

# Fruit Deformity Classification through Single-Input and Multi-Input Architectures based on CNN Models using Real and Synthetic Images

Tommy D. Beltran<sup>1</sup>, Raul J. Villao<sup>1</sup>, Luis E. Chuquimarca<sup>1,2</sup>, Boris X. Vintimilla<sup>1</sup>, and Sergio A. Velastin<sup>3,4</sup>

<sup>1</sup>ESPOL Polytechnic University, ESPOL, CIDIS, Guayaquil, Ecuador

<sup>2</sup>UPSE Santa Elena Peninsula State University, UPSE, FACSISTEL, La Libertad, Ecuador

<sup>3</sup>Queen Mary University of London, School of EECS, London, UK

<sup>4</sup>University Carlos III, Dept. of Computer Science, Madrid, Spain

{tomdbelt, rjvillao, lchuquim, boris.vintimilla}@espol.edu.ec, sergio.velastin@ieee.org

**Abstract.** The present study focuses on detecting the degree of deformity in fruits such as apples, mangoes, and strawberries during the process of inspecting their external quality, employing Single-Input and Multi-Input architectures based on convolutional neural network (CNN) models using sets of real and synthetic images. The datasets are segmented using the Segment Anything Model (SAM), which provides the silhouette of the fruits. Regarding the single-input architecture, the evaluation of the CNN models is performed only with real images, but a methodology is proposed to improve these results using a pre-trained model with synthetic images. In the Multi-Input architecture, branches with RGB images and fruit silhouettes are implemented as inputs for evaluating CNN models such as VGG16, MobileNetV2, and CIDIS. However, the results revealed that the Multi-Input architecture with the MobileNetV2 model was the most effective in identifying deformities in the fruits, achieving accuracies of 90%, 94%, and 92% for apples, mangoes, and strawberries, respectively. In conclusion, the Multi-Input architecture with the MobileNetV2 model is the most accurate for classifying levels of deformity in fruits.

**Keywords:** Fruit · Deformity · Multi-Input · CNN models · Real and Synthetic data.

## 1 Introduction

One of the most critical stages in the post-harvest process of fruits and vegetables is the selection and classification through quality inspection (parameters of ripeness, deformation, and defects) [1]. This process, typically meticulous and mechanized, is usually carried out manually by specialized personnel in the field. The significance of this stage lies in the commercial value attributed to the product once evaluated, as higher-quality products tend to command higher selling prices in the market. Additionally, both small-scale retailers and supermarkets, driven by their commercial needs, also consider various parameters when purchasing these foods [2]. Moreover, consumers often take into account the visual appeal of the fruit when making their purchases.

The issue lies in the fact that the quality assessment of fruits and vegetables is carried out through manual processes, turning it into a slow and inconsistent procedure due to the subjectivity associated with the evaluation criteria of the individuals responsible for inspection. This human factor diminishes the efficiency of the evaluation process and adversely impacts the commercial

value of the fruits, affecting both consumers in general and farmers/suppliers, who may experience decreased profits due to the quality of their products. Therefore, the purpose of this study is to propose an improvement in the fruit classification process by developing deep learning models that enable the automatic classification of one of the main quality assessment parameters: the level of fruit deformation. In this work, apples, mangoes, and strawberries are analyzed as the fruits under study.

This article is structured as follows. Section 2 provides a review of the state of the art regarding CNN models applied to the identification of fruit deformities. Section 3 describes the contributions and the proposed methodology for conducting the work. In Section 4, the results of identifying levels of deformities in fruits using Single-Input and Multi-Input architectures with public CNN models are presented. Finally, the study's conclusions are provided in Section 5.

## 2 Literature Review

This section describes the state of the art concerning the analysis of deformities in fruits, specifically apples, mangoes, and strawberries. It also reviews the techniques and models that have been employed in previous studies for deformity analysis in these fruits.

### 2.1 Fruit Quality Inspection

Quality inspection is an indispensable aspect of the food industry. Numerous food items and pre-processed products rely on fruits in their production, prompting both retailers and wholesalers to seek fruits of the highest quality and appearance. This approach ensures consumer satisfaction and may result in higher remuneration for the seller when marketing their products.

To assess fruit quality, two methodological approaches are employed: invasive and non-invasive methods [3]. Invasive methods, also known as destructive methods, allow for physicochemical analysis by assessing nutritional parameters such as sugar content, acidity, and others through the manipulation of the fruit. On the other hand, non-invasive methods enable the evaluation of external features such as deformities (shape, size), ripeness level [4], and defect detection without compromising the fruit's integrity [5,6].

International markets demand that fruits undergo rigorous quality control, leading to the establishment of international standards by organizations such as the Economic Co-operation and Development (OECD), United States Department of Agriculture (USDA), and the regulatory bodies of the People's Republic of China.

When it comes to deformities, it is crucial to assess both the shape and size of the fruit. While Chinese and American regulations address both characteristics, European standards focus solely on the fruit's shape. Incorporating the evaluation of both characteristics would ensure a more accurate prediction of deformities. However, considering size adds a level of complexity to data collection. Given the impracticability of establishing a controlled environment for image acquisition and the paucity of publicly available data concerning size evaluation, this project opted to focus solely on the parameter of shape rather than size. Consequently, only shape classes delimited by the standards set forth by the OECD for deformity assessment are taken into consideration. For this reason, the categories employed for each fruit were: Extra Class, First Class, Second Class, and Ungraded which will be the class where we will send all those images that are already outside of any limit allowed by the OECD.

## 2.2 Fruit Deformities

During the growth and development process of fruits, alterations in their shape and size can occur as a result of various factors, including genetic causes [7]. A deformed fruit exhibits an atypical shape that does not align with the characteristic features of its species. In this study, we investigate the detection of deformities in fruits using CNNs. It is essential to consider the nature of these deformities to achieve precise classification and enable CNNs to learn to recognize fruits with deformities.

When dealing with deformities, typically two aspects are evaluated: shape and size. International organizations responsible for establishing quality parameters assess these aspects. Some examples include the OECD [8], the USDA [9], and governmental entities that establish their criteria, such as the Republic of China [10]. Table 1 summarizes the primary parameters evaluated in most fruits.

**Table 1.** Parameters evaluated for classifying deformities according to various international organizations.

Organizations	Categories	Shape	Size
OECD	Extra-Class, Class 1, Class 2	✓ Typical Shape of fruit	✗ Does not assess
USDA	Extra-Fancy, Fancy, Utility	✓ Typical Shape of fruit	✓ Measurement of the fruit's major diameter.
Republic of China	Excellent, Class 1, Class 2, Substandard	✓ Typical Shape of fruit	✓ Measurement of the fruit's major diameter. Measurement of the fruit's circumference from the top view.

In [11] and [12], both the shape and diameter of the apple are considered for categorization. However, the limitation of size as an evaluation parameter is that it requires the measurement of fruit diameters. Therefore, it is necessary to capture images under controlled conditions, taking into account parameters such as the working distance between the camera and the target to facilitate diameter calculation.

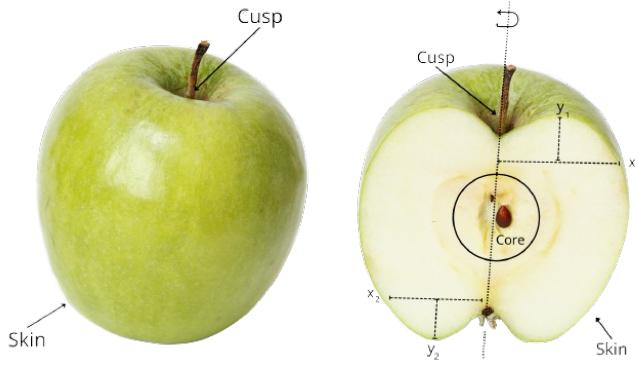
According to the OECD, for a fruit to be classified as Extra Class, Class I, or Class II, it must meet minimum requirements for each parameter in quality inspection such as ripeness, defects, and shape. If a fruit fails to meet the criteria for classification within these classes, it is deemed of low quality and categorized as Ungraded. The following sections elaborate on the permissible levels of deformities in apples, mangoes, and strawberries within international standards for commercialization.

**Apple Deformities** According to [13], the classification of deformities in apples is reflected in the classes depicted in Figure 1: Extra Class, Class 1, Class 2. The OECD stipulates that apples must have a symmetrical surface without evident deviations or bumps. This clearly describes the typical shape of an apple, which, during its development, should have a symmetrical form on both sides, as if it were a mirror image. When the apple lacks this symmetrical shape, it is considered deformed [14].

An apple whose distance between one end and its axis is similar to the distance between the axis and the other is considered symmetrical. As can be observed in Figure 2, both sides of the "top" part of the apple grow similarly.



**Fig. 1.** Apple Classification based on OECD standard.



**Fig. 2.** Apple Symmetry.

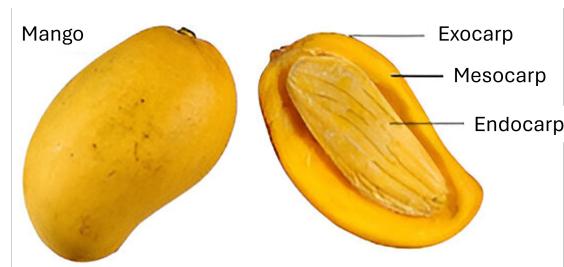
In the context of this research, all apples, regardless of their family, must have a symmetrical shape. Therefore, apples from different families can be included in the image dataset since they all contribute to the recognition of deformities in their symmetry.

**Mangoes Deformities** According to [13], the classification of deformities in mangoes is reflected in the classes depicted in Figure 3: Extra Class, Class 1, Class 2. According to the parameters set by the OECD, mangoes should have an oval shape. However, there are different mango families, so we cannot guarantee that every mango will have a similar shape. This study will focus on the properties of mangoes from *Mangifera Indica*, specifically the Tommy Atkins variety, and its traditional shape is ovoid (see Fig. 4).

**Strawberries Deformities** According to [13], the classification of deformities in strawberries is reflected in the classes depicted in Figure 5: Extra Class, Class 1, Class 2. Just like with apples and mangoes, there are various species and families of strawberries. Fortunately for this study, strawberry varieties are distinguished by their size and flavor characteristics, but the traditional conical shape is maintained [15]. As a result, any strawberry family will be included in the image collection (see Fig. 6). A strawberry is considered symmetrical if its shape from the calyx to the tip is like an inverted cone.



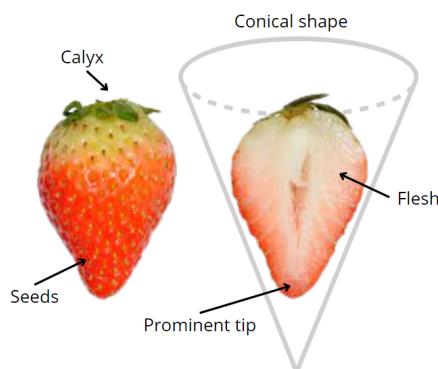
**Fig. 3.** Mangoes Classification based on OECD standard.



**Fig. 4.** Mango with elongated oval shape.



**Fig. 5.** Strawberries Classification based on OECD standard.



**Fig. 6.** Strawberry Symmetry.

### 2.3 Generation of datasets

When utilizing CNN models, it is essential to have an extensive database of images from which the model can extract meaningful features for each classification. However, having a large quantity of training images does not conclusively equate to obtaining a model with high accuracy. In [16], it was concluded that inadequate preprocessing of input images resulted in a model with low learning levels. From this perspective, the need to refine the dataset to ensure the quality of the input data for CNN models becomes evident.

Studies such as [11], [17], and [18] add value to their image datasets by employing preprocessing techniques, which involve segmenting the object of interest, applying enhancement processes, or using images in different visible spectra.

When acquiring images, it is common to turn to public platforms, although at times it is crucial to maintain strict control over the input data, as discussed in [11]. On the other hand, a proprietary acquisition process is often chosen, where the environment is prepared, and parameters such as luminosity and focus distance can be managed, among others. Furthermore, if the aim is to increase the number of training images, Data Augmentation techniques can be employed. The use of synthetic images is also an option. The work in [4] utilizes the Unreal Engine game engine to create realistic images of bananas at different stages of ripeness. In [19], on the other hand, DALL-E mini, a Text-to-Image artificial intelligence, was used to generate images of apples and mangoes with various types of defects such as rot, scabs, and blemishes. In both studies, the use of this type of image contributed to improving the model's performance.

### 2.4 Classification of deformities using CNN models

In recent years, the application of CNN models to images in the agricultural sector has gained significant attention, particularly in the classification of fruits based on various external quality parameters. While research has been conducted on identifying ripeness levels in fruits and classifying healthy or defective fruits using CNN models, there is a notable scarcity in the literature regarding the classification of fruit deformities. This study specifically addresses this issue by focusing on the parameter of deformation, which is crucial for ensuring the quality, categorization, and marketability of fruits.

The study in [12] focused on identifying the shape and size of apples. A modified VGG16 model was used, which consisted of two fully connected layers (traditional in VGG architectures), and an additional four convolutional layers were added. As a result of the modification, the model achieved an accuracy of 99.04%.

The study in [20] proposes an apple size classification system using their own CNN, the LightNet model. The model's generalization capacity was verified as it was tested with images of zizanias, yielding similar results in model accuracy. The proposed method achieves an accuracy of 99.31%, surpassing that obtained by the AlexNet, VGG11, ResNet18, MobileNet, MobileNetV2, and ShuffleNetV2 models.

Currently, there is a noticeable scarcity of published works on the classification of deformities in fruits using CNN models. This research gap highlights a significant opportunity for future investigations in the field. The application of deep learning techniques, such as CNNs, in the detection and classification of deformities in fruits represents a promising and relevant area, both for the agricultural industry and for the automation of quality inspection processes. Therefore, researchers are encouraged to explore and develop innovative approaches in this direction, aiming to enhance the efficiency and accuracy of fruit deformity classification.

## 2.5 Architectures Topology

Single-Input Architectures process a single type of input data at a time. This is the most common architecture and is used in a wide range of applications. Multi-Input Architectures can process two or more types of input data simultaneously. This allows the network to learn complex relationships between different types of data, such as RGB images and images in other spectrums. The architecture has multiple branches, each processing a type of input data, and then concatenates the features extracted from each branch to perform the final classification or prediction task [21,22].

## 3 Proposed Methodology

Throughout this section, the methodology employed in the development of this work is outlined. Additionally, it is important to highlight the contributions that this study offers to the scientific community:

- Three distinct datasets were developed for this study: Initially, a real dataset was compiled, sourced from public image repositories, and carefully refined according to international standards for shape classification. The second dataset, synthetic in nature, was generated using a blend of "text-to-image" and "image-to-image" tools, to produce high-fidelity images to enrich the training of CNN models. Lastly, a dataset comprising fruit silhouettes was created, wherein only the fruit shapes were meticulously extracted. Furthermore, the dataset generated by this study is made publicly available at the following link: (<https://github.com/luischuquim/Deformed-fruits.git>).
- The accuracies obtained by the three evaluated models: VGG16, MobileNetV2, and CIDIS are compared. It is important to highlight that the MobileNetV2 model has demonstrated superior performance compared to the other evaluated CNN models.
- The contribution of this study lies in the evaluation of two different training approaches: Single-Input architecture, which allows for progressive and linear learning in contrast to Multi-Input architecture training, involving the simultaneous training of two models with different input approaches. In the case of Multi-Input architecture training, two branches are trained separately, one with RGB images and the other with silhouette images.

### 3.1 Real Image Acquisition

As with any machine learning model, the availability of a large number of images is essential for CNN models to effectively learn and generalize results. In this regard, it was considered that the appropriate number of images for each deformity category would be around 5,000; meaning a total of 20,000 images per fruit type were required to classify the 4 types of deformities. To compile this dataset, various online sources were explored, such as Kaggle, Github, and Mendeley. Initially, datasets already classified similarly to the established categories (Extra, Class I, Class II, and Unclassified) were sought. However, the search did not yield any public datasets with these specific classes, although datasets intended for other types of classifications were found. Faced with this situation, it was decided to adopt any available fruit image datasets, such as apples, mangoes, and strawberries, and subsequently classify the images according to the deformity class. Following this, the manual classification of each image obtained from public platforms was refined. Each image was categorized according to the accepted limits for each class. At the same time, the dataset was

curated by removing fruit images that did not contribute significantly to the training process. This included images with non-uniform backgrounds that were difficult to remove, as well as images of fruits captured from the top or bottom perspectives.

At the end of the classification process, a significant imbalance was observed among the deformity categories, with the Extra Class having the highest number of images, in contrast to Class II and Ungraded. To address this issue and balance the image dataset, data augmentation techniques were applied, increasing the number of images by a factor of up to 8x for the classes that required it.

### 3.2 Synthetic Image Acquisition

The generation of synthetic images was carried out in multiple stages, aiming to reach 5,000 images per class for each type of fruit, resulting in a synthetic dataset of 20,000 images per fruit. The initial phase entailed generating near-perfect images for each fruit and category using Easy Diffusion tool version 2.5 [23]. This tool facilitated the creation of diverse image variants corresponding to the categories from texts. However, at a certain point, it ceased to provide significant variations, or the generated images diverged from the expected outcome. Consequently, the decision was made to transition to generating images from existing ones.

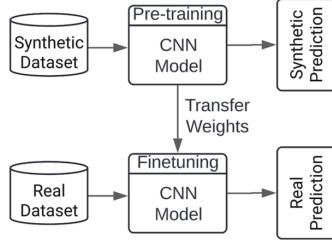
Following the approach outlined in [24], small sets of images (between 200 and 800) were created, depending on the different shape scales for each category (Extra, Class I, Class II, Ungraded) and fruit, especially for those classes that were poorer, such as Class II and Ungraded. Models were then trained using LORA with AUTOMATIC 1111 [25]. With the results obtained, the image generator Kohya [26] was used. Unlike Easy Diffusion, this tool allows us to load models trained and generated by AUTOMATIC 1111. This enables us to combine the trained image model with text and generate additional variations beyond what the model learned through LORA.

### 3.3 Model Pre-training Approach

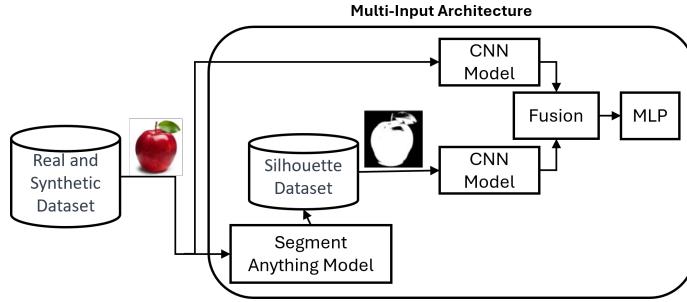
The training started from the CIDIS model [4]. This model, being custom-made, does not have an initial weight matrix. Therefore, it was first trained with the synthetic image dataset. Once it learned the characteristics of the synthetic data, the learning was transferred to further training using the real image dataset. Figure 7 illustrates the process graphically. In the Model Pre-training approach, we utilized the VGG16 and MobileNetV2 architectures with pre-trained weights, prepared for classification. Thus, in the Model Pre-training process of these architectures, only real images were employed, unlike the CIDIS model, which was used for the initial synthetic phase and subsequent Model Pre-training with real images.

### 3.4 Multi-Input Network Approach

In the analysis of implementing a Multi-Input architecture, the focus lies primarily on evaluating the differences between the input images of each branch [21,27,28]. Subsequently, the result obtained from each branch is unified through a feature fusion process, and classification is carried out after a Multilayer Perceptron (MLP) layer. Given that this approach prioritizes the study of fruit shape, it was necessary to disregard features typically recognized by CNNs, such as color and texture in the images. Therefore, it was decided that the input images for one of the branches would be RGB images of the fruits (from the real and synthetic datasets), while the other branch would receive the silhouettes of the same fruits in those images, as depicted in Figure 8.



**Fig. 7.** Model Pre-training Approach



**Fig. 8.** Multi-Input Network.

For this approach, the CIDIS, MobileNetV2, and VGG16 models were used again, with the difference that the CIDIS model was trained without initial weights, while the MobileNetV2 and VGG16 models were initialized with the same weights found before training with the Model Pre-training approach.

## 4 Experimental Results

In this section, we elaborate on the datasets of real and synthetic images, which were both acquired and generated to ensure high-quality data for evaluating CNN models. Subsequently, we provide a detailed presentation of the results obtained for each training approach, taking into account the various CNN models utilized.

After gathering the minimum required set of real fruit images, data augmentation techniques, including rotations, horizontal flips, and vertical flips, were employed to achieve the agreed-upon number of images for training, validation, and testing (80/10/10). The resulting image size was fixed at 224x224 pixels. In summary, a total of 20,000 images per fruit were obtained, evenly distributed among the four deformation categories (see Table 2). On the other hand, given the scarcity of suitable real images for fruit deformity analysis, synthetic datasets were generated leveraging the capabilities provided by the Stable Diffusion tool, as well as the technologies mentioned in Section 3.2. These tools produced high-quality synthetic images with dimensions of 512x512 pixels. In many cases, images from the real dataset were used as a basis for generating similar synthetic images. Subsequently, data augmentation techniques were employed to rotate and flip the images,

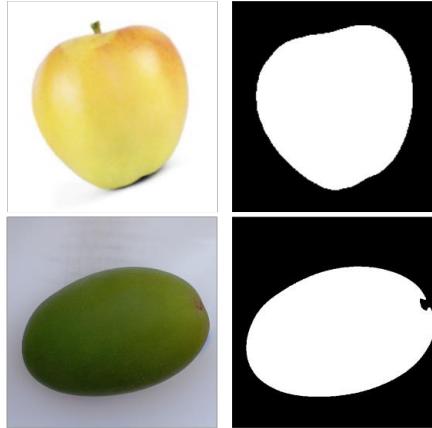
aiming to introduce variability into the dataset. Similar to the real image set, a total of 20,000 synthetic images per fruit were obtained, evenly distributed among the deformity categories.

**Table 2.** Number of real and synthetic images in the fruit dataset.

Fruits	Extra Class		First Class		Second Class		Out of Class	
	Real	Synthetic	Real	Synthetic	Real	Synthetic	Real	Synthetic
Apples	4050 (*950)	5000	1741 (*3259)	5000	651 (*4349)	5000	1961 (*3039)	5000
Strawberry	2782 (*2218)	5000	875 (*4125)	5000	759 (*4241)	5000	679 (*4321)	5000
Mango	4317 (*683)	5000	1889 (*3111)	5000	655 (*4345)	5000	734 (*4266)	5000

*Note: The numbers marked with (\*) represent augmented images.*

Furthermore, silhouette images were prepared, which are closely linked to the shape of an object. In our case, fruit analysis revolves around its shape. Therefore, it is crucial to train the models using fruit silhouettes. For the automatic generation of these silhouette images, the Segment Anything Model (SAM) tool was utilized. The generation of fruit silhouette images was successful in most cases (see Fig. 9). For fruit images with incorrect silhouettes, individual processing was applied, or they were excluded entirely, to obtain a refined and high-quality dataset.



**Fig. 9.** Obtaining the Silhouette of Each Fruit

Once the datasets were prepared, hyperparameters were set for training each CNN model. Each training was conducted with a different combination based on the obtained results. The use of various optimizers, such as Adam, Nadam, and RMSProp, along with variable learning rates, is

noted. Typically, trainings were configured to run until epoch 30, and the learned weights were restored from the epoch with the best performance.

#### 4.1 Results with Model Pre-training approach

When training and evaluating the CNN models using only real data, the results shown in Table 3 were obtained. These results provide a detailed assessment of the models' performance, including validation accuracy, validation loss, precision, recall, F1-score, and test accuracy. Table 3 offers a thorough comparison of these performance metrics across different fruits and model architectures, providing a clear understanding of each approach's effectiveness in image classification. However, the results shown in Table 4 are those obtained with the Model Pre-training approach, which achieves better results than the models trained only with real images. We based our selection on the model with the highest accuracy and the lowest loss.

**Table 3.** Training Results with CIDIS, MobileNetV2, and VGG16, Model Training Approach using only Real Data.

Fruit	Model	Val Accuracy	Val Loss	Precision	Recall	F1-score	Test Accuracy
Apple	CIDIS	0.6815	0.8491	0.6100	0.5900	0.5500	0.5915
	<b>MobileNetV2</b>	<b>0.7848</b>	<b>0.5851</b>	<b>0.6100</b>	<b>0.6100</b>	<b>0.5700</b>	<b>0.6095</b>
	VGG16	0.7747	0.6073	0.5600	0.5700	0.5200	0.5685
Mango	CIDIS	0.5155	1.2260	0.4400	0.4100	0.4100	0.4400
	<b>MobileNetV2</b>	<b>0.5155</b>	<b>0.9612</b>	<b>0.6700</b>	<b>0.6700</b>	<b>0.6700</b>	<b>0.6710</b>
	VGG16	0.5746	1.0673	0.4900	0.4800	0.4800	0.4770
Strawberry	CIDIS	0.5601	1.3034	0.4900	0.4100	0.4000	0.4110
	<b>MobileNetV2</b>	<b>0.6971</b>	<b>1.2005</b>	<b>0.7000</b>	<b>0.7000</b>	<b>0.7000</b>	<b>0.6980</b>
	VGG16	0.6694	1.0141	0.6300	0.6100	0.6000	0.6100

**Table 4.** Training Results with CIDIS, MobileNetV2, and VGG16, Model Pre-training approach.

Fruit	Model	Val Accuracy	Val Loss	Precision	Recall	F1-score	Test Accuracy
Apple	CIDIS	0.8956	0.3645	0.8900	0.8800	0.8800	0.8806
	<b>MobileNetV2</b>	<b>0.9325</b>	<b>0.2087</b>	<b>0.9100</b>	<b>0.9000</b>	<b>0.9000</b>	<b>0.9044</b>
	VGG16	0.9050	0.2511	0.9000	0.8900	0.8900	0.8931
Mango	CIDIS	0.08679	0.3819	0.8700	0.8700	0.8700	0.8694
	<b>MobileNetV2</b>	<b>0.9420</b>	<b>0.1594</b>	<b>0.9400</b>	<b>0.9400</b>	<b>0.9400</b>	<b>0.9425</b>
	VGG16	0.9022	0.2290	0.9600	0.9600	0.9600	0.9604
Strawberry	CIDIS	0.8010	0.5258	0.8200	0.8100	0.8200	0.8135
	<b>MobileNetV2</b>	<b>0.9189</b>	<b>0.1945</b>	<b>0.9200</b>	<b>0.9200</b>	<b>0.9200</b>	<b>0.9220</b>
	VGG16	0.8962	0.2646	0.9100	0.9100	0.9100	0.9050

When training and evaluating the CNN models using only real data, as shown in Table 3, the results were suboptimal due to the limited amount of available data and the imbalance in the number of images across classes. This imbalance negatively impacts the models' ability to generalize

and correctly classify images, resulting in lower performance metrics. However, by applying the data augmentation methodology through the inclusion of synthetic images, as detailed in Table 4, the models' performance was significantly improved. This approach allowed for the balancing of the datasets and provided the models with a greater diversity of examples during training, leading to increased accuracy and reduced loss. Therefore, the use of synthetic images not only mitigates the issues associated with small and imbalanced datasets but also optimizes the models' ability to achieve more precise classifications.

In this way, it is determined that the best architecture for this type of training has been MobileNetV2 for all fruits. It is worth mentioning that in the case of mango, VGG16 had the highest precision; however, the loss was very high. Therefore, we consider that a better result can be achieved with MobileNetV2.

#### 4.2 Results with Multi-Input Network Approach

We relied on studies that validated the results of Multi-Input networks for extracting features between two similar images. In our approach, we feed one branch with an RGB image and the other with the silhouette of the same image. This process effectively identifies similarities or differences in the shape of the object in the image, in our case, the shape of the fruit.

The Multi-Input network approach was employed with the complete dataset of real and synthetic images, resulting in a larger volume of images than Model Pre-training. All models obtained with this approach exhibited an accuracy above 90%, with the MobileNetV2 model achieving the highest accuracy for apples and strawberries, whereas for mangoes, the VGG16 model was superior (see Table 5).

**Table 5.** Training Results with CIDIS, MobileNetV2, and VGG16, Multi-Input Network Approach

Fruit	Model	Val Accuracy	Val Loss	Precision	Recall	F1-score	Test Accuracy
Apple	CIDIS	0.7626	1.6818	0.8639	0.8366	0.8107	0.8249
	<b>MobileNetV2</b>	<b>0.8068</b>	<b>2.2276</b>	<b>0.9257</b>	<b>0.9158</b>	<b>0.9161</b>	<b>0.9257</b>
	VGG16	0.8060	1.9643	0.8966	0.8920	0.8832	0.8844
Mango	CIDIS	0.9234	0.2577	0.9237	0.9232	0.9233	0.9232
	MobileNetV2	0.9440	0.1505	0.9394	0.9393	0.9390	0.9393
	<b>VGG16</b>	<b>0.9501</b>	<b>0.1696</b>	<b>0.9551</b>	<b>0.9539</b>	<b>0.9541</b>	<b>0.9551</b>
Strawberry	CIDIS	0.9188	0.2496	0.9145	0.9171	0.9154	0.9171
	<b>MobileNetV2</b>	<b>0.9197</b>	<b>0.2330</b>	<b>0.9455</b>	<b>0.9228</b>	<b>0.9213</b>	<b>0.9394</b>
	VGG16	0.8769	0.4013	0.9623	0.9204	0.9190	0.9204

Based on the results obtained, it is evident that the MobileNetV2 model with the Multi-Input architecture achieved the best outcomes for apples and strawberries, with an accuracy of 92% and 93%, respectively. However, the VGG16 model proved to be superior for mangoes, with an accuracy of 95%. Consequently, the most favorable results for the Multi-Input approach compared to the Single-Input approach are summarized in Table 6.

The results obtained in this study have significant implications for both research and the agricultural industry. The high accuracy achieved by the MobileNetV2 and VGG16 models in classifying deformed fruits suggests that these architectures can be implemented in automated quality classification systems, significantly reducing the reliance on manual and subjective evaluations. This

**Table 6.** Results Comparison

Fruits	Models	Accuracy
Apple	MobileNetV2 (Single-Input)	0.9044
	<b>MobileNetV2 (Multi-Input)</b>	<b>0.9257</b>
Mango	MobileNetV2 (Single-Input)	0.9425
	<b>VGG16 (Multi-Input)</b>	<b>0.9551</b>
Strawberry	MobileNetV2 (Single-Input)	0.9220
	<b>MobileNetV2 (Multi-Input)</b>	<b>0.9394</b>

not only improves the consistency and speed of the inspection process but also optimizes the supply chain by ensuring that higher-quality products reach consumers, which, in turn, could increase profitability for farmers and suppliers. Additionally, the insights gained on the effectiveness of Multi-Input architectures compared to Single-Input provide a solid foundation for future research in automating quality assessment in other types of fruits and agricultural products.

The findings of this study, focused on the parameter of fruit deformation, enhance and complement the results obtained in previous research, such as the classification of healthy and defective fruits [19], and the assessment of banana ripeness levels [4]. These advancements contribute to the development of a comprehensive fruit categorization system based on international standards for external quality control, thereby strengthening the ability to ensure consistent standards in the classification and commercialization of agricultural products.

The increase in computational cost of the multi-input approach compared to the single-input method may seem like a challenge, especially in contexts with limited resources. However, it is crucial to consider the tangible benefits this approach offers in terms of precision and efficiency in fruit classification. By utilizing multiple inputs, the model can learn from different features and patterns present in both RGB and silhouette images, thereby enhancing its ability to differentiate the levels of deformities in fruits. This improvement in accuracy not only reduces the error rate in classification but also minimizes the risk of deformed products reaching the market, which can negatively impact both consumers and producers.

## 5 Conclusions

A public dataset has been generated, presenting a wide range of images, including both real and synthetic images of apples, mangoes, and strawberries, thus addressing their morphological diversities. This dataset is evenly divided for each fruit into 5,000 samples for each of the four defined classes: Extra Class, First Class, Second Class, and Out of Class. Furthermore, it is supplemented with an additional set of 20,000 silhouette images generated from each RGB image, providing a more comprehensive and detailed representation of the morphological characteristics of the apples in the dataset. For this work, a dataset was obtained that enabled successful training. AI data generation tools such as Stable Diffusion played a fundamental role in achieving this accomplishment. Analyz-

ing the shape of fruits involves being very careful with the images used. The results obtained show that training with binary images where only shape details are present significantly aids learning, as is the case with the Multi-Input Network approach. To achieve the goal of having a good dataset, it is essential to obtain as many images as possible that are easy to segment, such as images with a consistent background. Evaluating different architectures such as MobileNetV2, VGG16, and the one proposed by CIDIS has helped us understand that for this type of task, it is more advantageous to use lightweight architectures like MobileNetV2, as it yielded the best accuracy results.

Both approaches, Model Pre-training, and the Multi-Input Network, have yielded good results in deformity classification. However, the Multi-Input Network approach stood out for its efficiency, grounded in the capability of the branches corresponding to RGB images and silhouette images to capture distinctive characteristics of fruit deformities. In the case of the branch processing RGB images, it benefited from detailed information regarding edges, textures, color patterns, and specific shapes, all of which are fundamental elements for the precise identification of deformities. Conversely, the branch associated with silhouette images allowed for the consideration of aspects such as shape, contours, edges, and pixel distribution within the silhouette, thereby enriching the representation of the structural characteristics of the fruits. For future work, a comparison with ViT and DINO ViT models is proposed, as they have demonstrated high efficacy in image classification tasks, to evaluate their performance in classifying fruit deformities compared to lightweight architectures like MobileNetV2. Additionally, it is essential to improve the quality of the dataset using advanced artificial intelligence tools to manage, clean, and select visual data quickly and at scale. These enhancements will ensure a more robust foundation for model training and evaluation, enhancing their ability to accurately and efficiently identify deformities.

## Acknowledgment

This work has been partially supported by the ESPOL-CIDIS-11-2022 project.

## References

1. Olorunfemi, B.J., Kayode, S.E.: Post-harvest loss and grain storage technology-a review. *Turkish Journal of Agriculture-Food Science and Technology* **9**(1), 75–83 (2021)
2. Shewfelt, R.L., Prussia, S.E.: Challenges in handling fresh fruits and vegetables. In: *Postharvest Handling*, pp. 167–186. Elsevier (2022)
3. Vetrekar, N.T., Gad, R.S., Fernandes, I., Parab, J.S., Desai, A.R., Pawar, J.D., Naik, G.M., Umapathy, S.: Non-invasive hyperspectral imaging approach for fruit quality control application and classification: case study of apple, chikoo, guava fruits. *Journal of Food Science and Technology* **52**, 6978–6989 (2015)
4. Chuquimarpa, L.E., Vintimilla, B.X., Velastin, S.A.: Banana Ripeness Level Classification Using a Simple CNN Model Trained with Real and Synthetic Datasets. In: *VISIGRAPP (5: VISAPP)*, pp. 536–543 (2023)
5. Chuquimarpa, L.E., Vintimilla, B.X., Velastin, S.A.: Classifying Healthy and Defective Fruits with a Multi-Input Architecture and CNN Models. In: *ICPRS*, accepted for presentation (2024)
6. Coello, O.M., Coronel, M.A., Carpio, D.N., Vintimilla, B.X., Chuquimarpa, L.E.: Enhancing Apple's Defect Classification: Insights from Visible Spectrum and Narrow Spectral Band Imaging. In: *ICPRS*, accepted for presentation (2024)
7. Behera, S.K., Rath, A.K., Mahapatra, A., Sethy, P.K.: Identification, classification & grading of fruits using machine learning & computer intelligence: a review. *Journal of Ambient Intelligence and Humanized Computing*, 1–11 (2020)

8. Lidror, A., Prussia, S.E.: Improving quality assurance techniques for producing and handling agricultural crops. *Journal of Food Quality* **13**(3), 171–184 (1990)
9. Huang, K.M., Guan, Z., Hammami, A.M.: The US fresh fruit and vegetable industry: An overview of production and trade. *Agriculture* **12**(10), 1719 (2022)
10. Zhou, J.H., Kai, L., Liang, Q.: Food safety controls in different governance structures in China's vegetable and fruit industry. *Journal of Integrative Agriculture* **14**(11), 2189–2202 (2015)
11. Wang, J., Huo, Y., Wang, Y., Zhao, H., Li, K., Liu, L., Shi, Y.: Grading detection of "Red Fuji" apple in Luochuan based on machine vision and near-infrared spectroscopy. *Plos one* **17**(8), e0271352 (2022)
12. Hu, G., Zhang, E., Zhou, J., Zhao, J., Gao, Z., Sugirbay, A., Jin, H., Zhang, S., Chen, J.: Infield apple detection and grading based on multi-feature fusion. *Horticulturae* **7**(9), 276 (2021)
13. Porat, R., Lichter, A., Terry, L.A., Harker, R., Buzby, J.: Postharvest losses of fruit and vegetables during retail and in consumers' homes: Quantifications, causes, and means of prevention. *Postharvest biology and technology* **139**, 135–149 (2018)
14. Chakrabarti, A., Michaels, T.C., Yin, S., Sun, E., Mahadevan, L.: The cusp of an apple. *Nature Physics* **17**(10), 1125–1129 (2021)
15. Liu, L., Ji, M.L., Chen, M., Sun, M.Y., Fu, X.L., Li, L., Gao, D.S., Zhu, C.Y.: The flavor and nutritional characteristic of four strawberry varieties cultured in soilless system. *Food Science & Nutrition* **4**(6), 858–868 (2016)
16. Vujović, Ž., et al.: Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications* **12**(6), 599–606 (2021)
17. Sun, L., Liang, K., Song, Y., Wang, Y.: An improved CNN-based apple appearance quality classification method with small samples. *Ieee Access* **9**, 68054–68065 (2021)
18. Garillos-Manliguez, C.A., Chiang, J.Y.: Multimodal Deep Learning via Late Fusion for Non-Destructive Papaya Fruit Maturity Classification. In: 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–6. IEEE (2021)
19. Pacheco, R., González, P., Chuquimarca, L.E., Vintimilla, B.X., Velastin, S.A.: Fruit Defect Detection Using CNN Models with Real and Virtual Data. In: VISIGRAPP (4: VISAPP), pp. 272–279 (2023)
20. Cao, J., Sun, T., Zhang, W., Zhong, M., Huang, B., Zhou, G., Chai, X.: An automated zizania quality grading method based on deep classification model. *Computers and Electronics in Agriculture* **183**, 106004 (2021)
21. Mesa, Armacheska Rivero, and John Y Chiang. "Multi-input deep learning model with RGB and hyperspectral imaging for banana grading." *Agriculture* 11.8 (2021): 687.
22. Pipitsunthonsan, P., Pan, L., Peng, S., Khaorapapong, T., Nakasathien, S., Channumsin, S. & Chongcheawchamnan, M. Palm bunch grading technique using a multi-input and multi-label convolutional neural network. *Computers And Electronics In Agriculture*. **210** pp. 107864 (2023)
23. Brade, S., Wang, B., Sousa, M., Oore, S., Grossman, T.: Promptify: Text-to-image generation through interactive prompt exploration with large language models. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–14 (2023)
24. Somepalli, G., Singla, V., Goldblum, M., Geiping, J., Goldstein, T.: Diffusion art or digital forgery? investigating data replication in diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6048–6058 (2023)
25. Hidalgo, R., Salah, N., Chandra Jetty, R., Jetty, A., Varde, A.S.: Personalizing text-to-image diffusion models by fine-tuning classification for AI applications. In: Intelligent Systems Conference, pp. 642–658. Springer (2023)
26. Masrouri, M., Qin, Z.: Towards data-efficient mechanical design of bicontinuous composites using generative AI. *Theoretical and Applied Mechanics Letters* **14**(1), 100492 (2024)
27. Dua, Nidhi, Shiva Nand Singh, and Vijay Bhaskar Semwal. "Multi-input CNN-GRU based human activity recognition using wearable sensors." *Computing* 103.7 (2021): 1461–1478.
28. Choudhary, Anurag, Rismaya Kumar Mishra, Shahab Fatima, and BK Panigrahi. "Multi-input CNN based vibro-acoustic fusion for accurate fault diagnosis of induction motor." *Engineering Applications of Artificial Intelligence* 120 (2023): 105872.