Meta-Analysis of Machine Learning Methods for Fruit Quality Prediction

Daniyal Nazir [1], Qazi Arbab Ahmed [1,*], Saba Khan [1], Eesha Khan [1], Shafaq Ejaz [2]

[1] Department of Software Engineering, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan.
[2] Department of Electrical Engineering, Mirpur University of Science and Technology (MUST), Mirpur, Pakistan.
[*]Corresponding author: qazi.arbab@ajku.edu.pk

Abstract

Considering the intentions towards diminishing fruit spoilage and the rising need for fruit spoilage detections; many of the models have been suggested which has been possible with the current rise of machine intelligence and computer vision. Nonetheless, the effectiveness of these suggested models is controversial once disclosed to unseen datasets and their adaptability is unsure when faced with diverse fruit types. The benefaction of this paper is to identify the ideal model for classifying defective fruits using a unified dataset and modifications to the existing models for better fruit spoilage detection for real-life implementations. Machine learning models like Support Vector Machine (SVM), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNNs) are examined and fairly trained along with their processing phases, whereas their performances were estimated and analyzed in both binary and multiclass classification issues. Consequently, after the proposed modifications CNN-based models are the perfect solution. The suggested modifications to the CNN architecture improve the classification accuracy, precision, recall, and F1 score metrics. The experiment results show that the CNN1 ¬method outperformed the other four state-of-art compared models in the prediction of fruit quality.

Keywords—Fruit Spoilage, proposed modification, controversial, defective fruits, unified dataset

1 Introduction

There is convincing proof that fruits lead to a sound body so, it is mandatory for people to add fruits and vegetables to their diet as by WHO [1]. Both fruits and vegetables have the maximum decay rate (45%) among all varieties of edible food products which really affects the agricultural industry as reported by the Food Agriculture Organization of the United Nations (FAO) [2]. Fruit spoilage occurred in microbial due to low pH values and consequently, spoilage might lead to the wastage of fruit [3]. Both Customer's perception and marketability about the quality of fruit would revolve around the state of the fruit [4]. The physical state of the affected fruit would influence by a set of biochemical conversions [5]. Customers might gain decayed fruits if they were not managed correctly [6]. Fruit would save by preliminary identification of decayed ones. Boshan et al. [7] suggested a technique for fabric flaw identification that relies on the decay of low-rank and gradient details. Liang et al. [8] suggested a novel technique for fabric standard investigation depending on lattice fragmentation with a statistics template. Noticeable drawbacks in manual investigations are incompatible judgments and overdue. Consequently, it is mandatory to partially alter the physical work with an automated fruit standard investigation that would be quick, compatible, precise, and low-budget [9]. Conventional image grouping techniques such as SVM [10] and K-nearest neighbor (KNN) [11] were applied in classifying types of fruits [12] [13]. These machine learning techniques utilize the physical state of fruits such as skin, appearance, structure, and surface flaws in determining their visible aspects. Recently, deep learning has appeared as one of the favorable techniques in the discipline of image processing [14] 15], [16]. The most accepted deep learning models are Convolutional Neural Networks (CNNs) and hence, D4Net [17] is a network suggested to control flaws identification issues by defective image and its reference image. Bing et al. [18] suggested a novel framework that depends on a visual gain strategy for textile flaws identification. Computer vision takes part merely in the outer examination of fruit, not of the inner part.

ISSN: 2523-0379 (Online), ISSN: 1605-8607 (Print)
DOI: https://doi.org/10.52584/QRJ.2002.17.

Apart from this, non-destructive techniques such as hyperspectral imaging [19] and near-infrared imaging [20] were used in identifying fruit decay. Above all, the most economical technique is a combination of computer vision and image processing. However, the effectiveness of these proposed models is questionable once exposed to unseen datasets. The aim of these analyses is to determine the best technique for fruit decay identification for an unbiased dataset. In this paper, we systematically review and analyze the Fruit Quality Prediction Systems (FQPS) which are based on machine learning and hence encourage enhancing the model's performances and underperforming models such as parameters, number of layers, optimizers, and pre-processing techniques for better outcomes. To the best of our knowledge, this is the first study to conduct a meta-analysis of machine learning methods for fruit quality prediction systems and propose a necessary modification to DNN parameters for improved classification.

The main contributions of this paper are:

- To determine the ideal model for classifying defective fruits using a unified dataset.
- Propose modifications to the existing underperforming models.
- Provide recommendations for a better fruit spoilage detection model for real-life implementations.

The rest of this paper is structured as follows: Section 2 gives perceptions about studies held and the production of the recommended architectures in investigating fruit decay. Section 3 explains the techniques and instruments used in experiments followed by their outcomes which are discussed in Section 4. Analysis of investigation on decayed fruits, their evaluation, comparisons, and suggestions are discussed in Section 5. Section 6 Summarizes the outcomes, limitations, and future suggestions for work in fruit spoilage identification.

2 Literature Review

In this section, most of the research articles on fruit spoilage detection are analyzed and reviewed. With the aid of image processing [21] and computer vision, several approaches have been used by researchers in determining spoilage in fruits with machine learning algorithms.

2.1 Traditional Fruit Classification Methods

In terms of technology and techniques early research was not compatible with the agriculture field. The authors in [22] proposed a grading system of apples depending on defect features having an accuracy of 73% however, there is a complexity in differentiating brown and marked ones. Furthermore, [23] introduced a computer vision-based apple surface defect sorting system that has taken into consideration the entire surface of the apple and distinguished stem and calyx from true defects. Hence, results were inconsistent because of the difficulty to differentiate rotten areas. The SVM model was suggested by [24] for mangoes. Two image segmentation methods, K-Means Clustering, and Fuzzy C-Means (FCM) clustering were also suggested. Similarly, in [12], authors recommended KNN and SVM for detection. Fine KNN has 96.3% and quadratic SVM has 98% accuracy. Besides, [25] compared the performance of three computer vision-based models in classifying and grading apples. SVM, Multi-Layer Perception (MLP), and KNN were trained and tested with a prepared dataset. The SVM obtained the best results with an accuracy of 92.5% and 89.2% for two categories and three quality categories classification, respectively. The only limitation of this research is that the system is designed to detect one type of fruit only. By changing the interest in the fruit type, [26] invented a new image processing technique in determining external olive fruit defects. The technique is operated based on texture analysis and the homogeneity between neighboring pixels in an image. The method is proven to be superior to traditional classification algorithms in aspects of accuracy and speed so, can accurately identify the region of the defect and calculate its area.

2.2 Deep Learning Methods

As deep learning in object detection has importance in the agriculture field so, an early trial to employ the ANNs for image segmentation was proposed in [27] where a computer vision-based system was developed and ANN requires an additional stem/calyx recognition method because it was not completely compatible. A more advanced deep learning model that employed the ANN was developed for sorting the grade of bananas [28] and it was the only fruit type that has been experimented with. Hence, further investigation of the model's performance on other fruit types is required. Many types of research for different fruit types such as raspberry [29], bananas [30], mangosteens [31], sour lemons [32], and apples [33] have developed a CNN algorithm to identify defective fruits and obtained results were quite satisfying. For instance, in identifying raspberry conditions [29], an accuracy of 86% was obtained for correctly predicting the state of the raspberry. Furthermore, CNN was used for the first

Fig. 1: (a) Samples of healthy and rotten images from the experimental dataset. (b) Image (Fruit) classification model phases

time to inspect the ripening stages as well as the quality of bananas in the research proposed by [30]. Similarly, the binary sorting system of mangosteen into fine and the defect was introduced with CNN [31]. In this research, validation of data accuracy was done with the implementation of the 4-fold cross-validation technique to expand the limited data samples. However, as data image was created by authors, some of the images differ in terms of light intensity. Deep CNN was implemented in the defect classification of sour lemons [32]. A stochastic pooling mechanism based on meagreness was used to improve the CNN framework. This improved CNN turned out to have the best results in identifying defects with an accuracy of 100% when compared to other traditional classification methods. A commercial packing line system of apples using computer vision with CNN was proposed in [33]. For instance, [34] compared the performance of pre-trained networks like AlexNet [35], MobileNet [36], GoogLeNet [37], and Xception [38] with self-built two layers of CNN architecture in identifying strawberry quality. Binary class (bad and good class) and four classes (three ranks of good class and one rank of bad class) of strawberry classification were the two tests conducted to identify the accuracy of those architectures. The binary bad class contains rotten, overripe and damaged strawberries while the remaining all lies in the binary good class. The results showed that GoogLeNet architecture has the least computational time and model size while VGGnet is the best in terms of accuracy. Additionally, other researchers proposed a combination of You Only Look Once version 3 (YOLOv3-dense) [39], [40] and Cycle-Consistent Adversarial Network (CycleGAN) [41] in detecting apple spoilage. In all the methods described above, each has achieved satisfactory results in fruit spoilage detection. However, those models have only been trained with their own datasets. There is no guarantee that the model will perform well as it is supposed to when it is exposed to a new dataset. The model will struggle to adapt to a real-life scenario if it cannot perform well on other datasets. Hence, in this paper, several recently proposed methods will be selected as models to be trained and tested with a newly unified dataset. The performance of the proposed methods which include traditional classification algorithms and deep learning algorithms will be evaluated and analyzed. A comparison of strengths and weaknesses among the models will be discussed. The objective is to determine which model is proven to be effective in classifying variation of fruit defect images. The versatility of the models is also being tested in this work as the models will be trained with three different types of fruits. The introduction of a versatile and effective image processing model will be beneficial to the agriculture field and increases the chances of the actual implementation of an ideal model with reasonable speed and accuracy. Moreover, the deployment of a computer vision-based image processing machine would decrease the need for time and cost-consuming physical labor with subjective judgment on the fruit condition. Lastly and most importantly, chances of fresh fruit being considered as spoilt fruit will be greatly reduced and decrease the fruit wastage rates.

3 Dataset

This section highlights the extensive experimentation carried out in the field of computer vision and image processing and the datasets used in these studies. The research findings indicate that numerous diverse architectures and image-processing techniques have been developed to attain optimal results.

3.1 Dataset Description

The selected dataset for models' evaluation in this paper is the Fruits Fresh and Rotten from Kaggle [42]. The dataset consists of three types of fruits: apple, orange, and banana. The states of the fruits have been labeled into two classes: fresh and rotten, as shown in Table 1. Some of the sample images from the dataset are shown in Figure 1(a). The sample images contain some common defects such as russet, rot, marks, and burns. Making up to six different classes of fresh fruit and rotten fruit. The images were captured by a camera on a clear white background. The images were taken under the same settings such as distance and illumination. Each image consists of only a single fruit. As the images are screen-shots of fruits, the resolution of the images varies from around 200×200 pixels to 400×400 pixels. This has totaled up the augmented dataset to have 13599 images of fruits across the six classes. The dataset was separated into 80% for training and 20% for testing. This dataset is unexposed previously to all selected architectures. The availability of three different fruit types in fresh and rotten states has made this dataset ideal to train and test the efficiency of the architectures. To justify the performance of each proposed architecture, every step including processing, augmentation, segmentation, feature extraction, and classification will be emulated as described by each work.

3.2 Selected Models for Experiments

In this section, a total of five different proposed models have been discussed and selected based on compatible results. For the sake of easy representation, the models have been given nicknames according to the employed architectures and arranged in the following order: The SVM model proposed in [24], ANN by [28], CNN1 by [29], CNN2 by [32], and CNN3 by [33]. A typical image classification model for fruit quality inspection would comprise the following steps as shown in Figure 1 (b). Table 2 summarizes all stages and the employed techniques in each model for easier visualization.

4 Experimental setup

In this section, we will briefly discuss the experimental setup and the detailed analysis of the machine learning-based FQPS.

4.1 Research Instruments

We have used Google Colaboratory (Colab) as the training and testing platform for all models. This is to ensure fairness to all models in terms of processing power and speed. The training processes are carried out using NVIDIA TESLA K80 with an Intel (R)Xeon(R) processor, 12GB RAM, and 358 GB of disk space offered by Colab. The dataset was uploaded to Google Drive and serves as input data for all the models.

4.2 Methodology

In this section, we evaluate the results of various fruit defect classification models using a unified dataset as mentioned in section 3. The processing steps of images and the model structure are followed and built based on the provided specifications in each work. Each proposed method has been tested with fruit images of apples, bananas, and oranges. Confusion matrices for each experiment are provided where each matrix evaluates the actual output versus the predicted output by providing the number of correct and incorrect predictions [43]. The matrix is summarized with True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values to show the performance of the classification model. In this work, TP indicates the correct prediction of a healthy class, TN refers to the correct predicted defective class, FP refers to the wrong indication of a predicted healthy class and FN predicts a healthy class wrongly as a defective class. Moreover, the accuracy, F1 score, precision, and recall values of each classification model are introduced in each table. Accuracy is determined by the number of correctly predicted images over the total number of the testing set and it is shown in Equation 1.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

Precision and recall are often used together in model evaluation metrics and their values are always opposed to each other. Precision refers to the accuracy while recall measures the percentage of correct actual positive predictions over all the positive classes. Equations 2 and 3 show the formulas for precision and recall, respectively.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{2}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3}$$

Table 1: Fruits Fresh and Rotten Dataset description

	Without Augmentation		With Augmentation	
Label	Number of training images	Number of test images	Number of training images	Number of test images
Fresh Apples	182	50	1693	395
Rotten Apples	253	74	2342	601
Fresh Banana	169	49	1581	381
Rotten Banana	238	68	2224	530
Fresh Oranges	163	43	1466	388
Rotten Oranges	177	45	1595	403

F1 score is a combined measure between precision and recall and can be obtained by computing the harmonic mean. The formula of the F1 score is shown in Equation 4

$$F1\ Score = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \tag{4}$$

It is worth mentioning that some methods did not provide information on certain experimental specifications such as the image size, epochs, etc. Hence, the constant values that were used throughout the experiments are $image\ size = 200 \times 200\ pixels$, $no.\ of\ epochs = 50$, $batch\ size = 64$, *optimizer= Stochastic Gradient Descent (SGD)*, and $lossfunction = categorical\_crossentropy$.

4.3 Performance Analysis with Data Augmentation

4.3.1 Apples

The augmented apple dataset consists of 4035 training images and 996 test images. Table 3 shows the classification results of the 395 healthy and 601 defect test images of apples by the competing models. CNN1 demonstrated a good performance in identifying apple defects. with an accuracy of 97.09% as compared to all other methods.

4.3.2 Bananas

Bananas contain 3805 training images and 911 images that were used for the testing process. Table 4 shows the detailed results of classifying 381 healthy bananas and 530 rotten bananas by the models and CNN1 obtained the highest classification accuracy of 99.89%.

4.3.3 Oranges

The augmented orange dataset contains 3061 images for training and 791 images for testing. Out of 791 test images, 388 are fresh oranges and 403 are rotten oranges. From Table 5, it is observed that CNN1 performs better than the other models with an overall accuracy of 94.56%.

4.3.4 All fruit types (2 classes)

The augmented dataset of all fruit types consists of 10901 training images and 2697 test images. Test images were divided into 1163 fresh fruit classes and 1534 rotten fruit classes. The CNN1 model obtained excellent results in this phase with an accuracy of 96.85% by misclassifying 28 healthy fruits and 57 rotten fruits as shown in Table 6.

4.3.5 All fruit types (6 classes)

The same augmented dataset of all fruits used in Section 4.1.4 is used in this section too, to evaluate the model's performance. Table 7 displays the classification results for all methods. Test images were divided into 395 fresh apples, 381 fresh bananas, 388 fresh oranges, 601 rotten apples, 530 rotten bananas, and 403 rotten oranges and the CNN1 model obtain a total classification accuracy of 95.40%.

4.4 Performance Analysis without Data Augmentation

4.4.1 Apples

The dataset without augmentation contains 435 training images and 124 test images of apples. 50 out of 124 test images are labeled as healthy and the rest are defective. CNN1 obtained the highest accuracy of 91.94% as compared to all other methods as shown in Table 3.

4.4.2 Banana

The models were trained with 407 images and tested with 177 images which were made up of 49 fresh bananas and 68 rotten bananas. CNN1 leads to a high accuracy of 99.15% with an accuracy of 51.28% and 58.12% for CNN2 and CNN3 respectively as shown in Table 4.

Table 2: Summary of all stages of the proposed methods

<table>
<tr><th>Methods</th><th>Ref.</th><th>Method</th><th>Pre-processing</th><th>Augmentation</th><th>Segmentation</th><th>Feature Extraction</th><th>Classification</th></tr>
<tr><td>Traditional Machine Learning</td><td>[24]</td><td>SVM</td><td>Median and mean filter, Histogram equalization</td><td>-</td><td>Image binarization, FCM, and K-means with 3 clusters</td><td>Color</td><td>SVM</td></tr>
<tr><td rowspan="4">Deep Learning</td><td>[28]</td><td>ANN</td><td>Weighted/luminosity method, Median filter
Background removal by thresholding
Resized to 32 × 32</td><td>Rotation by 0°, 90°, 180°</td><td>-</td><td>Edge detection, Morphological operation</td><td>ANN<table><tr><td>Layer</td><td>No of Neurons</td></tr><tr><td>Input</td><td>1024</td></tr><tr><td>Hidden</td><td>40</td></tr><tr><td>Output</td><td>2</td></tr></table></td></tr>
<tr><td>[29]</td><td>CNN1</td><td>Background removal by thresholding</td><td>Augmented by dividing a large image into smaller parts</td><td>Segmented into multiple smaller images</td><td colspan="2">CNN<table><tr><th>Layer</th><th>No of Filter</th><th>Filter Size</th><th>Stride</th></tr><tr><td>Conv1</td><td>16</td><td>20×20</td><td>4×4</td></tr><tr><td>Pool1</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv2</td><td>32</td><td>15×15</td><td>4×4</td></tr><tr><td>Pool2</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv3</td><td>64</td><td>6×6</td><td>4×4</td></tr><tr><td>Pool3</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>FC1</td><td>32</td><td>-</td><td>-</td></tr></table></td></tr>
<tr><td>[32]</td><td>CNN2</td><td>Background removal
Resized to 32 × 32</td><td>Rotation by 45°
Horizontal and vertical mirroring</td><td>-</td><td colspan="2">CNN<table><tr><th>Layer</th><th>No of Filter</th><th>Filter Size</th><th>Stride</th></tr><tr><td>Conv1</td><td>40</td><td>5×5</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>Pool1</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv2</td><td>50</td><td>4×4</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>Pool2</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>ReLU</td><td>-</td><td>-</td><td>-</td></tr><tr><td>Conv3</td><td>500</td><td>4×4</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>Pool3</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>ReLU</td><td>-</td><td>-</td><td>-</td></tr><tr><td>Conv4</td><td>300</td><td>1×1</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>FC1</td><td>15</td><td>-</td><td>-</td></tr><tr><td>Softmax</td><td>-</td><td>4×4</td><td>-</td></tr></table></td></tr>
<tr><td>[33]</td><td>CNN3</td><td>Background removal by Otsu's thresholding
Resized to 80 × 80</td><td>Rotation by 90°, 180°, 270°
Brightness ±0.1
Hue ±0.05</td><td>N/A</td><td colspan="2">CNN<table><tr><th>Layer</th><th>No of Filter</th><th>Filter Size</th><th>Stride</th></tr><tr><td>Conv1</td><td>16</td><td>3×3</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>ELU</td><td></td><td></td><td></td></tr><tr><td>Pool1</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv2</td><td>64</td><td>3×3</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>ELU</td><td></td><td></td><td></td></tr><tr><td>Pool2</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv3</td><td>16</td><td>3×3</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>ELU</td><td></td><td></td><td></td></tr><tr><td>Pool3</td><td>-</td><td>-</td><td>2×2</td></tr><tr><td>Conv4</td><td>2</td><td>1×1</td><td>1×1</td></tr><tr><td>BatchNorm</td><td>-</td><td>-</td><td>-</td></tr><tr><td>ELU</td><td></td><td></td><td></td></tr><tr><td>GAP</td><td>15</td><td>-</td><td>-</td></tr><tr><td>Softmax</td><td>-</td><td>4×4</td><td>-</td></tr></table></td></tr>
</table>

Table 3: The obtained results by the competing models on 2 classes of apple fruit with and without data augmentation

	With data augmentation/without data augmentation					
The Model	**Confusion Matrix**		**Precision**	**Recall**	**F1 Score**	**Overall Accuracy**
SVM	264/30	131/20	0.775/0.750	0.752/0.811	0.764/0.779	71.89/72.58%
	149/14	452/60				
ANN	302/31	93/19	0.815/0.743	0.812/0.729	0.814/0.733	82.23/75.00%
	84/12	517/62				
CNN1	386/42	9/8	0.968/0.927	0.972/0.906	0.970/0.914	97.09/91.94%
	20/2	581/72				
CNN2	20/0	375/50	0.808/0.298	0.525/0.500	0.429/0.374	62.35/59.68%
	0/0	601/74				
CNN3	379/0	16/50	0.602/0.298	0.530/0.500	0.378/0.374	44.18/59.68%
	540/0	61/74				

Table 4: The obtained results by the competing models on 2 classes of banana fruit with and without data augmentation

	With data augmentation/without data augmentation					
The Model	**Confusion Matrix**		**Precision**	**Recall**	**F1 Score**	**Overall Accuracy**
SVM	230/27	151/22	0.713/0.707	0.708/0.779	0.710/0.741	66.41/68.28%
	155/15	375/53				
ANN	312/37	69/12	0.891/0.872	0.878/0.848	0.883/0.856	88.80/86.32%
	33/4	497/64				
CNN1	380/48	1/1	0.999/0.993	0.999/0.990	0.999/0.991	99.89/99.15%
	0/0	530/68				
CNN2	0/49	381/0	0.291/0.731	0.500/0.581	0.368/0.455	58.18/51.28%
	0/57	530/11				
CNN3	0/0	381/49	0.291/0.291	0.500/0.500	0.368/0.368	58.18/58.12%
	0/0	530/68				

4.4.3 Oranges

The orange images without augmentation contain 340 training images and 88 test images. Table 5 shows the classification results for 43 oranges in fresh condition and 45 oranges in rotten condition. CNN1 and SVM achieved acceptable results with an accuracy of 82.95% and 76.14% respectively.

4.4.4 All fruit types (6 classes)

The dataset of all three fruit images consists of 1182 training images and 329 test images. Test images were divided into 50 fresh apples, 49 fresh bananas, 43 fresh oranges, 74 rotten apples, 68 rotten bananas, and 45 rotten oranges and here CNN1 has an accuracy of 86.63%. Table 7 presents detailed results.

5 Performance Analysis

This section focuses on the significant results obtained from the classification experiments. The efficacy of each method is analyzed, and its advantages and limitations are discussed in the subsequent sections.

5.1 Performance Analysis of Machine Learning Models

This section delves into the outcomes of the fruit defect identification task using classification techniques. The

Table 5: The obtained results by the competing models on 2 classes of orange fruit defect detection with data augmentation and without data augmentation

	With data augmentation/without data augmentation					
The Model	Confusion Matrix		Precision	Recall	F1 Score	Overall Accuracy
SVM	255/30	133/13	0.678/0.740	0.695/0.822	0.686/0.779	67.64/76.14%
	123/8	280/37				
ANN	294/27	94/16	0.729/0.581	0.729/0.581	0.728/0.579	72.82/57.95%
	121/21	282/24				
CNN1	374/37	14/6	0.946/0.831	0.946/0.830	0.946/0.830	94.56/82.95%
	29/9	374/36				
CNN2	388/43	0/0	0.245/0.244	0.500/0.500	0.329/0.328	49.05/48.86%
	403/45	0/0				
CNN3	174/0	214/43	0.385/0.255	0.387/0.500	0.384/0.338	38.56/51.14%
	272/0	131/45				

Table 6: The obtained results by the competing models on 2 classes of all fruit defect detection with and without data augmentatior

	With data augmentation/without data augmentation					
The Model	Confusion Matrix		Precision	Recall	F1 Score	Overall Accuracy
SVM	597/78	566/64	0.628/0.636	0.622/0.633	0.623/0.634	63.70/64.44%
	413/53	1121/134				
ANN	812/88	351/54	0.777/0.712	0.769/0.706	0.772/0.708	77.86/71.73%
	246/39	1288/148				
CNN1	1135/108	28/34	0.967/0.819	0.969/0.813	0.968/0.816	96.85/82.07%
	57/25	1477/162				
CNN2	1163/9	0/133	0.725/0.792	0.535/0.532	0.376/0.428	47.13/48.86%
	1426/0	108/187				
CNN3	960/112	203/30	0.395/0.437	0.454/0.464	0.340/0.377	40.30/41.95%
	1407/161	127/26				

Table 7: The obtained results by the competing models on 6 classes of all fruit defect detection with and without data augmentation

With data augmentation /without data augmentation				
The Model	Precision	Recall	F1 Score	Overall Accuracy
SVM	0.506/0.520	0.507/0.505	0.504/0.508	51.74/52.28%
ANN	0.627/0.410	0.626/0.409	0.625/0.406	63.38/42.86%
CNN1	0.960/0.867	0.951/0.859	0.955/0.862	95.40/86.63%
CNN2	0.620/0.082	0.361/0.231	0.353/0.105	35.47/18.84%
CNN3	0.068/0.071	0.179/0.185	0.097/0.087	18.31/23.10%

accuracy, precision, recall, and F1 score of each tested dataset were computed and used as the metrics for model performance evaluation. The performance of all methods is analyzed and described below.

5.1.1 Classification Results in Identifying Only One Fruit Type

Each image used in this experiment contains only one type of fruit among apples, bananas, and oranges and there are two states for each type of fruit, fresh or rotten. The comparison of the various methods in identifying individual fruit defects is shown in Table 8. The results show that the banana is the most identifiable fruit while the orange is the least accurately predicted one. CNN1 method is the best having an accuracy of 91.35%. CNN architecture is made up of only nine layers, but each layer performs well in its operation with the help of ReLU and softmax activation functions. 750×750 sized images were used as the input to the neuron layer to help the model to capture and sustain valuable information about the image. The disadvantage of this model is that it requires higher computational time and power than the other methods as it is trained by high-pixel images with 50 epochs. Except for the low accuracy (57.95%) on orange defect identification, ANN performed well with apples and bananas and not for oranges. For apple and banana defect classification, there were also slight differences in accuracy. Furthermore, the SVM method portrays rather consistent results in the classification of these defective fruits with an average accuracy of 72.33%. SVM has utilized the ability to distinguish defective and healthy fruit but it has average accuracy.

CNN2 can be identified as an overfitted model while training the dataset. For instance, the training accuracy and validation accuracy of the CNN2 were in the range of 90% and 80%, respectively, while the test accuracy was merely around 50%. Overfitting has caused the CNN2 unable to generalize well on unseen data which leads to poor test accuracy results. The reason behind the overfitting of this model may be due to the complexity of the model and the performance of CNN2 was unsatisfactory. The final training accuracy of CNN3 was 68.29% and 41.86% for validation accuracy in apple defect classification. 59.86% was the test accuracy of apple for CNN3. These results were considered low for both the training set and the testing set. CNN1 is the most ideal method for determining fruit defects. ANN and CNN2 methods were overfitted while CNN3 was under fitted.

5.1.2 Classification Results in Identifying All Fruit Types

Table 9 shows the results of 2 classes of classification and 6 classes of classification for all types of fruits. For 2 classes classification, the models only target identifying the fresh or rotten fruits out of all classes of fruits. For 6 classes classification, the models aim to predict the 6 classes of fruits into their belonging class. CNN1 method costs a lot of time and the ANN method came up second in terms of accuracy. The performance of SVM is better in predicting defective fruit images with the features extracted from all types of fruits. CNN2 was determined to be overfitted in this binary classification task. CNN2 correctly predicted all defective fruits but only managed to determine 9 out of 142 fresh fruits. Lastly, CNN3 has an unsatisfactory result among the five methods. Theoretically, CNN3 is under-fitted in this experiment. In 6 classes classification, the performance from all methods was not good at all except for CNN1. Although CNN1 achieved an above-average classification result, it still misclassified a small portion of data, especially in the rotten orange class. The performance of SVM was poor in this experiment. SVM barely estimated half of the classes correctly with a precision rate of 0.520. In this experiment, ANN suffers from the overfitting problem as the training accuracy was nearly 99% at the end of the 3000 epochs, but the test accuracy was merely just 42.86%. Also, overfitting is dominating the CNN2 method again as the model was extracting a high number of features from all the fruit images. With poor training, validation, and testing accuracies, CNN3 was identified to be under-fitted in the 6 classes classification task as well. These methods were unable to recognize the feature of fruit type and did not generalize well on the unseen data which causes the models to only be capable of predicting two classes out of six classes.

5.2 The Effect of Data Augmentation on Models' Performance

In this section, the effect of data augmentation on various methods is evaluated as shown in Table 10. Since two out of five methods were overfitted with the dataset without augmentation, therefore all methods were trained with augmented data to prevent overfitting in the models. The data augmentation techniques applied to the dataset were rotation, translation, vertical flip, and the addition of salt and pepper noises. This has increased the data size by at least eight times the original dataset. SVM is not suitable to be trained with a large number of samples. ANN and CNN1

Table 8: Comparison of all individual fruit type classification accuracy by the competing models (without augmentation)

Fruit type	SVM	ANN	CNN1	CNN2	CNN3	Average per fruit
Apple	72.58%	75.00%	91.94%	59.68%	59.68%	71.78%
Banana	68.28%	86.32%	99.15%	51.28%	58.12%	72.63%
Orange	76.14%	57.95%	82.95%	48.86%	51.14%	63.41%
Average per model	72.33%	73.09%	91.35%	53.27%	56.31%	-

Table 9: Comparison of all fruit type classifications (without augmentation)

Classification type	SVM	ANN	CNN1	CNN2	CNN3
All fruits (2 classes)	64.44%	71.73%	82.07%	48.86%	41.95%
All fruits (6 classes)	52.28%	42.86%	86.63%	18.84%	23.10%

methods performances were enhanced after training with the augmented dataset. The most noticeable increment of accuracy was the orange dataset with an increase of 14.87% and 11.61% for ANN and CNN1, respectively. After augmentation was applied to the orange dataset, the training images increased to 3061 which is sufficient to train models better. In the first experiment, ANN overfitted with the unaugmented dataset because of the small number of training data. The overfitting was solved by data augmentation. Although data augmentation has been applied to the dataset, the training of CNN2 with the augmented dataset did not significantly improve the classification results. A slight increase in accuracy can be seen in the apple, banana, and orange individual datasets. However, the model remained overfitted. Similarly, the performance of CNN3 did not improve even after data augmentation was implemented. As shown in Table 10, most of the datasets with augmentation applied performed better than those without augmentation in the classification of defective fruit for overfitted cases. While data augmentation is ineffective in applying to SVM or under-fitted models.

5.3 The Proposed Modifications to Improve the Models' Performance

The performance for fruit defect classification was not good enough to be implemented in real life. Hence, a few modifications have been applied to each model architecture to enhance the performance of classification. To overcome the uncertainty in model's versatility when faced with a unified dataset of different types some adjustments such as adding layers, removing layers, and tuning the parameters were made to each model to increase its detection accuracy. There are no adjustments made for the SVM model. As the ANN model was proven to be overfitted in the first experiment, adjustments have been made to prevent overfitting. Modifications have been done to reduce the overfitting of the model. The ReLU activation function was added to the input layer. The sigmoid activation function in the output layer was replaced by the softmax activation function. Since the training accuracy and loss were almost constant after 50 epochs, hence the 3000 epochs used in ANN are reduced to 100 epochs to reduce the computational time and power. This is called early stopping which is also one of the ways to stop overfitting. The performance of ANN in classification gained a slight increase in accuracy after the modifications. The results can be observed in Table 11.

As the performance of CNN1 in all categories of classification was satisfactory. Even though changes have been applied to the architecture, the results obtained were not better than the initial architecture. There is no alternative way to further improve the method as the suggested architecture was the most accurate method after all. Since the CNN2 model has been overfitted in most of the classification cases, the performance of this model was poor. Implementation of batch normalization in this model does not work well with the dataset. The introduction of batch normalization in the network was to speed up the training process and reduce the sensitivity of the CNN. However, the performance of the testing set did not improve with the existence of batch normalization layers. Therefore, all batch normalization layers were removed from the architecture. Besides, a dropout layer with a value of 0.4 was added before the fully connected layer. These actions were done to avoid overfitting from happening in the model. After the modifications, the results were significantly better and the model is no longer overfitted with the dataset. Table 11 shows

Table 10: Comparison between the classification accuracy of the augmented dataset and unaugmented dataset of all models

Dataset type	SVM	ANN	CNN1	CNN2	CNN3
Apple (augmented)	71.89%	82.23%	97.09%	62.35%	44.18%
Apple (unaugmented)	72.58%	75.00%	91.94%	59.68%	59.68%
Banana (augmented)	66.41%	88.80%	99.89%	58.18%	58.18%
Banana (unaugmented)	68.28%	86.32%	99.15%	51.28%	58.12%
Orange (augmented)	67.64%	72.82%	94.56%	49.05%	38.56%
Orange (unaugmented)	76.14%	57.95%	82.95%	48.86%	51.14%
All types (2 classes) (augmented)	63.70%	77.86%	96.85%	47.13%	40.30%
All types (2 classes) (unaugmented)	64.44%	71.73%	82.07%	48.86%	41.95%
All types (6 classes) (augmented)	51.74%	63.38%	95.40%	35.47%	18.31%
All types (6 classes) (unaugmented)	52.28%	42.86%	86.63%	18.84%	23.10%

the comparison results of the modification of CNN2.

CNN3 was the only under-fitted model in the experiment. Normally an under-fitted model is lacking complexity in its architecture, and so the number of parameters used in the model is increased. The kernel size of the last convolution layer in CNN3 has been increased from 2 to 16. As one of the data augmentation techniques, brightness adjustment used in the proposed work was limiting the training and validation accuracy of the experiment, therefore this implementation was removed. The inclusion of Nesterov's moment in the optimizer did not improve the accuracy of detection, hence it was not used in optimizing the model's parameters. As the ELU activation function was preventing the model from obtaining good test results, all ELU activation function layers were removed. In the case of all fruit, types defect detection, both 2 and 6 class classifications were found to be overfitted after the amendments being made to the architecture. To solve the overfitting issues, a dropout layer with a value of 0.5 was added before the GAP layer in the architecture. These modifications have notably enhanced the performance of CNN3 in fruit spoilage detection as shown in Table 11.

6 Conclusion

This work provides a systematic review of the design and analysis of machine learning methods used for the Fruit Quality Prediction System. The aim of this research has been achieved in such a way that the CNN1 method has been declared as the optimal solution in the classification of fruit spoilage against any fruit dataset from all other state-of-the-art machine learning methods. This method has an accuracy of above 80% in detecting fruit defects with the aid of computer vision. The results were exceptionally inappropriate when defects in three different fruit types were determined. Then classification was carried out in two parts; data with augmentation and without augmentation and this leads to better classification results as some modifications were also performed. L2 and dropout regularization were implemented in the models to solve the overfitting issues. While for the under-fitted model, the number of parameters used is increased. Through this research, we conclude that not every model is suitable to be implied at every dataset as the complexity of the particular model may be too high or low for the given data, hence the model may not be well trained. The limitation of this work is the quality of the fruit images and the size of the dataset. For future work, capturing fruit images at various angles is recommended as it can cover up the majority of the surface areas.

References

[1] World Health Organization. https://www.who.int/news-room/fact-sheets/detail/healthy-diet (Accessed on 10.2022.)

[2] D.Ewing-Chow. "Forbes." https://www.forbes.com/sites/daphneewingchow/2020/10/ 31/fruit-and-vegetable-spoilage-is-a-hidden-cause-of-underdevelopment/?sh=67c3bffc6923 (accessed on 10.2022)

[3] Brackett, R. E., and D. F. Splittstoesser. "Fruits and vegetables." Compendium of methods for the microbiological examination of foods 4 (2001): 515-520.

[4] Momeny, Mohammad, Ahmad Jahanbakhshi, Khalegh Jafarnezhad, and Yu-Dong Zhang. "Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach." Postharvest Biology and Technology 166 (2020): 111204.

[5] Fu, Yuhang. "Fruit freshness grading using deep learning." PhD diss., Auckland University of Technology, 2020.

[6] Eaton, Madison J., Shanice Edwards, Harrison A. Inocencio, Franklin J. Machado, Etta M. Nuckles, Mark Farman, Nicole A. Gauthier, and Lisa J. Vaillancourt. "Diversity and cross-infection potential of Colletotrichum causing fruit rots in mixed-fruit orchards in Kentucky." Plant Disease 105, no. 4 (2021): 1115-1128.

[7] Shi, Boshan, Jiuzhen Liang, Lan Di, Chen Chen, and Zhenjie Hou. "Fabric defect detection via low-rank decomposition with gradient information and structured graph algorithm." Information Sciences 546 (2021): 608-626.

[8] Jia, Liang, Chen Chen, Shoukun Xu, and Ju Shen. "Fabric defect inspection based on lattice segmentation and template statistics." Information Sciences 512 (2020): 964-984.

[9] Donis-González, Irwin R., Daniel E. Guyer, Gabriel A. Leiva-Valenzuela, and James Burns. "Assessment of chestnut (Castanea spp.) slice quality using color images." Journal of Food Engineering 115, no. 3 (2013): 407-414.

Table 11: ANN, CNN2, and CNN3 classification accuracy results on all fruit classification problems before and after the proposed modifications with augmentation

	ANN		CNN2		CNN3	
	Before Modification	After Modification	Before Modification	After Modification	Before Modification	After Modification
Apple	82.23%	84.23%	62.35%	94.18%	44.18%	91.77%
Banana	88.80%	89.46%	58.18%	98.24%	58.18%	98.79%
Orange	72.82%	78.89%	49.05%	88.24%	38.56%	91.78%
All type 2 classes	77.86%	80.35%	47.13%	85.47%	40.30%	88.84%
All type 6 classes	63.38%	67.75%	35.47%	88.47%	18.31%	51.33%

[10] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20 (1995): 273-297.

[11] MacQueen, J. "Some methods for classification and analysis of multivariate observations." In Proc. 5th Berkeley Symposium on Math., Stat., and Prob, p. 281. 1965.

[12] Nosseir, Ann, and Seif Eldin Ashraf Ahmed. "Automatic Classification for Fruits' Types and Identification of Rotten Ones using k-NN and SVM." International Journal of Online Biomedical Engineering 15, no. 3 (2019).

[13] Ilyas, Muhammad, Sami Ur Rahman, Muhammad Waqas, and Fakhre Alam. "A robust algorithm for fruits recognition system." Transylvanian Review 26, no. 32 (2018).

[14] Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." Computers and electronics in agriculture 147 (2018): 70-90.

[15] Saxe, Andrew, Stephanie Nelli, and Christopher Summerfield. "If deep learning is the answer, what is the question?." Nature Reviews Neuroscience 22, no. 1 (2021): 55-67.

[16] Wang, Hongming, Ryszard Czerminski, and Andrew C. Jamieson. "Neural networks and deep learning." In The machine age of customer insight. Emerald Publishing Limited, 2021.

[17] Xu, Xuemiao, Jiaxing Chen, Huaidong Zhang, and Wing WY Ng. "D4Net: De-deformation defect detection network for non-rigid products with large patterns." Information Sciences 547 (2021): 763-776.

[18] Wei, Bing, Kuangrong Hao, Lei Gao, and Xue-song Tang. "Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism." Information Sciences 541 (2020): 60-74.

[19] Chen, Yushi, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks." IEEE Transactions on Geoscience and Remote Sensing 54, no. 10 (2016): 6232-6251.

[20] Shao, Wenhao, Yanjie Li, Songfeng Diao, Jingmin Jiang, and Ruxiang Dong. "Rapid classification of Chinese quince (Chaenomeles speciosa Nakai) fruit provenance by near-infrared spectroscopy and multivariate calibration." Analytical and Bioanalytical Chemistry 409 (2017): 115-120.

[21] Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. "Machine learning in agriculture: A review." Sensors 18, no. 8 (2018): 2674.

[22] Leemans, Vincent, and M-F. Destain. "A real-time grading method of apples based on features extracted from defects." Journal of food engineering 61, no. 1 (2004): 83-89.

[23] Li, Qingzhong, Maohua Wang, and Weikang Gu. "Computer vision based system for apple surface defect detection." Computers and electronics in agriculture 36, no. 2-3 (2002): 215-223.

[24] Kumari, Neeraj, A. K. Bhatt, Rakesh Kr Dwivedi, and Rajendra Belwal. "Performance analysis of support vector machine in defective and non defective mangoes classification." International Journal of Engineering and Advanced Technology (IJEAT) 8, no. 4 (2019): 1563-1572.

[25] Moallem, Payman, Alireza Serajoddin, and Hossein Pourghassem. "Computer vision-based apple grading for golden delicious apples based on surface features." Information processing in agriculture 4, no. 1 (2017): 33-40.

[26] Hussain Hassan, Nashat M., and Ahmed A. Nashat. "New effective techniques for automatic detection and classification of external olive fruits defects based on image processing techniques." Multidimensional Systems and Signal Processing 30 (2019): 571-589.

[27] Unay, Devrim, and Bernard Gosselin. "Artificial neural network-based segmentation and apple grading by machine vision." In IEEE International Conference on Image Processing 2005, vol. 2, pp. II-630. IEEE, 2005.

[28] Olaniyi, Ebenezer Obaloluwa, Oyebade Kayode Oyedotun, and Khashman Adnan. "Intelligent grading system for banana fruit using neural network arbitration." Journal of Food Process Engineering 40, no. 1 (2017): e12335.

[29] Prakash, Karthik Kuchangi Jothi. "Spoilage Detection in Raspberry Fruit Based on Spectral Imaging Using Convolutional Neural Networks." DOI 10 (2018): D7D23R.

[30] Zhang, Yan, Jian Lian, Mingqu Fan, and Yuanjie Zheng. "Deep indicator for fine-grained classification of banana's ripening stages." EURASIP Journal on Image and Video Processing 2018, no. 1 (2018): 1-10.

[31] Riyadi, Slamet, Laila Ma'rifatul Azizah, Cahya Damarjati, and Tony Khristanto Hariadi. "Evaluation of mangosteen surface quality using discrete curvelet transform." In 2018 International Conference on Information and Communication Technology Convergence (ICTC), pp. 475-479. IEEE, 2018.

[32] Jahanbakhshi, Ahmad, Mohammad Momeny, Majid Mahmoudi, and Yu-Dong Zhang. "Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks." Scientia Horticulturae 263 (2020): 109133.

[33] Fan, Shuxiang, Jiangbo Li, Yunhe Zhang, Xi Tian, Qingyan Wang, Xin He, Chi Zhang, and Wenqian Huang. "On line detection of defective apples using computer vision system combined with deep learning methods." Journal of Food Engineering 286 (2020): 110102.

[34] Sustika, Rika, Agus Subekti, Hilman F. Pardede, Endang Suryawati, Oka Mahendra, and Sandra Yuwana. "Evaluation of deep convolutional neural network architectures for strawberry quality inspection." Int. J. Eng. Technol 7, no. 4 (2018): 75-80.

[35] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural net-

works." Communications of the ACM 60, no. 6 (2017): 84-90.

[36] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[37] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015.

[38] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251-1258. 2017.

[39] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[40] Tian, Yunong, Guodong Yang, Zhe Wang, En Li, and Zize Liang. "Detection of apple lesions in orchards based on deep learning methods of cyclegan and yolov3-dense." Journal of Sensors 2019 (2019).

[41] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. 2017.

[42] S. R. Kalluri. "Fruits fresh and rotten for classification, Apples Oranges Bananas." https://www.kaggle.com/sriramr/fruits-fresh-and-rotten-for-classification (accessed 10.2022)

[43] Brownlee, J. "What Is A Confusion Matrix In Machine Learning.[online] Machine Learning Mastery."