Jubair, M. Alauddin, and R. Ranjan, "Implementation of deep learning methods toidentify rotten fruits," in2021 5th International Conference on Trendsin Electronics and Informatics (ICOEI), pp. 1207–1212, IEEE, 2021.

[3] L. Zhu and P. Spachos, "Support vector machine and yolo for a mobilefood grading system,"Internet of Things, vol. 13, p. 100359, 2021.

[4] V. Bhole and A. Kumar, "A transfer learning-based approach to predictthe shelf life of fruit,"Inteligencia Artificial, vol. 24, no. 67, pp. 102–120, 2021.

[5] O. M. Lawal, "Yolomuskmelon: quest for fruit detection speed andaccuracy using deep learning,"IEEE Access, vol. 9, pp. 15221–15227,2021.

[6] Y. Yu, K. Zhang, H. Liu, L. Yang, and D. Zhang, "Real-time visual localization of the picking points for a ridgeplanting strawberry harvestingrobot,"IEEE Access, vol. 8, pp. 116556–116568, 2020.

[7] N. Stasenko, E. Chernova, D. Shadrin, G. Ovchinnikov, I. Krivolapov,and M. Pukalchik, "Deep learning for improving the storage process:Accurate and automatic segmentation of spoiled areas on apples," in2021 IEEE International Instrumentation and Measurement TechnologyConference (I2MTC), pp. 1–6, IEEE, 2021.

[8] C. C. Foong, G. K. Meng, and L. L. Tze, "Convolutional neural networkbased rotten fruit detection using resnet50," in2021 IEEE 12th Controland System Graduate Research Colloquium (ICSGRC), pp. 75–80, IEEE,2021.

[9] J. Ni, J. Gao, L. Deng, and Z. Han, "Monitoring the change processof banana freshness by googlenet,"IEEE Access, vol. 8, pp. 228369–228376, 2020.

[10] S. S. S. Palakodati, V. R. R. Chirra, D. Yakobu, and S. Bulla, "Freshand rotten fruits classification using cnn and transfer learning.,"Rev.d'Intelligence Artif., vol. 34, no. 5, pp. 617–622, 2020.

[11] K. Roy, S. S. Chaudhuri, and S. Pramanik, "Deep learning based real-time industrial framework for rotten and fresh fruit detection usingsemantic segmentation,"Microsystem Technologies, vol. 27, no. 9,pp. 3365–3375, 2021.

[12] L. Wu, H. Zhang, R. Chen, and J. Yi, "Fruit classification using convolutional neural network via adjust parameter and data enhancement,"in2020 12th International Conference on Advanced ComputationalIntelligence (ICACI), pp. 294–301, IEEE, 2020.

[13] H. B.¨Unal, E. Vural, B. K. Savas¸, and Y. Becerikli, "Fruit recognitionand classification with deep learning support on embedded system(fruitnet)," in2020 Innovations in Intelligent Systems and ApplicationsConference (ASYU), pp. 1–5, IEEE, 2020.

[14] L. Wu, J. Ma, Y. Zhao, and H. Liu, "Apple detection in complex sceneusing the improved yolov4 model,"Agronomy, vol. 11, no. 3, p. 476,2021.

[15] M. O. Lawal, "Tomato detection based on modified yolov3 framework,"Scientific Reports, vol. 11, no. 1, pp. 1–11, 2021.

[16] V. Bhole, A. Kumar, and D. Bhatnagar, "A texture-based analysis andclassification of fruits using digital and thermal images," inICT Analysisand Applications, pp. 333–343, Springer, 2020.

[17] H. Mures¸an and M. Oltean, "Fruit recognition from images using deeplearning,"Acta Universitatis Sapientiae, Informatica, vol. 10, no. 1,pp. 26–42, 2018.

[18] F. Valentino, T. W. Cenggoro, and B. Pardamean, "A design of deeplearning experimentation for fruit freshness detection," inIOP Conference Series: Earth and Environmental Science,

vol. 794, p. 012110, IOPPublishing, 2021.

[19]  Y. Kumar, A. K. Dubey, R. R. Arora, and A. Rocha, "Multiclassclassification of nutrients deficiency of apple using deep neural network,"Neural Computing and Applications, pp. 1–12, 2020.

[20]  D. T. P. Chung and D. Van Tai, "A fruits recognition system based ona modern deep learning technique," inJournal of physics: conferenceseries, vol. 1327, p. 012050, IOP Publishing, 2019.

[21]  D. Karakaya, O. Ulucan, and M. Turkan, "A comparative analysis onfruit freshness classification," in2019 Innovations in Intelligent Systemsand Applications Conference (ASYU), pp. 1–4, IEEE, 2019.

[22]  J. Feng, L. Zeng, and L. He, "Apple fruit recognition algorithm based onmultispectral dynamic image analysis,"Sensors, vol. 19, no. 4, p. 949,2019.

[23]  A. F. Agarap, "Deep learning using rectified linear units (relu),"arXivpreprint arXiv:1803.08375, 2018.

[24]  S. Afaq and S. Rao, "Significance of epochs on training a neuralnetwork,"International Journal of Scientific and Technology Research,vol. 19, no. 6, pp. 485–488, 2020.

**Table 1**

Some Related Work

| Author | Dataset | Dataset Used | Methods used | Merits | Demerits |
|---|---|---|---|---|---|
| **N**. Ismail et al. [1] | Apple and Banana images | Internal feeding worm dataset of CASC | ResNet50, MobileNet V2, DenseNet-121, NASNet-A and EfficientNet B0-B2 | Deep learning based on machine learning model has cost low for grading fruits. | Classifier confuse between yellowish green and green |
| **S**ovon et al. [2] | Apple, banana and orange images | Kaggle | MobileNetV2 | Performance is very high | Loss is very high |
| **Lili** Zhu et al. [3] | 150 banana images | Author created from online source | SVM and YOLOv3, CYcleGAN | Less network communication and save computational resource | Too much small dataset is taken |
| **V**. Bhole et al. [4] | 4560 Thermal and RGB images of mangoes each | Author created Real images of three type of mangoes | MobileNetv2, ShuffleNet and SqueezeNet. | proposed a versatile system can work RGB as well as thermal image for prediction | Thermal imaging is expensive |
| **O**. M. Lawal et al. [5] | 410 muskmelon images | collected from greenhouse in various province of China and labeled by Github | ResNet43, SPP, FPN, DIoU-NMS and CIoU | Model is robust and fast. | To mentioned accuracy and too much deep network is used. |
| **K**. Zhang et al. [6] | 2000 Strawberry images | Downloaded from Internet and real captured of strawberry | R-YOLO improved YOLOv3 with MobileNet-V1, K-Means | Proposed model was excellent real-time performance and 3.6 time faster comparable to YOLOv3 | Performance is poor for multiple fruits, occlusion and overlap. |
| **N**. Stasenko et al. [7] | 12000 apple images | Captured using Testbed | U-Net and DeepLab | mIoU for Deeplab was remarkable. | Not focus on accuracy and Precision. |
| **Chai** C. Foong et al. [8] | 2100 apples, bananas and oranges | Kaggle | ResNet50 | Taking less time to be trained. | Not always accurate for green apples. |
| **Jiangong** N. et al. [9] | 618 banana images | Author photographed images during 11 days | Transfer learning and GoogleNet | Model is fast and scalable, can be deployed on mobile | Recognition was not correct |
| **S**. Bulla et al. [10] | 5989 apples, banana and oranges images | Kaggle | Transfer learning and CNN | Model is feasible because being used of less computational time and memory space. | More number of epochs are used to get remarkable accuracy. |

**Table 2**

Comparison of Various Authors on the base of Sensitivity, Specificity, F1 Score and Accuracy

| Author | Sensitivity or recall | Specificity | Precision | F1-score | Accuracy |
|---|---|---|---|---|---|
| N. Ismail et al. [1] | 99.6 % for apple 98.9% for banana | 99.2 % for apple 99.1% for banana | NA | NA | 99.5 % for apple 99.2% for banana |
| Sovon et al. [2] | 98.6% | NA | 98.6% | 98%(average) | 99.6% |
| Lili Zhu et al. [3] | 96.8 | NA | 98.5% | 97.6 % | 98.5 % |
| V. Bhole et al. [4] | RGB 97% and Thermal 98.1% | NA | 97.5% and 98.3 % | RGB 97.3% and Thermal 98.2% | RGB 97.1% and Thermal 98.5% |
| O. M. Lawal et al. [5] | 82% | NA | 96.3% | 84% | NA |
| K. ZHANG et al. [6] | 93.5% | NA | 94.4% | 93.9% | 84.3% |
| Jiangong N. et al. [9] | 99.6% | NA | 100% | 99.8 | 98.9% |
| S. Bulla et al. [10] | NA | NA | NA | NA | 97.8% |
| L. Wu et al. 14] | 97.4% | NA | 95.5% | 96.5% | NA |