

Classifying Healthy and Defective Fruits with a Multi-Input Architecture and CNN Models

Luis Chuquimarca
CIDIS

ESPOL Polytechnic University
Guayaquil, Ecuador
lchuquim@espol.edu.ec,
and

FACSISTEL

UPSE Santa Elena Peninsula State University
La Libertad, Ecuador

Boris Vintimilla
CIDIS

ESPOL Polytechnic University
Guayaquil, Ecuador
boris.vintimilla@espol.edu.ec

Sergio Velastin
School of EECS

Queen Mary University of London
London, UK
sergio.velastin@ieee.org,
and

Dept. of Computer Science
University Carlos III
Madrid, Spain

Abstract—This study presents an investigation into the utilization of a Multi-Input architecture for the classification of fruits (apples and mangoes) into healthy and defective states, employing both RGB and silhouette images. The primary aim is to enhance the accuracy of CNN models. The methodology encompasses image acquisition, preprocessing of datasets, training, and evaluation of two CNN models: MobileNetV2 and VGG16. Results reveal that the inclusion of silhouette images alongside the Multi-Input architecture yields models with superior performance compared to using only RGB images for fruit classification, whether healthy or defective. Specifically, optimal results were achieved using the MobileNetV2 model, achieving 100% accuracy. This finding suggests the efficacy of this combined methodology in improving the precise classification of healthy or defective fruits, which could have significant implications for applications related to external quality inspection of fruits.

Index Terms—Fruit Defects, Defects Classification, CNN, Multi-Input, Silhouette Images

I. INTRODUCTION

In today's post-harvest processes, fruit classification is carried out through quality inspection conducted by specialized agents, who consider parameters such as ripeness, deforms, and defects [1], [2]. The considerations adhere to international standards for defects, as attributed by regulatory bodies governing international norms. This process is traditionally performed manually, and due to the commercial value attributed to fruits based on their quality level, higher quality implies higher profits and access to better markets [3]. Additionally, both small-scale retailers and supermarkets, as well as consumers, take various factors into account in the purchase and sale of these foods. However, visual appeal is a paramount factor in fruit selection. In the current global market, ensuring the quality control of agricultural products is paramount, with indispensable international standards for producing high-quality products essential for human health [4], [5]. For this study, the considered parameter is fruit defects (using apples and mangoes), which aids us in classifying whether a fruit is healthy or defective.

In recent years, there has been a significant advancement in the development of automated systems based on computer vision technology to replace manufacturing operations [6], [7]. Specifically, the classification of healthy or defective fruits has been addressed using convolutional neural network (CNN) models, which utilize fruit images as input [8], [9]. These images are evaluated using a series of weights derived from the training characteristics of the CNN model. The primary types of images in this context are binary and true color images, representing the predominant forms of digital images. To obtain specific information from a digital image, an essential process called segmentation is employed [10], [11]. Following this, a name or label is assigned to the image based on certain visual criteria, with the two classes of healthy or defective fruits being considered in this study [12], [13]. In this research, two simple CNN models, MobileNetV2 and VGG16, are used and analyzed.

The key contributions of this scientific article are as follows:

- Improvement of the performance of CNN models used in the study [14] by exclusively utilizing real images, no longer involving the generation of synthetic images.
- Implementation of image preprocessing through fruit segmentation in silhouette format, which is a well-known shape descriptor. This technique allows for a more detailed observation of the presence of defects (rot, bruises, scabs, and black spots) in fruits, significantly enhancing the accuracy of the CNN classification model.
- Introduction of a Multi-Input architecture that leverages both the features of RGB (Red-Green-Blue) images and silhouettes respectively in each branch. This innovative technique surpasses the results obtained in a previous study [14], which relied solely on a Single-Input architecture. The utilization of the Multi-Input architecture enables a better capture of the distinctive features of fruits, resulting in higher classification accuracy.
- The methodology employed in this study integrates the utilization of a dataset comprised of RGB images along-

side their corresponding silhouettes to feed into a multi-input architecture. This architecture undergoes training utilizing both the MobileNetV2 and VGG16 models. As a result, a remarkable precision of 100% was attained in the classification of both healthy and defective apples using the MobileNetV2 model. Similarly, in the classification of healthy and defective mangoes, the MobileNetV2 model achieved an impressive precision of 99.53%. These results stand in favorable comparison to those reported in the previous study [14], where the MobileNetV2 model yielded a precision of 97.50% for apples and 92.90% for mangoes. This comparative analysis illuminates the relative efficacy of various architectures in this domain, offering invaluable insights for future research endeavors.

The structure of this article is as follows: Section II provides a comprehensive review of the literature related to the classification of healthy and defective fruits. Subsequently, Section III details the proposed methodology for conducting the study. Section IV showcases the results obtained by the CNN (Multi-Input) models. Finally, in Section V, the conclusions derived from this work are presented.

II. RELATED WORKS

This section outlines previous studies related to the use of CNN models for classifying healthy or defective fruits, providing an overview of the criteria to consider for the current research. Common defects that may manifest in fruits include bruising, spots, and rot. These anomalies, often attributable to inadequate storage conditions or damage during handling, have a significant impact on the sensory and nutritional quality of fruits. The accurate classification of defective fruits is essential for the food industry, as it allows for the separation of products of optimal quality from those that may be deemed unfit for human consumption. In this context, the application of CNN models for the automatic classification of healthy or defective fruits has emerged as a promising solution, offering an efficient and objective assessment of fruit quality at various stages of the supply chain.

The study conducted by [13] employs deep learning techniques to identify physiological disorders (bitter pit, shriveling, superficial scald) in apples. Leveraging CNN models, which enable efficient feature extraction from fruit images, they classify images of apples affected by post-harvest physiological disorders. Additionally, they highlight the Xception model with the best performance, achieving an accuracy, precision, recall, and F1-score average of 0.996, 0.994, 0.996, and 0.998 respectively, outperforming other models such as ResNet and MobileNet. However, it is important to note that the study does not address the detection of additional defects (bruising, spots, rot) in apples, nor does it utilize color balance models to enhance images, or explore the use of architectures or techniques to strengthen CNN models.

The study reported in [15] focuses on detecting diseases in apples, such as blotch, scab, and rot, using deep learning and machine learning approaches. A CNN model with three convolution layers was utilized to classify healthy and

defective apples. Additionally, pre-trained models were compressed using Differential Evolution, achieving a maximum compression rate of 82.19% for the VGG16 model without significant performance loss, thus facilitating its implementation on mobile devices. Binary classification of fruits with defects, alongside a healthy class, proves suitable for deep learning techniques. The proposed CNN model achieved a classification accuracy of 93.75%, while the best performance among machine learning classifiers was 99% using Random Forest, although its functionality is specifically geared towards detection. However, it is important to note that the research does not address disease detection in other fruits.

The aim of the research in [12] is to construct and evaluate various CNN models, including GoogLeNet, AlexNet, and a proposed CNN model, to categorize fresh and rotten appearances using a dataset of apple images. The results obtained reveal that GoogLeNet achieved an accuracy of 100%, surpassing AlexNet, which achieved 99.69%, and the proposed CNN model, which reached 87.24%. However, it is worth noting that a more thorough review of the dataset and an increase in the number of images (greater than 1000 images per class) could allow for the generalization of findings and the development of even more robust CNN models.

The study [14] aims to evaluate different CNN models in identifying various defects in apples and mangoes to ensure the quality of fruit production. Utilizing the InceptionV3, MobileNetV2, VGG16, and DenseNet121 models, trained with a dataset comprising real and synthetic images of apples and mangoes, as well as both healthy and defective fruits exhibiting rot, bruises, scabs, and black spots. MobileNetV2 demonstrated the highest accuracy, achieving 97.50% for apples and 92.50% for mangoes, making it the optimal choice for classifying healthy or defective fruits. In contrast, InceptionV3 and DenseNet121 exhibited accuracy above 90%, while VGG16 had a less satisfactory performance, failing to surpass 80% accuracy for either type of fruit. The robust dataset comprises 20k images (10k for each fruit), encompassing real and synthetic images. However, there is a need to enhance the CNN classifier model for defect classification and explore advanced architectures as potential future research directions.

Another article [16] introduces a proprietary algorithm based on deep learning designed for the identification and assessment of the state of fruits, distinguishing between those that are healthy and those with defects. Eight types of fruits were addressed, resulting in a total of 16 categories for classification. A CNN model with a Directed Acyclic Graph (DAG) structure was proposed for fruit recognition and classification. This CNN model demonstrated an accuracy of 94.43% in classifying 1600 test images, with processing times of approximately 45 to 55 milliseconds. The DAG structure of the CNN allows each branch to learn characteristics of different sizes, enabling increased network depth without compromising processing speed. However, it lacks an analysis of the types of defects present in the fruits, as well as the standards considered for conducting quality inspection thereof.

In [17], a computer-assisted classification system has been

presented for detecting defects in mangoes using a deep learning approach. Specifically, the CNN model is employed with a publicly accessible mango dataset as input. Experimental results reveal that the proposed method achieved an accuracy of 98%, tested on a set of 800 mango images. The proposed model has not been compared with other publicly available models, which could be an aspect to consider in future research.

Following a thorough review of the state of the art, four types of defects commonly found in apples and mangoes were identified, which are categorized into the following groups: rot, bruises, scabs, and black spots [14]. In the specific case of apples, studies such as that conducted by [13] identify three types of defects such as bitter pit classified within the black spots category, shriveling associated with rot, and superficial scald belonging to the scabs category. Conversely, research such as that carried out by [15] identifies three types of defects a blotch classified within the scabs category, scab, and rot.

Upon reviewing studies [16] and [17], which investigate defects in mangoes, it is evident that the specific types of defects are not detailed. In contrast, study [14] explicitly identifies defects such as rot, bruises, scabs, and black spots, which are also addressed in the present work.

After examining the reviewed studies, we observe a trend toward strengthening existing CNN models by exploring available datasets and implementing novel architectures to enhance their performance. This leads to obtaining superior results in the evaluation metrics of the CNN models.

III. PROPOSED APPROACH

This section presents the dataset acquisition along with its respective preprocessing, which includes generating silhouette images using segmentation techniques. The proposed approach aims to enhance the accuracy of CNN models in binary classification of apples, distinguishing between healthy and defective ones, which include bruises, blemishes, rot, and black spots. Therefore, we propose two strategies in this work. The first one is based on using silhouette images to complement RGB images, resulting in improved binarization of the dataset. The second strategy takes advantage of the first by incorporating it into the Multi-Input architecture where in one of its branches silhouette images are used, while in the other the corresponding RGB images are used, both serving as inputs for the CNN models used. The following sections provide more details of the proposed methodology.

A. Data Acquisition

The input dataset was sourced from a public platform (<https://github.com/luischuquim/Healthy-Defective-Fruits>), encompassing a combination of real and synthetic images. However, only 10k real images of apples and mangoes were utilized. Notably, the dataset is carefully labeled, ensuring an equitable distribution between healthy and defective fruits for each fruit type.

The acquired image dataset has been prepared for segmentation using the Segment Anything Model (SAM) tool [18],

allowing for the extraction of fruit silhouettes from the images. These silhouettes facilitate the identification of the object's shape. In our case, healthy fruits are binarized, and the complete silhouette of the fruit can be observed in white. Conversely, defective fruits are also binarized, but a difference can be observed when defects are present. These defects manifest in the white fruit silhouette with dark black pigmentations (see Fig. 1 and Fig. 2).

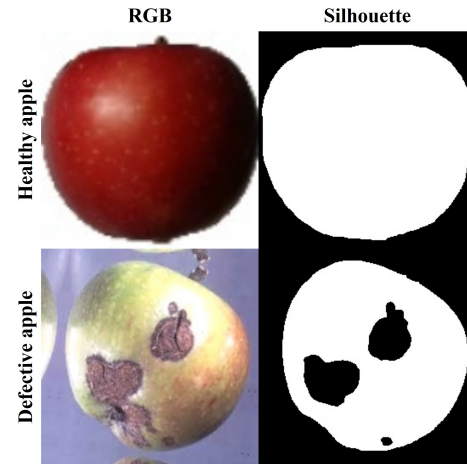


Fig. 1. Apple silhouette.

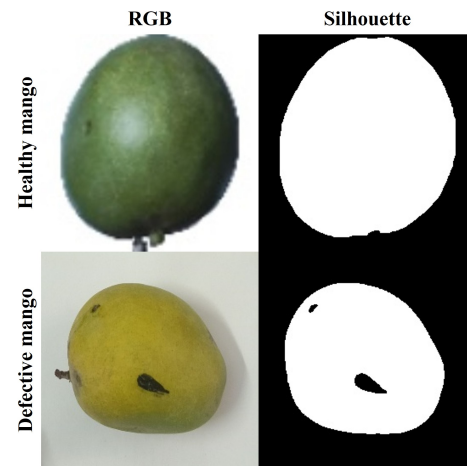


Fig. 2. Apple silhouette.

After collecting the dataset consisting of 5k images per fruit, the SAM model was employed to generate segmented images. It was discovered that approximately 25% of the images were segmented incorrectly. Consequently, a manual refinement of the dataset was carried out, resulting in the total number of images as indicated in Table I.

B. CNN Models Implemented

For training (80%), validation (10%), and testing (10%), the dataset is distributed as shown in Table I. The images were organized into directories by fruit type (apple and mango),

TABLE I
DATASET DISTRIBUTION.

Fruits	Category	Evaluation	# Original Images	# Refined Images
Apple	Healthy	Train	2000	1823
		Validation	250	218
		Test	250	214
	Defective	Train	2000	1459
		Validation	250	170
		Test	250	173
Mango	Healthy	Train	2000	1624
		Validation	250	203
		Test	250	221
	Defective	Train	2000	1478
		Validation	250	177
		Test	250	202

initially with 5k images each. However, after refining the dataset, the total number of apple images was reduced to 4,057 and the total number of mango images to 3,905.

The CNN models selected for this study are MobileNetV2 and VGG16, based on the state-of-the-art review.

MobileNetV2 has been specially designed for computer vision applications on devices with limited resources, such as mobile devices, which could be used by fruit inspectors, producers, or consumers in our case. This model stands out for its ability to significantly reduce the number of parameters and computational operations by using building blocks known as Inverted Residual Blocks, allowing it to maintain high performance in image classification, specifically in distinguishing between healthy and defective fruits. Additionally, MobileNetV2 employs a technique called Expansion Factor to control the number of channels in the intermediate layers of the network, facilitating the adaptation of the model to various tasks and computational resource requirements [19].

VGG16 is a deeply stacked model consisting of 16 layers of convolution and pooling, followed by three fully connected layers, and it is notable for its simplicity and relative depth. The model employs small-sized filters (3x3) in all convolutional layers, followed by max-pooling layers to reduce dimensionality. VGG16 has demonstrated good performance in image classification tasks on standard datasets, as is the case in our study [20].

C. Multi-Input architecture

The multi-input architecture using CNN models combines multiple streams of different types of input datasets, each processed by an independent CNN model, and as a final stage merges the learned features to perform a specific machine learning task [21]–[23]. The implemented multi-input architecture considers the utilization of two distinct input branches. In the first branch, an RGB dataset is employed, while in the second branch, a silhouette image dataset corresponding to each RGB image is utilized. Furthermore, independently trained models are deployed in each branch, namely MobileNetV2 and VGG16. Finally, the results obtained from each branch are unified through a feature fusion process, and a classification is achieved after a Multilayer Perceptron (MLP) layer. The

following figure 3 illustrates the methodology employed in this study.

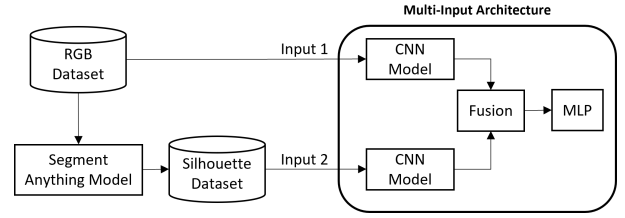


Fig. 3. Multi-Input architecture used for healthy and defective fruits classification.

IV. EXPERIMENTAL RESULTS

This section outlines the outcomes derived from evaluating Multi-Input architectures in conjunction with CNN models, following the methodology previously elucidated. For each fruit type, the entire dataset, including their respective silhouettes, was utilized to assess the CNN models. To perform the comparison with the study [14], the same fruits with their datasets and the types of defects they considered, namely rot, bruises, scabs, and black spots, are taken into account. It was noted that both models attained a 100% accuracy rate in fruit classification. Nevertheless, upon reevaluation of the CNN models using refined datasets, equally high outcomes were achieved. In this instance, only the MobileNetV2 model achieved a 100% accuracy rate.

The hyperparameters were established to train the architectures using input images sized 224x224. We employed the initial weights pre-trained with ImageNet for the CNN models. We opted for the Adam optimizer, with a low learning rate of 0.00001 and a batch size of 16. Additionally, a loss function designed for multiclass classification, categorical cross-entropy, was utilized, despite our classification being binary. Typically, training was configured to run up to epoch 60, yet it is noteworthy that due to achieving satisfactory results within the first 30 epochs, the number of epochs was reduced. As usual, the learned weights from the epoch exhibiting the best performance were selected as the best model.

A. Results with Apples

For the classification of healthy and defective apples, two CNN models were considered in training the Multi-Input architectures: MobileNetV2 and VGG16. Below, the training and validation plots for each model are presented (see Fig. 4 and Fig. 5).

Upon reviewing the training and validation plots of the Multi-Input architectures, it is evident that they consistently achieve a high level of accuracy. This performance can be attributed to the contribution of silhouette images, as the segmentation process enables the Multi-Input architectures to accurately discern whether the fruit is healthy or defective. Furthermore, both Multi-Input architectures were subjected to testing, with the results depicted in Fig. 6, Fig. 7, and Table II.

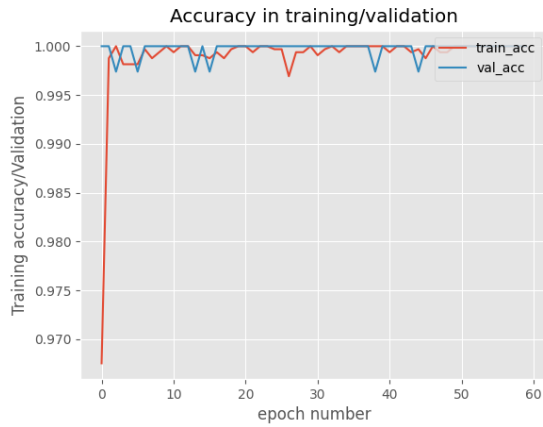


Fig. 4. Multi-Input architecture: Training/Validation of the MobileNetV2 model with Apples.

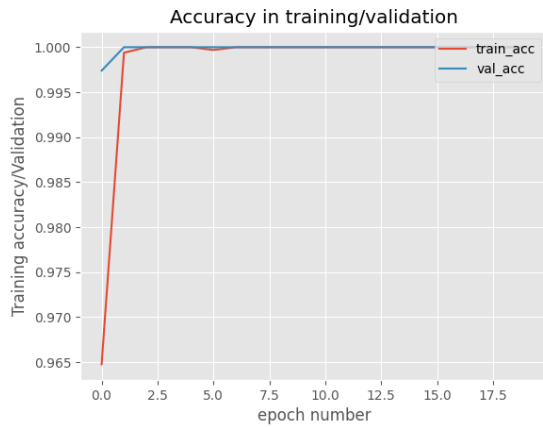


Fig. 5. Multi-Input architecture: Training/Validation of the VGG16 model with Apples.

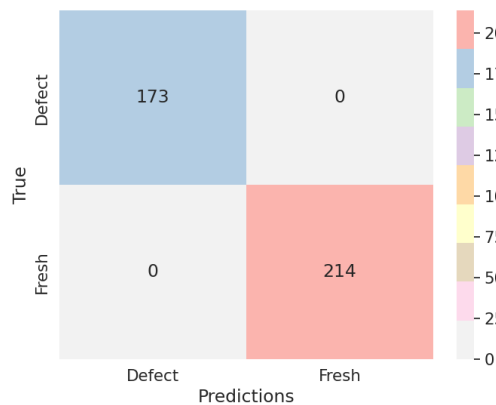


Fig. 6. Multi-Input architecture: Confusion Matrix of the MobileNetV2 model with Apples.

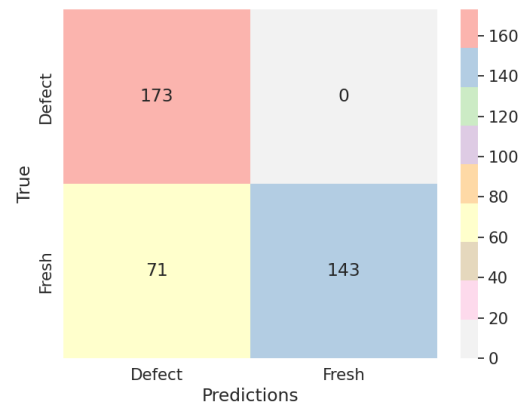


Fig. 7. Multi-Input architecture: Confusion Matrix of the VGG16 model with Apples.

During the testing phase of the Multi-Input architecture with the MobileNetV2 model, it became evident that all images of both healthy and defective apples were classified correctly, resulting in perfect accuracy. However, the Multi-Input architecture with the VGG16 model encountered challenges in accurately identifying images of healthy apples. This could be attributed to its design, which is optimized for classifying natural images containing rich details and textures. Consequently, it struggles when classifying silhouettes in binary images due to the loss of relevant information during conversion to a single channel and the lack of adaptation to the inherent simplicity of silhouettes.

B. Results with Mangoes

For the classification of healthy and defective mangoes, two CNN models were considered in training the Multi-Input architectures: MobileNetV2 and VGG16. Below, the training and validation plots for each model are presented (see Fig. 8 and Fig. 9).

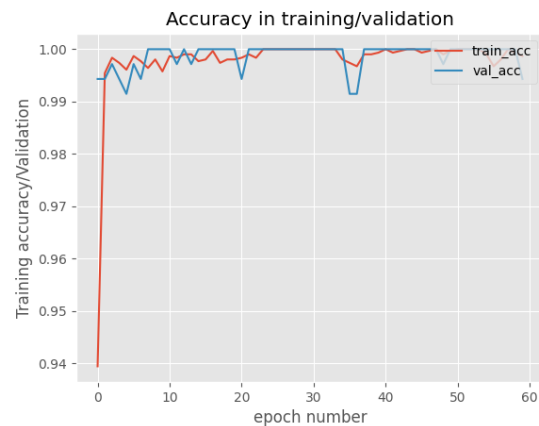


Fig. 8. Multi-Input architecture: Training/Validation of the MobileNetV2 model with Mangoes.

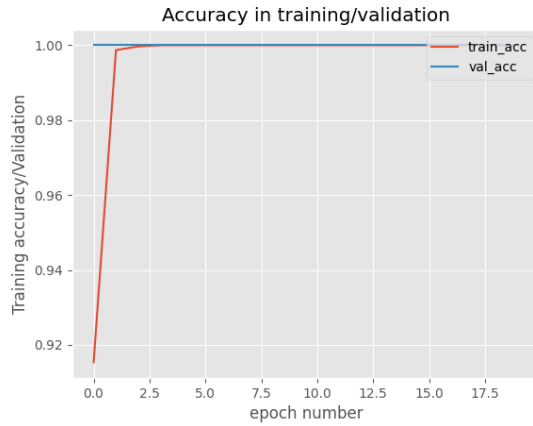


Fig. 9. Multi-Input architecture: Training/Validation of the VGG16 model with Mangoes.

The training and validation plots of the Multi-Input architectures for the classification of healthy and defective mangoes exhibit a high level of accuracy, similar to what was observed with apples. This performance can also be attributed to the contribution of silhouette images, demonstrating the applicability of this technique across various fruit types. Furthermore, both Multi-Input architectures underwent testing, with the results depicted in Figures 10, 11, and the Table II.

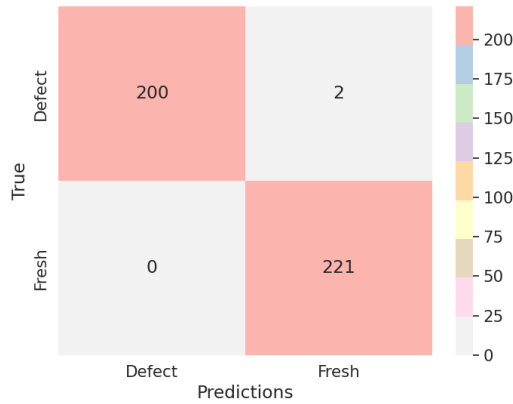


Fig. 10. Multi-Input architecture: Confusion Matrix of the MobileNetV2 model with Mangoes.

Upon reviewing the confusion matrices from the testing phase of the Multi-Input architecture using the MobileNetV2 model, it was observed that nearly all images of defective mangoes were classified correctly, and all images of healthy mangoes in the test dataset were accurately classified, resulting in high accuracy. Conversely, the Multi-Input architecture with the VGG16 model correctly classified almost all images of both healthy and defective mangoes.

Upon reviewing the metrics obtained by the models in Table II, it is evident that the best-performing model is MobileNetV2 (Multi-Input) for the classification of healthy

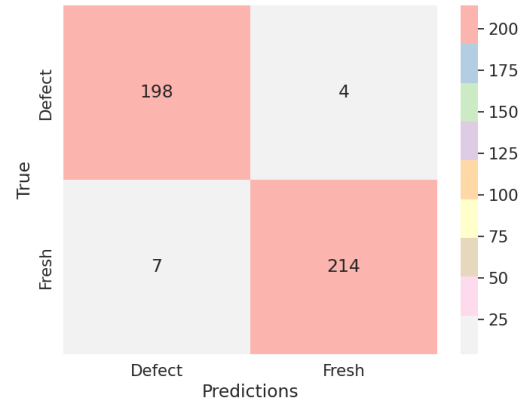


Fig. 11. Multi-Input architecture: Confusion Matrix of the VGG16 model with Mangoes.

and defective fruits. An accuracy of 100% was achieved for apples and 99.53% for mangoes. Comparing these metrics with those of the reviewed work [14], which achieved 97.50% accuracy for apples and 92.90% for mangoes using only the MobileNetV2 model, it is clear that the MobileNetV2 (Multi-Input) model implemented in this study achieves higher efficiency than the models reviewed in the state of the art. It is worth mentioning that the same real dataset from the reviewed work was used. Furthermore, the MobileNetV2 (Multi-Input) model is not only suitable for silhouette images but also for any type of segmented binary images.

V. CONCLUSIONS

Robust models were successfully implemented and investigated, surpassing the performance reviewed in state-of-the-art, accurately classifying healthy or defective fruits using a Multi-Input architecture that utilizes both RGB and silhouette datasets in each branch. The inclusion of a binarized dataset (silhouette) significantly improved the performance of the Multi-Input architecture, enabling more precise classification and optimal performance. After comparing several models, it was determined that the best performer, for both apples and mangoes, is the MobileNetV2 Multi-Input), achieving an accuracy of 100% for apples and 99.53% for mangoes. Furthermore, it is recommended for future work to implement this model and binary segmentation for the classification of other fruits, considering their healthy or defective state. In addition, the dataset used will be augmented to strengthen the accuracy and generalization of the models employed. Additionally, other types of architectures such as Siamese architectures will be evaluated.

ACKNOWLEDGMENT

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

TABLE II
RESULTS COMPARISON.

Fruits	Models	Accuracy	Precision	Recall	F1-Score
Apple	MobileNetV2 (Multi-Input)	1.0	1.0	1.0	1.0
	VGG16 (Multi-Input)	0.81653	0.8545	0.8341	0.8154
Mango	MobileNetV2 (Multi-Input)	0.9953	0.9955	0.9950	0.9953
	VGG16 (Multi-Input)	0.9739	0.9738	0.9743	0.9740

REFERENCES

- [1] M. Palumbo, M. Cefola, B. Pace, G. Attolico, and G. Colelli, "Computer vision system based on conventional imaging for non-destructively evaluating quality attributes in fresh and packaged fruit and vegetables," *Postharvest Biology and Technology*, vol. 200, p. 112332, 2023.
- [2] X. Liu, L. Wei, C. Miao, Q. Zhang, J. Yan, S. Li, H. Chen, and W. Qin, "Application of exogenous phenolic compounds in improving postharvest fruits quality: Classification, potential biochemical mechanisms and synergistic treatment," *Food Reviews International*, pp. 1–20, 2023.
- [3] Y. Lu, L. Harvey, and M. Shankle, "Survey and cost-benefit analysis of sorting technology for the sweetpotato packing lines," *AgriEngineering*, vol. 5, no. 2, pp. 941–949, 2023.
- [4] L. E. Chuquimarca, B. X. Vintimilla, and S. A. Velastin, "Banana ripeness level classification using a simple cnn model trained with real and synthetic datasets," in *VISIGRAPP (5: VISAPP)*, 2023, pp. 536–543.
- [5] G. Devika, C. S. R. Bandaru, R. V. Srinivas, N. Girdharwal, M. N. Devi *et al.*, "Real-time quality assurance of fruits and vegetables using hybrid distance based mkelm approach," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*. IEEE, 2023, pp. 728–734.
- [6] S. K. Chakraborty, A. Subeesh, K. Dubey, D. Jat, N. S. Chandel, R. Potdar, N. G. Rao, and D. Kumar, "Development of an optimally designed real-time automatic citrus fruit grading-sorting machine leveraging computer vision-based adaptive deep learning model," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105826, 2023.
- [7] G. Dhanush, N. Khatri, S. Kumar, and P. K. Shukla, "A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce," *Scientific African*, p. e01798, 2023.
- [8] G. Li and Y. Yang, "Smart vision for quality apple classification using surf-harris optimizing techniques," *Wireless Networks*, pp. 1–13, 2023.
- [9] H. H. C. Nguyen, A. T. Luong, T. H. Trinh, P. H. Ho, P. Meesad, and T. T. Nguyen, "Intelligent fruit recognition system using deep learning," in *International Conference on Computing and Information Technology*. Springer, 2021, pp. 13–22.
- [10] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fredes, and A. Valenzuela, "A review of convolutional neural network applied to fruit image processing," *Applied Sciences*, vol. 10, no. 10, p. 3443, 2020.
- [11] N. Gururaj, V. Vinod, and K. Vijayakumar, "Deep grading of mangoes using convolutional neural network and computer vision," *Multimedia Tools and Applications*, vol. 82, no. 25, pp. 39 525–39 550, 2023.
- [12] M. S. Ibrahim, S. A. Hamid, Z. Muhammad, N. A. M. Leh, S. Abdullah, S. J. A. Bakar, M. K. Osman, and S. Y. Fadhlullah, "Cnn comparative study for apple quality classification," in *2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 2022, pp. 53–58.
- [13] B. Buyukarikan and E. Ulker, "Classification of physiological disorders in apples using deep convolutional neural network under different lighting conditions," *Multimedia Tools and Applications*, vol. 82, no. 21, pp. 32 463–32 483, 2023.
- [14] R. Pacheco, P. González, L. E. Chuquimarca, B. X. Vintimilla, and S. A. Velastin, "Fruit defect detection using cnn models with real and virtual data," in *VISIGRAPP (4: VISAPP)*, 2023, pp. 272–279.
- [15] M. Agarwal, R. K. Kaliyar, and S. K. Gupta, "Differential evolution based compression of cnn for apple fruit disease classification," in *2022 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2022, pp. 76–82.
- [16] C. Pachón-Suescún, J. Pinzón-Arenas, and R. Jiménez-Moreno, "Fruit identification and quality detection by means of dag-cnn," *Int J Adv Sci Eng Inf Technol*, vol. 10, no. 5, pp. 2183–2188, 2020.
- [17] R. Nithya, B. Santhi, R. Manikandan, M. Rahimi, and A. H. Gandomi, "Computer vision system for mango fruit defect detection using deep convolutional neural network," *foods*, vol. 11, no. 21, p. 3483, 2022.
- [18] A. Dikshit, A. Bartsch, A. George, and A. B. Farimani, "Robochop: Autonomous framework for fruit and vegetable chopping leveraging foundational models," *arXiv preprint arXiv:2307.13159*, 2023.
- [19] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit image classification based on mobilenetv2 with transfer learning technique," in *Proceedings of the 3rd international conference on computer science and application engineering*, 2019, pp. 1–7.
- [20] C. C. Ukwuoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent advancements in fruit detection and classification using deep learning techniques," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–29, 2022.
- [21] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input cnn-gru based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 7, pp. 1461–1478, 2021.
- [22] A. Choudhary, R. K. Mishra, S. Fatima, and B. Panigrahi, "Multi-input cnn based vibro-acoustic fusion for accurate fault diagnosis of induction motor," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105872, 2023.
- [23] A. R. Mesa and J. Y. Chiang, "Multi-input deep learning model with rgb and hyperspectral imaging for banana grading," *Agriculture*, vol. 11, no. 8, p. 687, 2021.