From pixels to patterns: AI-driven image analysis in multiple domains

Javanmardi, S.

CHAPTER 3

Mulberry Ripeness Detection via Deep Learning

This chapter is based on the following publication:

S.-H. M. Ashtiani, S. Javanmardi, M. Jahanbanifard, A. Martynenko, and F. J. Verbeek, Detection of mulberry ripeness stages using deep learning models, IEEE Access, vol. 9, pp. 100380–100394, 2021.

3.1 Chapter Summary

In the previous chapter, we focused on using deep learning to classify different varieties of corn seeds. As we move to Chapter 3, we turn our attention to a vital part of agricultural production: determining the ripeness of mulberry fruit. This step is key for efficient postharvest management. The traditional method of manually sorting the fruit is prone to microbial spoilage and human error, leading to significant losses. Determining the ripeness of mulberry fruit is critical because it directly impacts the quality and safety of the fruit at the time of consumption. Ripe fruits are at their peak nutritional value and taste, making them more desirable for consumers. However, if fruits are not accurately sorted by ripeness, unripe or overripe fruits may reach the consumer, affecting the consumer's experience and potentially leading to health risks associated with spoiled fruit. Moreover, accurate ripeness detection can help in planning the distribution and sale of the fruits more effectively, reducing wastage due to spoilage. Efficient postharvest management, facilitated by accurate ripeness detection, can lead to better market prices, reduced losses, and higher profitability for farmers. It also ensures that consumers receive fresh, high-quality produce, enhancing food security and nutrition.

This chapter presents a novel automated sorting mechanism that employs computer vision and Convolutional Neural Networks (CNNs) to classify the ripening stages of mulberries with precision. A particular emphasis is placed on utilizing transfer learning to enhance the performance of CNN architectures like DenseNet, Inception-v3, ResNet-18, ResNet-50, and AlexNet, thereby improving classification accuracy while curtailing training expenses. The integration of these refined models into automated systems signifies a transformative leap forward for postharvest fruit management, promising to elevate both the efficiency and dependability of the sorting processes.

3.2 Introduction

Mulberry (Morus spp., Moraceae family) is one of the fruit species, widely distributed from temperate to subtropical zones of the northern hemisphere to the tropical zones of the southern hemisphere (Saracoglu, 2018), (Chen et al., 2015). Among the 24 known Morus cultivars, white mulberry (Morus alba L.), black mulberry (Morus nigra L.), and red mulberry (Morus rubra L.) are the most cultivated species in the world (Nguyen and Nguyen, 2018), (Jelled et al., 2017), (Sánchez-Salcedo et al., 2015). They are an excellent source of many nutritive compounds such as vitamins, minerals, polysaccharides, fatty acids, and amino acids (Nguyen and Nguyen, 2018), (Lou et al., 2012), as well as phenolic compounds including carotenoids, anthocyanins, flavonoids, and phenolic acids (Huang et al., 2020), (Calín-Sánchez et al., 2013) with the health-promoting and pharmacological effects, such as anti-cancer, anti-cholesterol, anti-inflammation, anti-diabetic, anti-aging, antioxidant, anti-obesity and neuroprotection (You et al., 2018), (Han et al., 2017). In addition to high nutritional value and bioactivity, the low acidity and very sweet taste of white mulberry and the slightly acidic flavor and attractive dark color of black mulberry led to a rapid increase in their production and consumption (Nayab et al., 2020), (Donno et al., 2015).

The fully ripened fruits are usually consumed either fresh or used as ingredients in marmalade, tea, vinegar, wine, juice, jam, ice cream, jelly, syrup, food colorant, natural dyes, dried fruits, and other food and cosmetic products (Jelled et al., 2017), (Sánchez-Salcedo et al., 2015), (Rohela et al., 2020). Fruit ripening is accompanied by the color change because of the pigment concentration in the fruit skin (Jelled et al., 2017), (Lou et al., 2012). During growth period (25 to 30 d), mulberry skin color changes from green (unripe) to white, red, purple-black (fully ripe) (Tabakoglu and Karaca, 2018). Mulberry is a non-climacteric product, therefore, its harvest in the suitable ripening stage is highly significant from a nutritional and economic perspective (Nayab et al., 2020).

Mulberry harvest operation is usually carried out by spreading a sheet under a tree and shaking the branches mechanically or manually (Assirelli et al., 2019), (Tabakoglu and Karaca, 2018). Due to alternate ripening patterns (Figure 3.1), harvest includes a mixture of fruits at various ripening levels (Tabakoglu and Karaca, 2018), (Azarmdel et al., 2020). This inconsistency of the product negatively affects its commercial value and marketability (Yang et al., 2017). On the other hand, fresh mulberry is available only in the short term and it is hardly commercialized. Fresh, fully ripened fruits decay fast after harvesting due to the soft structure, high moisture content, respiration, growth, and proliferation of microorganisms on the surface (Chen et al., 2015), (Nguyen and Nguyen, 2018), (Han et al., 2017). Although their shelf-life can be extended by up to six weeks at cold storage, they cannot be preserved for a long time at ambient temperature (Hojjatpanah et al., 2011). Therefore, further processing is required to extend

Figure 3.1: Distribution of white (a) and black (b) mulberries with different ripening levels on a branch (Adapted from https://www.trees.com/mulberry-trees).

the availability and utilization of health benefits. Unfortunately, fruits with a low ripening index are discarded and wasted during quality control (Nayab et al., 2020). As a result, the classification of fruits from a ripening stage perspective is essential for initial sorting. At this stage, unripe or unmarketable fruits could be separated from the batch for the next processing, while the best quality fresh fruits are delivered to consumers (Saracoglu, 2018), (Yang et al., 2017).

The popularity of mulberry fruits among consumers and processing companies increases due to their recently discovered nutritional and nutraceutical value (Li et al., 2020). Since the nutritional and functional compounds decrease and/or increase during ripening, a proper classification approach would extend the range of applications of mulberries in food and pharmaceutical industries (Lee and Hwang, 2017).

3.2.1 Problem Statement

Nowadays, the mulberry industry has significant problems with mulberry classification. Usually, identification of the ripening degree of mulberry fruits is done by trained personnel through visual inspection of fruit color. This process is exhausting, time-consuming, subjective, and costly (Castro et al., 2019), (Wan et al., 2018). Alternatively, identification of ripening stages can be done by chemical or physicochemical methods (Zhang et al., 2018). Although these methods have better accuracy, they are time-consuming, costly, destructive, and sometimes require complex analytical equipment. They allow quantifying total flavonoids, anthocyanins, and total soluble solids (TSS), but are limited to a certain amount of samples, which is not suitable for automatic sorting systems (Zhuang et al., 2019), (Pereira et al., 2018).

Therefore, the need for an automated non-destructive sorting system to increase fruit utilization and supply high-quality mulberry products to consumers is indispensable. Recently, smart analytical tools such as spectroscopy and spectral imaging, electronic noses, and computer vision have been utilized to evaluate ripening levels of fruits (Sabzi et al., 2019), (Pourdarbani et al., 2020). Some of them use machine learning and pattern recognition techniques (Pereira et al., 2018), (Cárdenas-Pérez et al., 2017), (Harel et al., 2020). Table 3.1 summarizes the relevant research for various horticultural products using traditional machine learning algorithms and handcrafted features.

Despite the progress in the classification, these applications are limited because of the following reasons:

- The complexity of these methods due to the manual choice of features (Zhou et al., 2019).
- Inability to differentiate the subtle differences between subordinate classes (Zhang et al., 2018).

3.2.2 Novelty

Despite recent advancements in computer vision, only one research has been conducted to classify the ripening stage of mulberries (Azarmdel et al., 2020). In this work, traditional handcrafted features and two traditional machine learning classifiers, i.e., ANN and SVM, have been used. The complexity and low accuracy of these models motivated us to test deep learning models for the same purpose. Our study is the first time attempt to solve the challenge of accurate detection of the mulberry ripening stage by using deep learning. This approach is more accurate, enabling the classification of 4 ripening stages, which is more challenging compared to the previous research (3 ripening stages). It was demonstrated that deep learning can automatically extract features without human intervention and accurately classify the ripening stage of mulberries. Another novelty of this chapter is the calculation of the time required to complete the classification process, which plays a key role in the design of a smart mulberry sorting system. This is in contrast to (Azarmdel et al., 2020), which did not examine the classification time.

3.2.3 Related Work

So far, the application of deep learning models in precision agriculture has shown advantages for automatic feature extraction and learning, transfer learning, quick adaptation to a new problem, dealing with heterogeneous big data, and obtaining higher accuracy and excellent performance (Zhou et al., 2019), (Javanmardi et al., 2021). Convolutional neural networks (CNN) and their derivatives have shown to be among the most successful techniques in image classification and recognition (Zhang et al., 2018), (Ge et al., 2019).

Recently, several studies have applied CNNs for ripeness classification by analyzing RGB images of fruits. For example, Zhang et al. (Zhang et al., 2018) designed a CNN structure for fine-grained classification of banana maturity. The proposed CNN achieved a 95.6% classification accuracy, which was higher than conventional strategies such as Gabor + SVM, Wavelet + SVM, Wavelet + Gabor + SVM, and combined features + SVM approaches. Another CNN model was developed to classify five stages of tomato maturity based on skin color, i.e., green, light pink, pink, light red, and red. The model was able to detect 100 images in less than 0.01 s with 91.9% accuracy.

Halstead et al. (Halstead et al., 2018) developed a robotic vision system for classifying the sweet pepper ripeness into three classes (unripe, partially ripe, and ripe) based on the parallel Faster R-CNN technique. The framework yielded a classification accuracy of 82.1%. Ge et al. (Ge et al., 2019) employed the Mask Region-CNN model to detect and classify different ripening levels (raw, pink, and ripe) of strawberries in farm conditions. Mohtar et al. (Mohtar et al., 2019) adopted an Inception-v3 model to classify six stages of ripening of mangosteen fruit with a classification accuracy of 91.9%. Liu et al. (Liu et al., 2020) proposed a modified densely-connected convolutional network (DenseNet), aiming to detect the maturity of tomatoes in complicated images. The detection rate of the improved DenseNet network was superior compared to the residual network (ResNet), DenseNet, and single-shot detector (SSD) frameworks.

Huang et al. (Huang et al., 2020) developed a fuzzy Mask R-CNN model to classify the ripeness levels of cherry tomatoes into 4 categories. Their model was

Table 3.1: Comparative analysis of classical machine vision techniques applied for ripening stage classification.

Product	Feature	#classes	Data Analysis	Accuracy%	References
Red-white mulberries	Color, Texture, Shape	3	ANN, SVM	98.26-99.13	Azarmdel et al. (Azarmdel et al., 2020)
Cape gooseberry	Color	7	ANN, DT, SVM, KNN	85.90-92.65	Castro et al. (Castro et al., 2019)
Tomato	Color	3	BPNN	99.31	Wan et al. (Wan et al., 2018)
Banana	Color, Texture, Shape	4	NB, LDA, SVM	82.60-100	Zhuang et al. (Zhuang et al., 2019)
Papaya	Color	3	RDF	94.3	Pereira et al. (Pereira et al., 2018)
Apple	Color	4	Hybrid ANN-GA	97.88	Sabzi et al. (Sabzi et al., 2019)
Apple	Color	4	SVM, KNN, ANN-GA, ANN-PSO, ANN-FA	92.56-97.16	Pourdarbani et al. (Pourdarbani et al., 2020)
Apple	Color	3	MDA	100	CÃąrdenas-PÃľrez et al. (Cárdenas-Pérez et al., 2017)
Yellow-red sweet peppers	Color, Shape	4	RF, LR	89.5-97.3	Harel et al. (Harel et al., 2020)

able to achieve an accuracy of 98%. In another study, Ramos et al. (Ramos et al., 2021) attempted to classify the ripening stage of two grape cultivars. In their work, they employed two CNN architectures containing 10 convolutional layers and VGG-19. The authors reported that changing the number of ripening classes from three to eight would improve the classification accuracy from 65.30 to 93.41%. To detect different maturity levels of date fruit, Faisal et al. (Faisal et al., 2020) used three pre-trained architectures: VGG-19, Inception-v3, and NASNet. They achieved a correct classification rate greater than 99%.

3.2.4 Main Contributions and Chapter Structure

Our study fills the gap in the knowledge, makes the following significant contributions:

- To best of our knowledge and literature survey, this research is the first attempt to determine the ripening stages of mulberries using CNN-based deep learning architecture.
- The performance of different CNN architectures including DenseNet, Inception-v3, ResNet-18, ResNet-50, and AlexNet has been evaluated for this classification problem.
- To minimize the number of training images and reduce the training time, CNN models have been fine-tuned and optimized on our target data sets.

The remainder of this chapter is structured as follows: In Section 3.3, the materials and methods used in the study including the computer vision system and detection models, are presented and explained. In Section 3.4, the results of the testing of proposed frameworks are reported. In Section 3.5, the main results are discussed. In Section 3.6, the design of a computer vision-based sorting system for automatic detection of mulberry ripening stages is proposed. Finally, in Section 3.7, conclusions and suggestions for further studies are presented.

3.3 Materials and Methods

3.3.1 Dataset Composition

Fruits of two mulberry genotypes, i.e., black and white mulberry, grown under the same environmental conditions were collected at four successive ripening stages, i.e., unripe, semi-ripe, ripe, and overripe (Figure 3.2). These samples were hand-

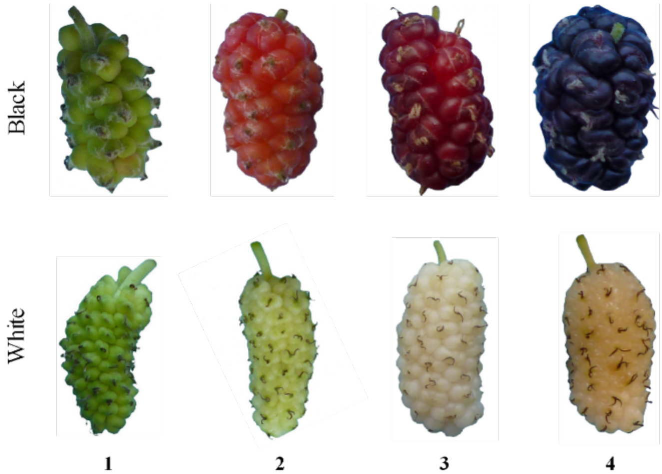

Figure 3.2: Different ripening stages of black and white mulberries: (1) unripe, (2) semi-ripe, (3) ripe, and (4) overripe.

harvested by an expert in the morning from a commercial orchard located in Shahriar (35°36′43″ N, 51°7′27″ E), Tehran Province, Iran, in the period May–June 2020. Fruits packed in plastic punnets were immediately transported to the experimental laboratory under refrigeration at approximately 5 °C. Only healthy fruits without disease and mechanical damages were selected. In total, 1000 samples of mulberry (250 samples per ripening stage) were used for imaging under controlled conditions at 20 ± 1 °C and $60 \pm 5\%$ RH. The TSS value of the juice produced from the fruits at each ripening stage was measured using a hand-held refractometer (Master-53PT, Atago, Japan, ±0.2% accuracy). The measurements were performed ten times for each class. The TSS values of mulberry fruits at different ripening stages are presented in Table 3.2.

3.3.2 Acquisition of Images and Preprocessing

Image acquisition was carried out by a digital camera (Nikon D3200 24.2 MP CMOS, Japan) placed at 25 cm from the samples in a 30L cm × 30W cm × 40H cm illumination chamber, equipped with an 18 W circular fluorescent lamp with a color temperature of 6500 K. Each picture was acquired with the blue background of the chamber; no zoom nor flash were used. Images were stored in RGB format with a resolution of 300 dpi and an image size of 4320×3240. Image segmentation

Table 3.2: Changes in total soluble solids (TSS; °Brix) of mulberry fruits during different stages of ripening.

Ripening stage	Genotype	
	White	Black
Unripe	7.2 ± 0.4	6.7 ± 0.3
Semi-ripe	10.6 ± 0.6	9.4 ± 0.5
Ripe	13.9 ± 0.2	12.1 ± 0.8
Overripe	17.1 ± 0.7	14.8 ± 0.4

Mean TSS values and their standard deviation over 10 measurements

was accomplished by the modified unsupervised segmentation algorithm introduced by Aganj et al. (Aganj et al., 2018). This procedure includes converting RGB images into other color spaces including CIE Lab, HSV, and YCbCr. Within these color spaces, the strongest contrast between the mulberries and background was obtained in YCbCr color space.

3.3.3 CNN Models

A typical structure of a deep CNN is composed of input and output layers, as well as multiple hidden layers.

The hidden layers of a CNN are generally made up of convolutional, pooling, activation, and fully connected layers and in some cases a Softmax layer (Too et al., 2019), (Momeny et al., 2020). The CNNs have as a common characteristic that they can extract the features automatically from data and visualize the extracted features. For example, Figure 3.3 shows visualization features of the first, intermediate, and last convolutional layers of the ResNet-18 model, as well as the activation of these features. The first convolutional layers mainly extract the primary features like colors and edges. The filters in the intermediate layers mostly contain texture information which is made of a combination of edges and colors. With the deepening of layers, their outputs become increasingly abstruse and less visually interpretable.

Along with ResNet-18, four other well-known CNNs, such as AlexNet, Inception-v3, ResNet-50, and DenseNet, have been studied. These networks have been successfully used for a range of different image recognition tasks, such as leaf disease classification ResNet by Deeba and Amutha (Deeba and Amutha, 2020)

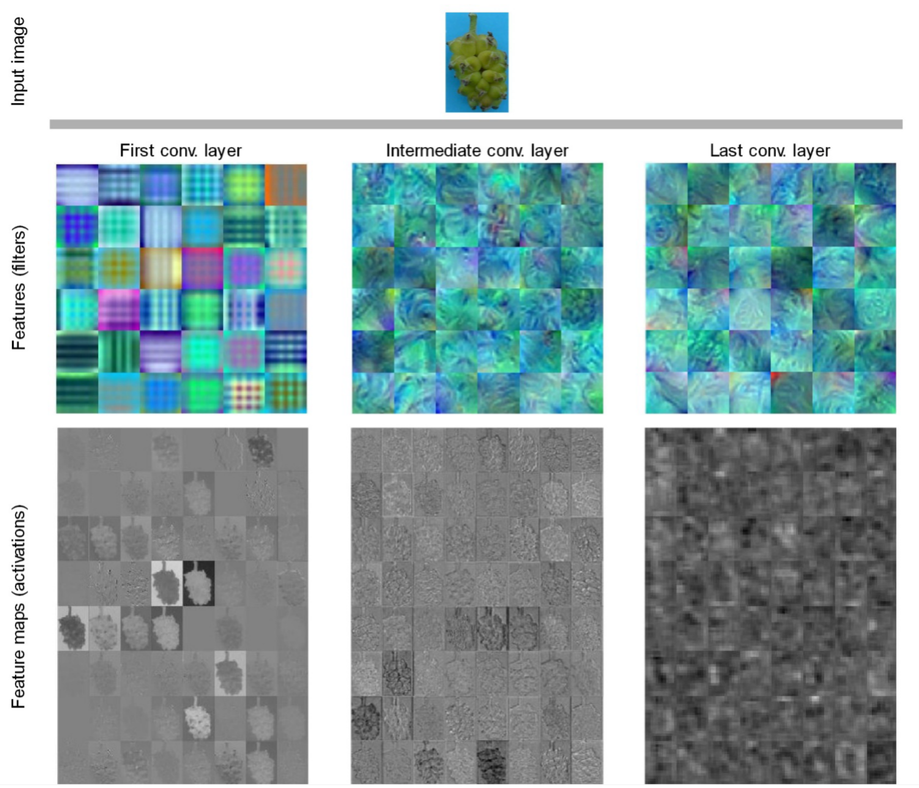

Figure 3.3: An example of visualization results of the ResNet-18 model. The figure shows the output features of the first, intermediate, and last convolutional layers at the top as well as feature maps of these features at the bottom.

and Inception-v3 by Qiang et al. (Qiang et al., 2019), remote sensing (AlexNet, ResNet-34, ResNet-50, ResNet-101 ResNet-152, VGG-16, VGG-19 and DenseNet-121 by Rohith and Kumar (Rohith and Kumar, 2020) and freshwater fish detection (DenseNet by Wang et al. (Wang et al., 2020). Further, a short explanation of each CNN included in our study is provided hereafter. The workflow of our research with respect to testing of CNNs for the new task of classification of mulberry ripeness is illustrated in Figure 3.4).

1) AlexNet

AlexNet was introduced by Alex Krizhevsky (Krizhevsky et al., 2012) to compete in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC2012). The original design utilized two graphics processing units (GPUs) to speed up the

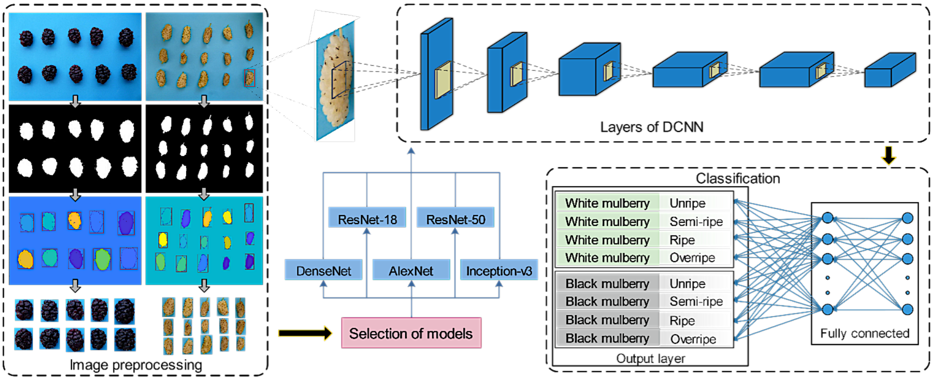

Figure 3.4: Schematic presentation of the workflow for classification of mulberry fruits.

training, but in this study, the single GPU processing version is used as it is more efficient with the newer GPUs. The AlexNet includes five convolutional and three fully-connected layers to process 227 × 227 pixels of images. In the first two fully connected layers, the Dropout regularization technique was applied to reduce the overfitting. Rectified Linear Units (ReLU) was utilized for all the hidden layers and Softmax for the output layer as the activation functions.

2) ResNet

He et al. (He et al., 2016) introduced this CNN architecture to push the depth of convolutional networks to its limits. Due to a network-in-network (NIN) architecture, ResNet is theoretically capable of having an infinite depth without losing accuracy. In practice, it can have up to 152 layers by stacking residual blocks throughout the network. The NIN networks use blocks that have few convolutional layers but more complex structures (known as micro neural networks). These blocks help the whole network to extract better features by focusing on a smaller receptive field, instead of the usual convolutional networks, which scan the input image using linear filters (Lin et al., 2013). ResNet has many stacked residual blocks, including a set of convolution and pooling layers. Although it has a similar architecture to AlexNet, it is about 20 times deeper due to the overcoming of the so-called degradation problem. ResNet has several implementations with different depths. In this study, ResNet-18 and ResNet-50 have been used.

3)Inception-V3

In 2015, Google introduced a network called GoogLeNet, also known as inception-v1

(Szegedy et al., 2015) in order to achieve the performance of a deep network with a light-weight structure (Emmert-Streib et al., 2020). Inception-v1 has different kernel sizes (1 ×1, 3 × ×3, 5 ×5) to extract feature maps in different scales, and by stacking them, the model can extract more features in total. This also reduces the parameters and accordingly reduces the computation (Qiang et al., 2019). Inception-v3 breaks down a large-scale convolution kernel into smaller convolution kernels (for instance breaks 3×3 kernels into two (1×3, 3×1) kernels). In this manner it contributes to further reduce network parameters and results it faster to run without sacrificing overall performance. At the same time this enables to extend the depth of the network (Emmert-Streib et al., 2020).

4)Densenet
In 2017, the idea of densely connected CNNs was proposed by Huang et al. (Huang et al., 2017). This architecture is introduced to solve a notorious problem regarding very deep networks known as the vanishing-gradient. The layers in DenseNet are connected to every other layer feed-forwardly and the feature-maps of each layer are used as inputs into all subsequent layers. This means that for a given network with n layers, there are 2n(n + 1) direct connections between each layer and its subsequent layers, to compare with traditional CNNs with n layers which have n connections. In addition to mitigating the vanishing-gradient problem and reducing the number of parameters, DenseNet strengthens feature propagation and encourages feature reuse.

3.3.4 Fine-Tuning The Models

A transfer learning approach was utilized to benefit from the pre-trained network by adjusting its parameters to our data set; this procedure is also known as fine-tuning. Fine-tuning is faster than training from scratch as a pre-trained network already has established weights. These weights are the result of learning over a data set (usually ImageNet) and help the network to train the features faster (Too et al., 2019). In order to realize this, the convolutional layers were frozen and the dense layer after those layers was trained. Furthermore, the last fully connected layer of networks was modified to have four output classes according to four levels of ripeness for either black or white mulberry respectively. If both data sets were combined, the last connected layer was modified to eight outputs for the genotype and ripeness detection. Given that one of the most obvious approach to avoid overfitting is initializing all the weights of components in CNNs to a pre-trained model, ImageNet pre-trained CNNs have been used in all the experiments. It improves the generalization and the performance of the model (Vinyals et al., 2016).

3.3.5 Software And Hardware Platform

All these networks are implemented in the MATLAB Deep Learning Toolbox (MATLAB R2020b, Mathworks Inc.) and Python 3.6. The parameters for each network are summarized in Table 3.3. The training was done using a machine with AMD Ryzen 9 3900 12cores/24thread 3.7GHz CPU, 128 GB DDR4 RAM, and a GeForce RTX 2080Ti GPU card with 11GB memory. The machine was installed with 2 GPUs but for the experiments only one was used.

Table 3.3: Specific parameters of the models in the evaluation.

Parameters	**AlexNet**	**Inc-v3**	**ResNet18**	**ResNet50**	**DenseNet**
Depth	8	48	18	50	201
Image size	227x227	299x299	224x224	224x224	224x224
optimizer	SGDM	ADAM	ADAM	ADAM	SGDM
Loss function	cross ent.	cross ent.	cross ent.	cross ent.	cross ent.
Batch size	64	64	64	64	64
Learn Rate(LR)	0.001	0.001	0.001	0.001	0.001
LR drop factor	0.1	0.1	0.1	0.1	0.1
LR drop period	10	10	10	10	10
Momentum	0.9	0.9	0.9	0.9	0.9
Gradient Thresh.	L2norm	L2norm	L2norm	L2norm	L2norm
Parameters	61 M	23.9 M	11.7 M	25.6 M	20 M

SGDM: Stochastic gradient descent with momentum; Adam: Adaptive momentum estimation; M: Millions

3.4 Results

3.4.1 Performance Evaluation

The performance of the selected deep learning models for the recognition and classification of mulberry fruits was evaluated based on multiple indicators: training accuracy, validation accuracy, training loss, and validation loss in each epoch. Training accuracy is a measure of model correctness during the training phase, whereas validation accuracy is defined as the percentage of test data truly classified by the trained model. The cross-entropy error was used as the loss function. A training-validation strategy for training and testing the classifier's performance was developed.

The image data set was split into two independent groups: 70% for training and 30% for validation of the trained model. The reason for splitting the data set into two subsets is that in small data sets, the additional split might lead to a smaller training set which may be exposed to overfitting (Féré et al., 2020). To provide enough data for training, the validation set was used to assess the performance of the models. In this regard, 5-fold cross-validation was involved to tune model hyperparameters. As a general rule, the higher the number of iterations, the higher the detection accuracy. However, after a certain number of epochs, the accuracy of network recognition is not increasing and sometimes even reducing (Wu et al., 2020).

As a result, the best detection accuracy can be achieved by selecting an optimal number of epochs for training the CNN models (Suh et al., 2018). In our experiments, the model accuracy and the loss function were monitored during the training process. Each of the experiments ran for a total of 100 epochs, where the number of epochs was defined as the number of times the network had to cycle through the data set. One hundred training epochs were enough since the validation accuracy and loss function demonstrated the best results around 20 to 75 epochs. This is further discussed in detail in later subsections. Another criterion, considered in the model evaluation, was classification time, calculated as the time required to classify all the validation samples by the trained algorithm.

3.4.2 White Mulberry Classification Results

In this section, a quantitative assessment of the deep learning models for the classification of white mulberry ripeness from images is presented. The classification performance of AlexNet, DenseNet, Inception-v3, and ResNet with 18 and 50 layers have been compared based on the same set of metrics. Figure 3.5a-d illustrates losses and accuracies of training and validation of the chosen deep learning models during the training procedure. As can be deduced from Figure 3.5a and Figure 3.5c, all models achieved high accuracy in both the training and the validation phase, respectively. ResNet-50 achieved the maximum classification accuracy without overfitting in 60 epochs. After the 60th epoch, the curves of training accuracy and

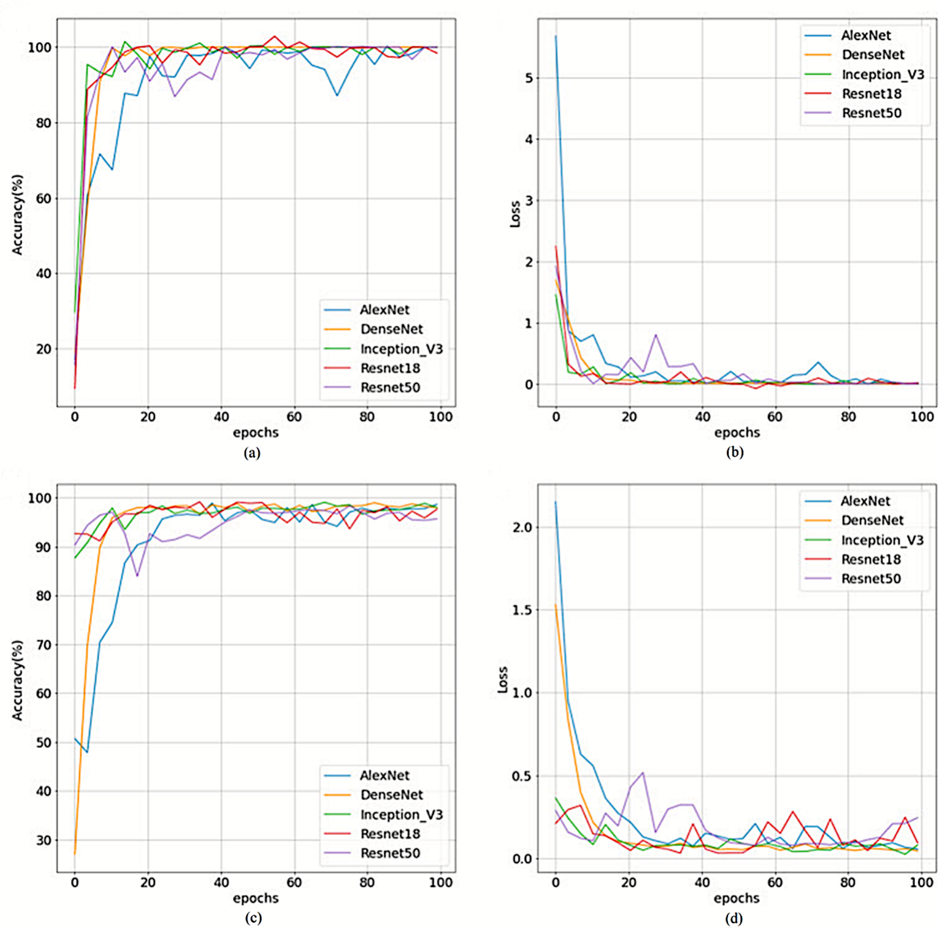

Figure 3.5: The behavior of the training (a, b) and validation (c, d) accuracies and losses of the white mulberry classification models.

loss reached a plateau and no significant change was observed in both loss function and accuracy (Figure 3.5a and Figure 3.5b). As seen in Figure 3.5c and Figure 3.5d, significant fluctuation of validation accuracy and loss curves started after the 83rd epoch, which is the sign of overfitting.

The behavior of Inception-v3 was similar to ResNet-50, converging at around epoch 60. In the case of ResNet-18, the loss and accuracy of the training set did stabilize after about 74 epochs, while the accuracy and loss of the validation set started to converge at around the same epoch with tiny fluctuations.

The DenseNet reached its optimal performance at epoch 34. After this point, the training and validation accuracy and loss curves started to become constant.

In AlexNet, the training procedure converges while attaining reasonable accuracy and loss after around 38 epochs (Figure 3.5a and Figure 3.5b). Similarly, the validation accuracy and loss curves saturated after the same epoch. It is worth noting that the very close results for training and validation accuracy indicate that the overfitting did not happen during the training process (Mumtaz and Qayyum, 2019).

The overall accuracy and loss of the five models for ripening classification of white mulberry are presented in Figure 3.6a and Figure 3.6b, respectively. Overall, all models achieved an accuracy of more than 96%, with DenseNet having the highest accuracy of 98.67% and the lowest loss of 0.0497. On the other hand, the highest loss of 0.182 and the lowest classification accuracy of 96.33% were obtained with ResNet-50. Figure 3.7 shows the classification time of the five CNN models.

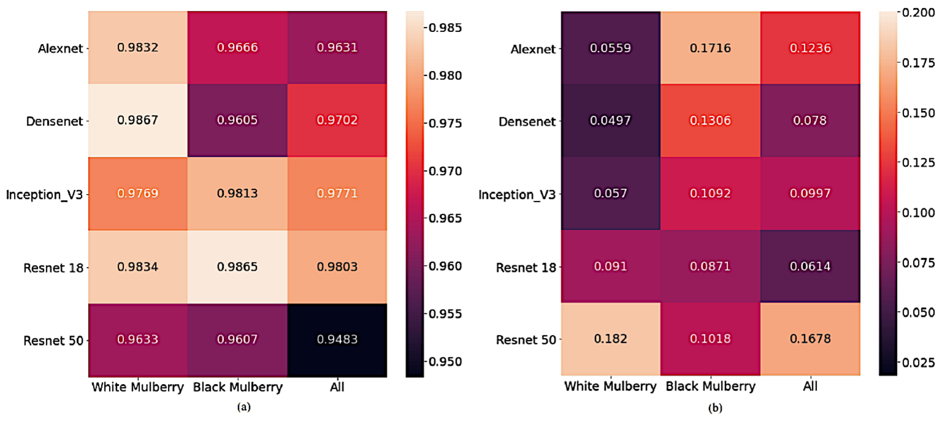

Figure 3.6: Comparison of overall accuracy (a) and loss (b) of proposed models.

It follows that the DenseNet, despite having the highest classification accuracy, required the longest classification time compared to other models. AlexNet and ResNet-18 needed significantly less time for classification than others, with AlexNet faster than ResNet-18 by a very small margin. These results for classification time can be considered reasonable since DenseNet contained the largest number of layers, whereas AlexNet had the fewest number of layers (Suh et al., 2018). It can be concluded that among the compared networks, AlexNet was the best

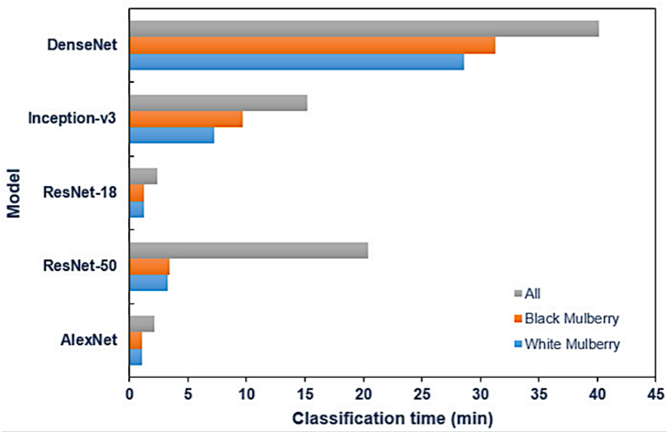

Figure 3.7: Classification time of proposed models.

because of overall accuracy above 98%, very close to DenseNet. Although the loss was slightly higher, it was 26 times faster than DenseNet.

3.4.3 Black Mulberry Classification Results

In this section, the performance of deep learning models for the classification of the black mulberry ripening stage from images is evaluated. The graphical representation of the training and validation accuracy and loss of the classifiers for each experimental run is shown in Figure 3.8a-d. In the case of the ResNet-50 model, we observed that with the number of the training epochs the algorithm gradually converged and the best results were achieved for 71 epochs (Figure 3.8 a and Figure 3.8b). At the beginning of the training, the classification accuracy of the algorithm was relatively low and then gradually improved. The high accuracy rate and low loss rate were achieved after about 20 epochs and convergence was achieved in approximately 70 epochs. With increasing the number of epochs, the training and validation losses decreased from about 2.6 to 0.0002 and 0.4 to 0.09, respectively. As displayed in Figure 3.8a and Figure 3.8c, the training and validation accuracy increased from about 23% to 100% and 86% to 98% with the progression in training epochs, respectively.

In Inception-v3, the training accuracy increased with the number of epochs and then stabilized after the 75th epoch. After this epoch, the validation accuracy began to decline. Likewise, the training loss became stable after 75 epochs and

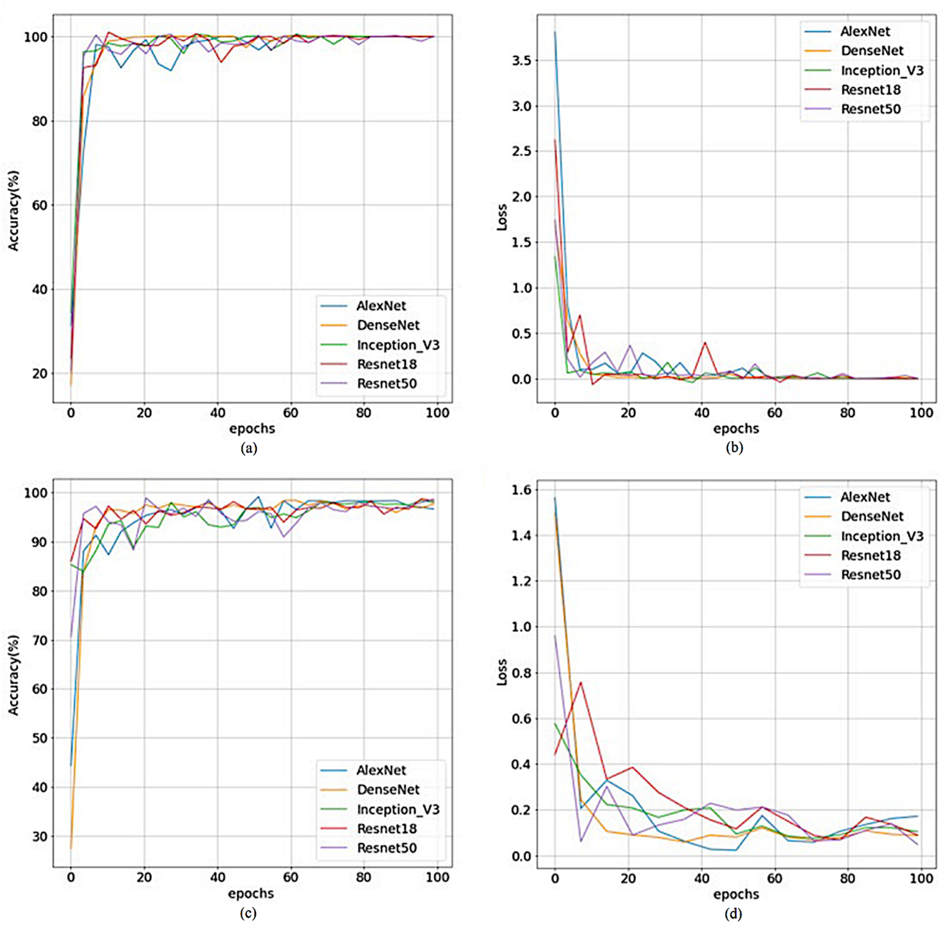

Figure 3.8: The behavior of the training (a, b) and validation (c, d) accuracies and losses of the black mulberry classification models.

the validation loss began to increase (Figure 3.8b and Figure 3.8d), implying that the model performed better on the training data set than on the validation data set. The training and validation losses of the model reduced from almost 1.3 to 0.00008 and 0.6 to 0.1 at the end of the training, respectively.

For DenseNet architecture, the training accuracy reached saturation at 20 epochs, but the validation accuracy fluctuated between 95% and 99% starting from about the 30th epoch. The training loss for DenseNet rapidly decreased from the first to twenty epochs and then became steady, while the validation loss values rapidly reduced from the first to fifteen epochs and then reasonably stabilized although fluctuating between 0.06 and 0.15. As illustrated in Figure 3.8a and b, the training of the AlexNet model stopped after 60 epochs as the training accuracy and loss started to plateau. After this epoch, the validation accuracy of the model decreased and the loss parameter increased (Figure 3.8d). A comparison of the results in Figure 3.6a and b depicts that all proposed CNN models can efficiently classify the ripening stages of black mulberry. For all scenarios, the training loss and accuracy were higher or approximately equal to the validation loss and accuracy, indicating that the networks were able to generalize well without overfitting.

The ResNet-18 model showed the best performance in both accuracy and loss function. The accuracy of ResNet-18 was 98.65%, which is 2.6% higher than that of DenseNet. Although DenseNet performed the best in classifying ripening levels of white mulberry, it was the relative worst in classifying black mulberry.

Compared with the other CNN models, the AlexNet had the highest loss value of 0.1716, which is twice higher than that of ResNet-18. As shown in Figure 3.7, classification time varied from 1.05 to 31.25 min with AlexNet requiring the shortest classification time and DenseNet requiring the longest classification time because of an extensive number of layers (Wu et al., 2020). Although AlexNet was the champion in the classification time, the ResNet-18 required only 8 seconds more, which is negligibly small compared to the overall classification time. Therefore, by considering all the aforementioned results, ResNet-18 is considered the best option among all networks for classifying the ripening stage of black mulberry.

3.4.4 Classification Results Of Combining Both Genotypes

The training and validation loss and accuracy curves of different models for the classification of both genotypes with the ripening stage are shown in Fig. 9a-d and explained in detail hereafter.

ResNet-50: The best performance of the ResNet-50 model was reported around epoch 63 because of the greatest training accuracy and the lowest training loss (Figure 3.9a and Figure 3.9b). The training loss and accuracy saturated after 63 epochs, so the slope of both training plots was close to zero. The accuracy of the validation data set increased continuously up to epoch 63 but then started to decrease. In the same manner, the loss decreased continuously up to epoch 63 but then started to increase (Figure 3.9c and Figure 3.9d).

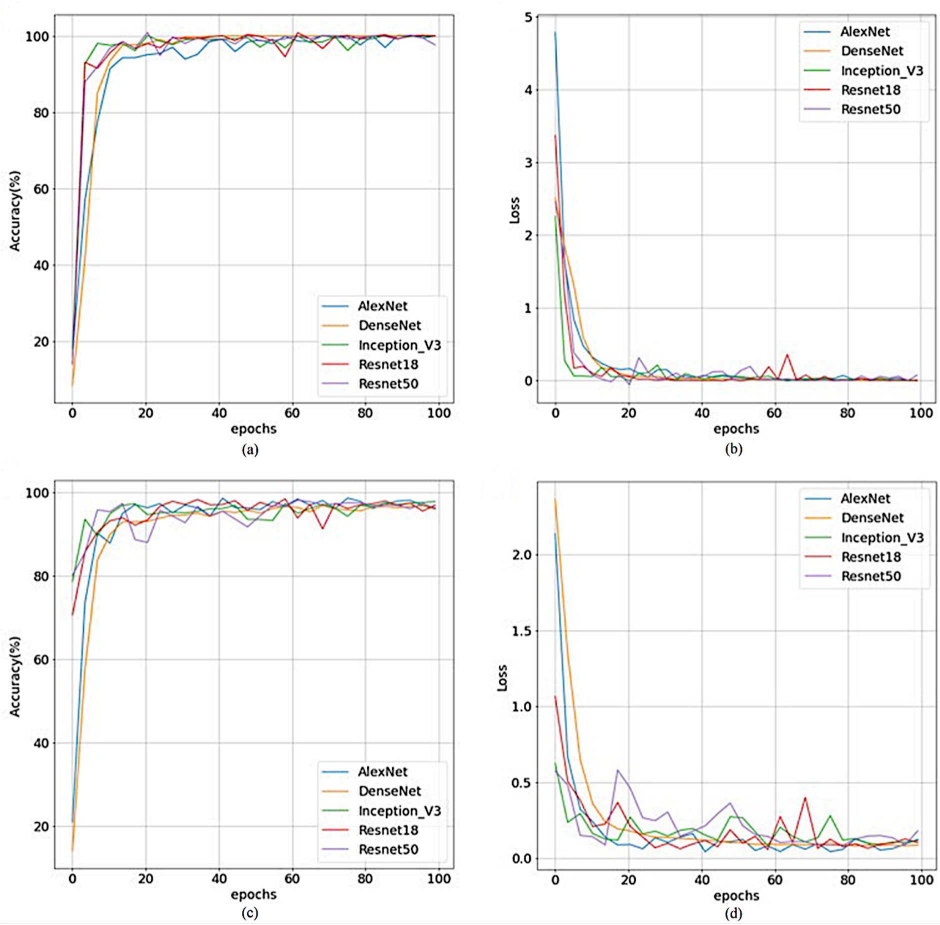

Figure 3.9: Training (a, b) and validation (c, d) accuracy and loss curves of the CNN models for classification of both genotypes.

ResNet-18: The ResNet-18 model required a lower number of epochs (about 33) to reach the desired performance than the ResNet-50. Figure 3.9b and Figure 3.9d show strong fluctuation in the training and validation losses between the 46th and 78th epochs, indicating overfitting. However, after the 78th epoch, there were no excessive fluctuations, which tells that overfitting was considerably decreased.

Inception-v3: As shown in Figure 3.9b and Figure 3.9d, both training and validation loss of Inception-v3 decreased with the number of epochs and tended to flatten. The accuracy of the training and validation sets of the model reached optimal accuracy at about 30 epochs (Figure 3.9a and Figure 3.9c).

DenseNet: For the DenseNet model, it took around 43 epochs for the network to converge appropriately. It is visible from the trends in Fig. 9a and c, initially, the training and validation accuracy values rose sharply. Later, the growth was gradual and reached a plateau. Training and validation losses dropped consistently and converged, indicating a well-fitting model (Figure 3.9b and Figure 3.9d).

AlexNet: The training set loss in AlexNet rapidly declined in the first ten epochs and then slightly fluctuated as the number of epochs increased (Figure 3.9b). From Figure 3.9d it follows that the validation loss became relatively stable after 42 epochs. After the 42nd epoch, the fluctuations in the accuracy of the validation set were negligible: around 97% with 3% tolerance (Figure 3.9c). No discrepancy between the training and validation accuracies implies that the proposed model had no evident overfitting.

ResNet-18 performed better than the other four models in our study, reaching an accuracy of 98.03% and loss of 0.0614, in contrast to ResNet-50, which had the lowest accuracy (94.83%) and the highest loss (0.1678). Figure 3.6a and Figure 3.6b illustrate the overall accuracy and loss of the proposed models for the classification of both genotypes based on their ripening stage. It follows that the classification capability of the proposed networks is generally maintained similarly, despite the combination of data sets and the increasing complexity of the problem. From Figure 3.7 it follows that the classification time of all models for the mixed batch increased compared to other scenarios. For example, DenseNet required the longest classification time (40.14 min), while AlexNet had the shortest classification time (2.1 min). This is logical since the models perform two simultaneous tasks of classifying the genotype and degree of ripeness. Although AlexNet had the shortest classification time, the time taken by ResNet-18 to complete the classification process was only 15 sec longer. Given these results, ResNet-18 was selected as the best model for the classification of genotype and ripeness of mulberries.

The superiority of the ResNet-18 over other models is related to its well-designed topology and structure. One of the common problems among CNN models is performance saturation specifically in Deep Neural Networks (Kwan et al., 2019). The ResNet-18 overcomes this issue by implementing an identity shortcut connection, which skips one or more layers and performs identity mapping of the layer than the original mapping (Zhang et al., 2019). Through the residual connections, all the inputs can forward propagate faster across the layers (He et al., 2016). On the other hand, ResNet-18 architecture has a shallow depth which reduces the overfitting problem, parameters and overhead of computing resources (Dutta et al., 2018). In general, ResNets are easy to optimize and can easily obtain accuracy gains from considerably increased depth.

3.5 Discussion

Manual classification of mulberry fruit ripening stages at different points of the production chain, i.e., farmers, manufacturers, distributors, and retailers, is a challenging task. The introduction and application of novel technologies may help in solving the problem of real-time detection and classification of fruits according to ripening level. Our research has been a first step into developing a new intelligent sorting system, based on computer vision and deep learning techniques, to enable recognizing and classifying white and black mulberries into four classes, i.e., unripe, semi-ripe, ripe, and overripe. The potential of the deep learning technique to solve this problem is substantiated by the excellent performance of the CNN models that were evaluated. From the literature, only one attempt is known to classify the mulberry fruit based on the ripening stage. In this study, the ripeness of red and white mulberry fruit was classified into three categories, i.e., unripe, ripe, and overripe. In addition to those three classes, the feasibility of classifying the semi-ripe class was also examined. For the classification task, the two traditional machine learning methods, i.e., ANN and SVM, were applied with handcrafted features. The results of this study are, unfortunately, difficult to compare with ours, because of differences in materials, techniques, data sets, as well as classification criteria.

Our approach is based on deep learning frameworks, i.e., AlexNet, DenseNet, ResNet-18, ResNet-50 and Inception-v3, which do not require manual definition of features. The performance of the trained CNN models in detecting the mulberry type and ripeness demonstrated high potential with accuracies between 95 and 98%. Moreover, the size of the sample set in our study increased the level of classification reliability. Considering the fact that ResNet-18 and AlexNet have fewer layers compared to the other architectures, the classification accuracy achieved with these shallow networks was surprisingly good in comparison with deeper networks such as DenseNet. Although deeper models can improve the classification accuracy (Sun and Ge, 2021), increasing the number of layers raises challenges of degradation, computational cost, internal covariate shifts, and vanishing gradients (Too et al., 2019). Low sample size in a given data set in deep networks can lead to overfitting. In this case, shallow networks can be a possible solution (Webb and Reis, 1994). The methodologies presented in this chapter result in a significantly better classification, adding to the knowledge of ripening detection. From an economic perspective, potentially it is an extremely cost-effective system because it is based on images from low-cost digital cameras without the need for sophisticated imaging equipment, such as hyperspectral imaging, laser backscattering imaging, multispectral imaging, fluorescence imaging, magnetic resonance imaging, etc.

3.6 Scale-Up and Integration of Technology

In Figure 3.10 a proposed design of an automatic mulberry sorting machine is depicted; it consists of a mulberry feeding unit, conveying units, imaging equipment, and a sorting mechanism. Mulberry fruits poured into the container are directed to

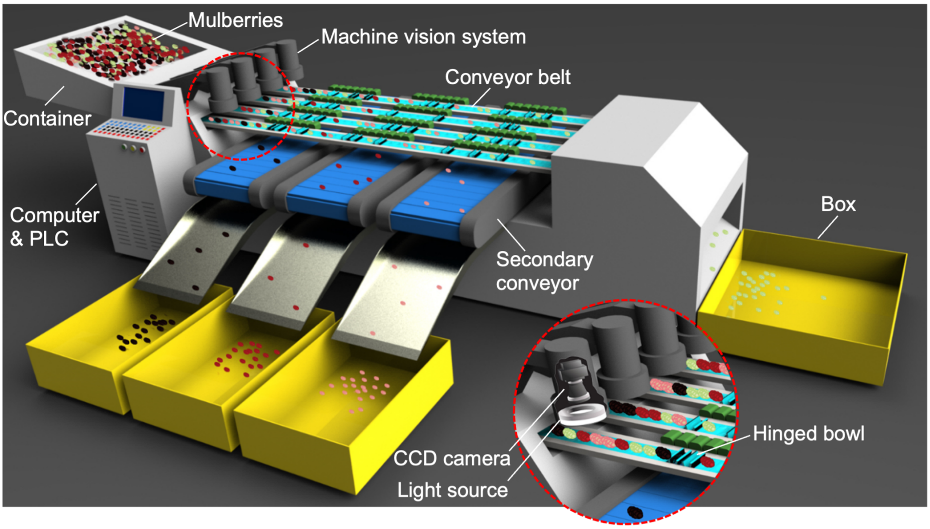

Figure 3.10: The conceptual design of an automatic mulberry sorting system.

the flat conveyors which are equipped with hinged small shallow bowls. The fruits on the conveyor belt continuously pass through the image-capturing area with a camera and illumination system mounted exactly perpendicular to the center of the conveyor. Data from the camera are transferred to a computer in real time to determine the fruit ripening class. The capacity of the system is increased due to multiple classification lines. When mulberry fruit reaches the exact location, determined by the speed of the conveyor, a signal is sent by the Programmable Logic Controller (PLC) (Hudedmani et al., 2017) to open the bowl, which will deliver fruit to the second conveyor to be discharged into a specified box.

To prevent bruising damage, mulberry fruits, are classified as overripe with the least mechanical strength falling into the box at the end of the first conveyor without dropping on the second conveyor. From the product quality perspective, the factors affecting the mechanical behavior of the fruit, including the trajectory of falling, the material of the conveyor, and static and dynamic forces applied to the fruit, need to be carefully evaluated before constructing such a system.

3.7 Conclusion

This study presents a comparison of different CNN-based models for the classification of two genotypes of mulberry, namely white and black, according to their ripening stages. The models that were evaluated include DenseNet, Inception-v3, ResNet-18, ResNet-50 and AlexNet. These models have been tested on a data set of 2000 fruit images (1000 per genotype), where the model was trained using 70% of the data set. The performance analysis has been done by comparing three performance metrics: accuracy, loss, and classification time. The major contributions of this study are as follows:

1. Although all CNN architectures achieved high accuracy, the AlexNet model outperformed the other models in the classification of white mulberries with an accuracy of 98.32%, loss of 0.0559, and classification time of about 1 min.

2. The experimental results demonstrated that the ResNet-18 model seems to be more reliable for the classification of black mulberry ripening with the best accuracy, minimal loss, and short classification time (98.65%, 0.0871, and 1.2 min, respectively).

3. The overall performance of the ResNet-18 was the best when the data sets of both genotypes were combined. The model was neither overfitted nor underfitted. The recognition of the fruit genotype and classification of the ripening stage of 600 testing samples showed that the overall accuracy, loss rate, and processing time were 98.03%, 0.0614, and 2.36 min, respectively.

The results of the current study could be extended to the classification of more than four ripeness stages. This study also provides the foundation for the design of an automated sorting machine for mulberry fruit. Moreover, the deep learning frameworks applied in this study can serve as a template for other types of horticultural commodities.