

Original Research Paper

Fine-Tuned MobileNet Classifier for Classification of Strawberry and Cherry Fruit Types

Venkatesh, Nagaraju Yallappa, Siddhanth Udayashankar Hegde and Sangeetha Raj Stalin

Department of Computer Science and Engineering,
University Visvesvaraya College of Engineering, Bangalore University, Bangalore 560 001, India

Article history

Received: 19-09-2020

Revised: 16-01-2021

Accepted: 19-01-2021

Corresponding Author:

Venkatesh

Department of Computer
Science and Engineering,
University Visvesvaraya
College of Engineering,
Bangalore University,
Bangalore 560 001, India
Email: venkateshm.uvce@bub.emet.in

Abstract: This paper proposed an accurate, fast and reliable strawberry, cherry fruit detection and classification system for the automated strawberry cherry yield estimation. State-of-the-art deep learning-based fine-tuned MobileNet Convolutional Neural Network is developed to detect and classify strawberry and cherry fruit types in the outdoor field. The proposed CNN model is trained on 4250 strawberry fruit images, 3878 Cherry fruit images and tested on 990 strawberry fruit's images, 1012 Cherry fruit images. To capture features and classify fruit type, a fine-tuned MobileNet Convolutional Neural Network model is presented in this study. The original MobileNet CNN model has 88 layers, which is computationally intensive and has more parameters. In the fine-tuned MobileNet CNN model, top layers are frozen and few layers are replaced with other layers such as a depthwise layer, pointwise layer, ReLu and Batch normalization layer, global average pooling layer. The fully connected layer is removed. The fine-tuned MobileNet CNN model performs quite well with higher accuracy of classification of fruit at less computation cost. The proposed CNN Model performs classification and labels them as Blueberry, Huckleberry, Mulberry, Raspberry, strawberry, strawberry wedge, Cherry Brown, Cherry Red, Cherry Rainier, Cherry wax Black, Cherry wax Red, Cherry wax Yellow. The proposed model's average validation accuracy is about 98.60% and the loss rate is about 0.38%. The fruit images are acquired from the cultivation field include fruits that are occluded by foliage, under the shadow and some degree of overlap of strawberry, cherry flowers.

Keywords: Strawberry, Cherry Fruit, Accuracy, MobileNet, Fine-Tuned MobileNet

Introduction

In recent years, farmers in India eventually lost yield in the field due to incessant rain or intermittent rain, or relentless plant disease. The plant disease is a parameter to measure disease type, level and predict yield. Fast, accurate analysis of disease and recommend treatment will assist the farmer to get maximum yield (Bock *et al.*, 2010). The plant disease is due to intensive farming for higher food production and excessive use of chemicals to control pests and weeds (Somasundaram *et al.*, 2020). To identify the plant disease, there various methods/techniques are adopted. Traditionally, physical observation by plant pathologists has been conducted to identify plant diseases. In recent years, various spectroscopic and imaging and image processing

methods have been used extensively to categorize fruits and plant epidemics. But, accurate instruments and unwieldy sensors are required; these instruments are high cost, which is a burden for farmers (Yuan *et al.*, 2014). With the accessibility of high definition smart modern cameras, accurate and automatic plant disease detection methods based on machine learning algorithms are proposed in the paper (Liakos *et al.*, 2018). Nevertheless, Machine learning methods are appropriate for the recognition of consistent, clear plant images; it requires complex image pre-processing and feature extraction methods. In the most recent year, Convolutional Neural Networks (CNN) and Deep learning algorithms have made a prodigious breakthrough in recognizing and classification patterns, object identification, visual localization and segmentation

(Singh *et al.*, 2016; Ren *et al.*, 2016; Ünal, 2020; Cao *et al.*, 2018). Because Deep Learning and CNN based model capable of extracting features, finding hidden structures within the unlabelled image and data that is unorganized. The deep learning-based model does not require human involvement to excerpt a set of features from the input images and does not perform complex pre-processing on images (Elavarasan and Vincent, 2020; Lin *et al.*, 2019 Zhang *et al.*, 2020). Inspired by the breakthroughs of CNNs and Deep learning algorithms in object detection, image classification, the research and applications of deep learning and CNN in plant disease detection currently a research hotspot.

Contributions

- We created our dataset consisting of 5240 strawberry fruit images, 4890 cherry fruit images. The strawberry, cherry fruit images are downloaded from the publicly available dataset. We annotated all images manually
- The proposed model is a modified version of MobileNet that freeze top layer and include layer such as depthwise layer, pointwise layer, ReLu and Batch normalization layer, global average pooling layer
- All the dataset images are pre-processed to eliminate background-noise, unevenly illuminated intensity and identified image edges through image-processing. The new set of images are generated by the image augmentation technique. The color and textual features are used in the proposed model
- The model can detect and classify different types of strawberry and cherry fruits with an accuracy of not less than 98.60%. The proposed approach will assist cultivators, fruit-trader to classify strawberry and cherry fruits

Organization of Paper

This paper contains different sections. Section 1 gives an introduction. The detailed literature survey is presented in section 2. Section 3 explains the problem statement. The discussion on the proposed solution is given in section 4. Section 5 elaborates on the experiment results and analyze the results. Section 6 presents conclusions.

Literature Survey

The principles of the artificial neural network are used in machine learning techniques. Machine-learning techniques have been applied to shallow architecture to solve well-structured, well-labeled problems, but machine learning techniques are incompetent to solve real-world unlabelled and unstructured problems such as image classification and identification, human speech recognition, natural voice detection and object detection.

With the rapid development of deep learning, it is feasible to solve real-world problems using deep learning techniques (Deng and Yu, 2014). The neural networks that use deep learning have a significant contribution in recognizing and classification of images with higher accuracy. The neural networks with deep architecture can recognize plant leaves disease and locating diseased leaves with higher accuracy. Plant disease detection and categorization methods are not the same for plants, fruits, crops. In literature, plant-related disease identification and classification algorithms are proposed. Lu *et al.* (2017) proposed a convolutional neural network with a deep learning framework to identify wheat disease in fields. The residual neural network model to recognize diseases on tomato fruit and identify tomato infected leaf locations (Zhang *et al.*, 2020; Fuentes *et al.*, 2017). Liu *et al.* (2020) designed a deep learning-based method to recognize and classify the most common leaf diseases on a grape leaf. Yu *et al.* (2020) designed a localization technique-based MobileNet model to identify and localize strawberry fruits in cultivation. The author also designed a method to detect picket points for harvesting robots to improve efficiency and reduce manual intervention. To identify and categorize diseases on paddy leaves, the author (Ramesh and Vydeki, 2020) has proposed a deep neural network-based algorithm. Ge *et al.* (2019) developed a localization algorithm to localize strawberries. The localization algorithm is based on coordinate transformation and clustering. The region segmentation technique is designed to segment strawberry fruits that are occluded by foliage, under the shadow and some degree of overlap. Identification of epidemic spores in microscopic images is critical for early disease, a deep learning-based model proposed in (Zhao *et al.*, 2019). Coulibaly *et al.* (2019) proposed a transfer learning technique to recognize the mildew epidemic of pearl millet. The deep convolutional neural network with GoogLeNet inception CNN model is developed to detect and categorize apple leaf disease (Jiang *et al.*, 2019). Deep Residual Neural Network-based algorithm is proposed for multiple plant disease recognition and classification in real-time acquisition conditions (Picon *et al.*, 2019). The performance of MobileNet is compared with VGGNet and GoogLeNet in recognition of different fruits; the experiment results illustrate that VGGNet performs better with higher accuracy and GoogLeNet takes less training time and memory compared to the MobileNet model (Sustika *et al.*, 2018).

The MobileNet model is an open-source, small, low-latency, parameterized model. MobileNet architecture is developed to function on mobile, handheld applications (Howard *et al.*, 2017). The MobileNet is built on depthwise disconnected convolutions; it splits combining and filtering functions into two separate to reduce the

cost of computation and number of parameters. MobileNet has difficulties interpreting are *clementine* and *kiwi* (Femling *et al.*, 2018).

Problem Statement

It is difficult for a cultivator or fruit-trader to manually classify and distinguish count strawberry and cherry fruits based on fruits' shape and color. The traditional methods deploy high definition cameras for the detection and classification of strawberry and cherry fruits and their types. But, the conventional approaches are labor-intensive, time-consuming and it may lead to misclassification or misjudgment while the classification of strawberry and cherry fruits. The color-based image-processing methods are not suitable for changing the environment and change in the color of fruit in subsequent weeks. Design a deep learning-based convolutional neural network model that classifies strawberry, cherry fruits accurately as possible and label them distinctly. The proposed convolutional neural networks should be computationally efficient and simple. The proposed CNN model should identify and categorize variants of strawberry, cherry fruits.

Proposed Solution

A. Materials and Methods

In this research work, we collected a set of images of strawberry and cherry fruits from a publicly available dataset. There are 5240 color images of strawberry fruits, 4890 color images of Cherry fruit in the dataset. We

labeled all images manually and images are in 256×256 pixels dimension.

B. Methodology

The proposed system is depicted in Fig. 1. The proposed system consists of different phases. It includes image acquisition and collection from a publicly available dataset, pre-processing of the image, image annotation, segmentation, feature extraction and classification data set into a training, validation, testing. The fine-tuned MobileNet model is trained with a trained dataset. The fine-tuned Mobilenet Model is validated using a testing dataset.

C. Features Extraction

Extraction of Color Feature

The texture and color features are extracted from captured images of Strawberry and Sweet-cherry. The RGB values are extracted from each image. The standard deviation, mean values are determined from RGB values. The alternative representation of RGB values such as *hue(H)*, *saturation(S)*, *brightness(B)* is extracted and the mean value is computed. Similarly, the mean value is computed from *a*, *b* and *Lightness* of the LAB model. P_{yx} represents pixels:

$$M_y = \frac{1}{n} \sum_x^n P_{yx} \tag{1}$$

$$S_y = \sqrt{\frac{1}{n} \sum_x^n (P_{yx} - M_y)^2} \tag{2}$$

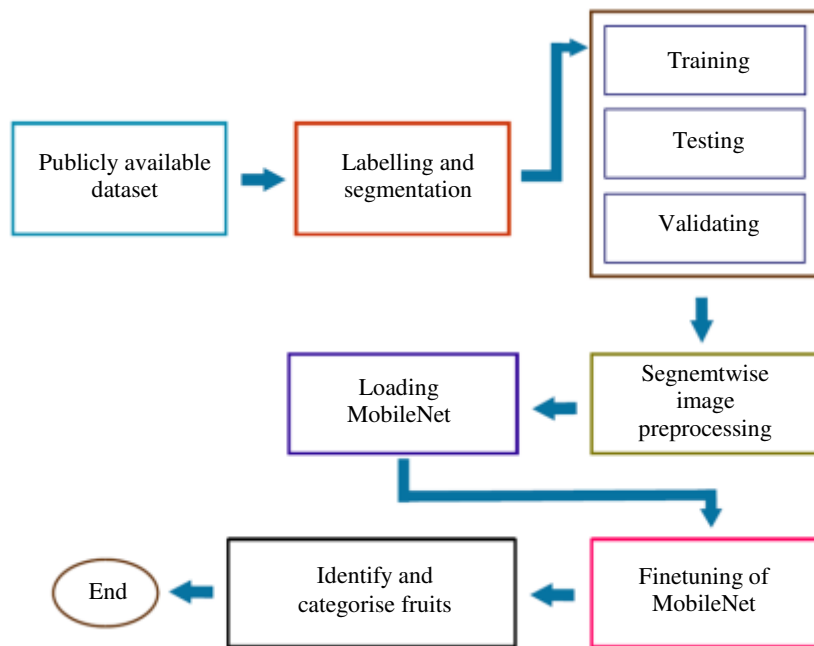


Fig. 1: System architecture

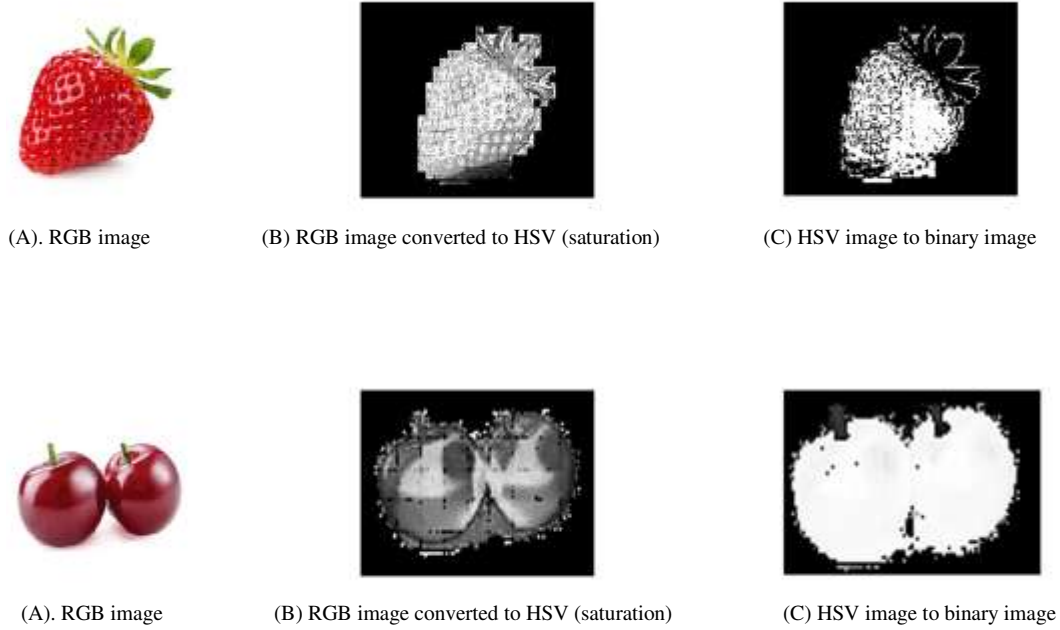


Fig. 2: Image pre-processing steps for input images

Extraction of Texture Feature

The texture feature of an image represents the spatial correlation between pairs of pixels and specific values in an image. The texture feature represents homogeneity(H_y), correlation(C_r), energy(E_n) and contrast(C_t) in an image and these values are computed using Eq. 3. Image pre-processing steps are shown in Fig. 2:

$$h = \sum_{x=0}^n \frac{P_{yx}}{1+(y-x)^2} \quad (3)$$

The correlation(C_r) is computed by Eq. 4:

$$C_r = \sum_{x=1}^n P_{yx} \frac{(y-M)(x-M)}{S_y} \quad (4)$$

The contrast(C_t) is computed by Eq. 5:

$$C_t = \sum_{x=0}^n P_{yx} (y-x)^2 \quad (5)$$

The energy(E_n) is computed by Eq. 6:

$$E_n = \sum_{x=0}^n (P_{yx})^2 \quad (6)$$

D. Bird-View of Proposed Fine-Tuned MobileNet Architecture

The MobileNet architecture model small in size, it has few parameters and has multiplication, addition

operation. Therefore, it is widely accepted and used in mobile-based computer vision applications. The MobileNet architecture can be fine-tuned with depthwise separable convolution, *Width Multiplier* α and *Resolution Multiplier* ρ . The original MobileNet architecture has 88 layers, it is shown in Fig. 3.

In this research paper, the fine-tuned MobileNet architecture is proposed. In fine-tune, MobileNet architecture, the representation of features at the higher-order layer is frozen shows that features are not updated at a given layer. The frozen feature can be used appropriately for processing a particular task. The basic idea freezing some layer of original model is to accustom features are more specific to the dataset that is considered. The proposed fine-tune MobileNet will not include the last one layer of the original MobileNet but it includes everything up to and including the last global average pooling layer. Through the experiment, we found that removing the last one layer from the original MobileNet architecture will improve classification accuracy. The proposed fine-tuned MobileNet architecture is shown in Fig. 4.

The fine-tuned MobileNet will be a smaller model size and satisfies the resource constraints of the application. The proposed fine-tuned MobileNet architecture has optimized latency and better accuracy in classifying fruits/objects. The proposed fine-tuned MobileNet architecture use factorization of convolutions. The convolution is factorized as a depthwise convolution, a 1×1 convolution. The 1×1 convolution is also referred as pointwise convolution. One filter for one input channel

concept is employed in a depthwise convolution. The 1×1 convolution is used to produce a blended output of the depthwise layer. Two-layer factorization is followed in depthwise separable convolution for filtering and grouping. The two-layer factorization will help to minimize computation time and design a model with a smaller size. The one filter for one input channel is called input depth and it is represented as in Eq. 7:

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i,l+j,m} \quad (7)$$

The variable \hat{K} indicates kernel of the size of depthwise convolutional $DK \times DK \times M$, a filter m_{th} in \hat{K} is used to channel the m_{th} in F to get a channel of the filtered output feature map \hat{G} . The cost of

computational for a depthwise convolution is given in Eq. 8:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F \quad (8)$$

Depthwise convolution is simple yet efficient because it filters out input channels and does not group them to produce new features. The new features are produced when the output of Depthwise convolution is grouped with 1×1 convolution and is referred to as depthwise separable convolution (Sifre and Mallat, 2014). The computation cost of depthwise separable convolution is given in Eq. 9:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \quad (9)$$

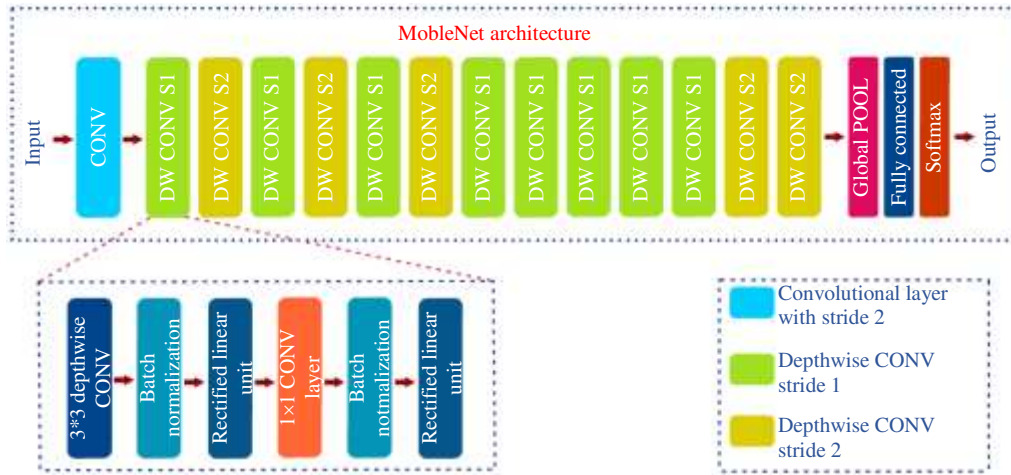


Fig. 3: MobileNet CNN model

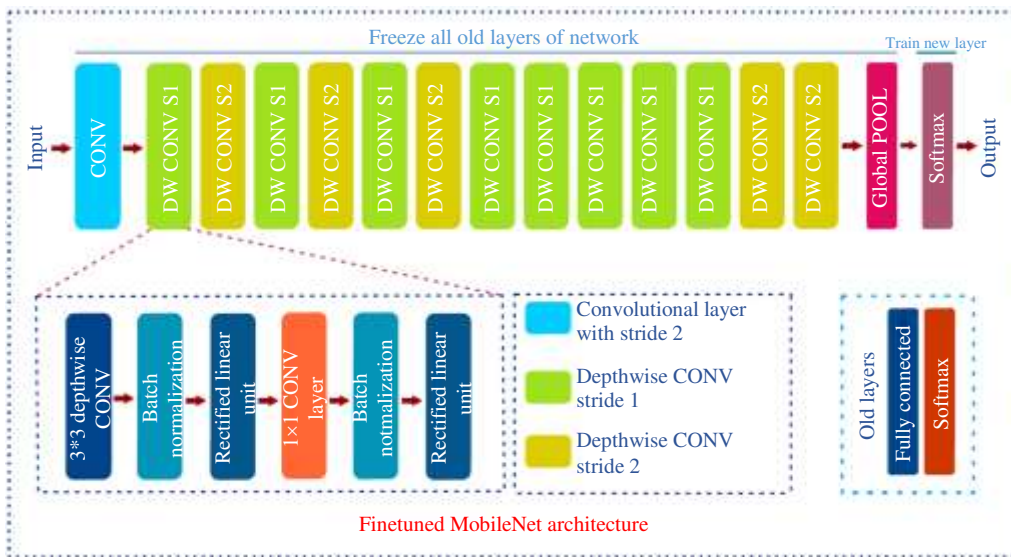


Fig. 4: Proposed MobileNet fine-tuned MobileNet

Table 1: Parameters in fine-tuned MobileNet CNN

Layer type	Stride	Kernel	Feature map
Conv.Layer	2	3×3	3×32
Depthwise Conv.	1	3×3	32 dw
Conv.Layer	1	1×1	32×64
Depthwise Conv.	2	3×3	64 dw
Conv.Layer	1	1×1	64×128
Depthwise Conv.	1	3×3	128 dw
Conv.Layer	1	1×1	128×128
Depthwise Conv.	2	3×3	128 dw
Conv.Layer	1	1×1	128×256
Depthwise Conv.	1	3×3	256 dw
Conv.Layer	1	1×1	256×256
Depthwise Conv.	2	3×3	256 dw
Conv.Layer	1	1×1	256×512
5× Depthwise	1	3×3	512 dw
5× Conv.Layer	1	1×1	512×512
Depthwise Conv.	2	3×3	512 dw
Conv.Layer	1	1×1	512×1024
Depthwise Conv.	2	3×3	1024 dw
Conv.Layer	1	1×1	1024×1024
Avg. Pool	1	7×7	7×7×1024
Softmax	1	Classifier	12 classes

Two-layer factorization is followed in depthwise separable convolution for filtering and grouping and the cost of computation is given by Eq. 10:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (10)$$

In fine-tuned MobileNet architecture, 3×3 depthwise separable convolutions are used. The fine-tuned MobileNet architecture has optimized computation cost. However, additional factorization does not optimize computation costs further. The fine-tuned MobileNet architecture is based on depthwise separable convolution and fine-tuning of some layers. The layers of fine-tuned MobileNet architecture is shown in Table 1. The first layer of fine-tuned MobileNet architecture is the full convolutional layer and every layer is followed by batch normalization and ReLU layer. The final layer is the softmax layer. The original MobileNet architecture is shown in Fig. 3. The fine-tuned MobileNet architecture is shown in Fig. 4. The fine-tuned MobileNet architecture has depthwise convolution, batch-normalization, ReLU and 1×1 pointwise convolution layer after each convolutional layer. The stride convolution manages the downsampling in depthwise convolution. The average pooling layer cut down spatial convolution to 1. All multiplication-addition computations are placed in dense 1×1 convolution to implement Multiplication-Addition efficiently. The dense 1×1 convolution has specific matrix multiply functions and reorders functions in memory. Most of the computation is performed in dense

1×1 convolution and it has most of the parameters. To make fine-tuned MobileNet architecture less computational, the parameter α called width multiplier is introduced. Where α can have value between 0 and 1. At each layer, the total input channels M is represented as αM and the total number of output channels N is represented as αN . To make architecture further less computation, another parameter called the resolution parameter is introduced. The resolution parameter is applied to images to reduce internal representation. The depthwise separable convolution with α is given by Eq. 11:

$$D_K \cdot D_K \cdot \alpha M \cdot D_F \cdot D_F + \alpha M \cdot \alpha N \cdot D_F \cdot D_F \quad (11)$$

With α the computational cost and number of parameters are minimized by α^2 . The parameters ρ is another parameter that has a direct influence on computation cost. The computation cost of depthwise separable convolutions layer with parameters α ρ is given by Eq. 12:

$$D_K \cdot D_K \cdot \alpha M \cdot \rho D_F \cdot \rho D_F + \alpha M \cdot \alpha N \cdot \rho D_F \cdot \rho D_F \quad (12)$$

Experiment Results and Discussion

All experiments are implemented using Python3.8(Anaconda3), OpenCV-Python, Keras library, etc. All experiments were conducted on the Intel i7 processor at 2.34GHz, 12GB RAM and NVIDIA graphic card.

Table 2 differentiates dissimilar CNN models concerning computation cost and accuracy rate. The fine-tuned MobileNet CNN Model is compared with VGGNet16 and GoogleNet.

The VGG-16 model is used for image classification. The model uses more filters to reduce the number of parameters, but the model has 138 Billionparameters. VGG-16 is a slower and bigger model and it uses the ReLU activation function (Simonyan and Zisserman, 2014). The GoogLeNet is a widely accepted CNN Model used for classification and detection. The GoogLeNet uses 11 convolutions to reduce the total number of parameters, but 11 convolution increases the depth of the model. It uses global average pooling to reduce trainable parameters. All convolution and max pooling operations are carried in a parallel way and output is stacked together to generated the final output (Szegedy *et al.*, 2015). The fine-tuned MobileNet model is less computation-intensive, more accurate, while GoogleNet and VGG16 are more compute-intensive and less accurate.

Table 3 shows the training accuracy, validation accuracy and loss rate of the proposed fine-tuned MobileNet architecture and other State-of-art CNN Models. It also gives other metrics used to evaluate the identification

and classification of various fruits. The results include sensitivity, precision, specificity and average precision. The higher precision value in Table 3 reflects the performance of the proposed fine-tune MobileNet model. The precision value is about 99.59%; therefore, our proposed model is better than other models. The performance of fine-tuned GoogLeNet (Steinbrener *et al.*, 2019) is compared with the proposed fine-tuned MobileNet for the same dataset that is considered in this study. The experimental result on this dataset shows that the proposed fine-tuned MobileNet achieves an accuracy of 98.60% and GoogLeNet achieves an average accuracy of 86.91%. The GoogLeNet is computationally intensive and requires more training time. The fine-tuned GoogLeNet achieves better accuracy if input images are taken from a hyperspectral camera and constructed 3D matrices for images.

The performance of AlexNet (Zhu *et al.*, 2018) is compared with the proposed fine-tuned MobileNet on the same dataset used in this study. The experiment results illustrate that AlexNet performs better than the proposed fine-tuned MobileNet if input images are 100×100 pixels, classification accuracy of AlexNet decreases as image size is increased to 256×256 pixels. The blended version of Inception-ResNet and Autoencoder models (Tran *et al.*, 2019) were compared with the proposed fine-tune MobileNet model. The experimental result shown that the blended version of Inception-ResNet, Autoencoder models has an accuracy of 89.28, 87.09% and the fine-tuned MobileNet model has an average accuracy of 98.60% for the dataset considered in this study. The proposed model reduces complexity by factorizing filtering and combining inputs into distinct layers. In the proposed fine-tune MobileNet, the features are more specific to the dataset on which the model was trained; therefore, more accuracy in the classification of fruit types. The performance of the proposed model is also compared with the modified version of AlexNet (Wu *et al.*, 2020) to detect and categorize the fruit types. The modified version AlexNet has an accuracy of 93.20% since it extracts a set of features with LBP and histogram gradients. The modified version of AlexNet does not adapt quickly to the specialized features of the new dataset.

Table 2: Comparison of different CNN models

Model name	Val. Accuracy (%)	Cost of computation (Minutes/Training)
Fine-Tuned MobileNet	97.60	36
GoogLeNet	86.91	223
VGGNet16	83.68	250

Table 3: Comparison of proposed fine-tuned MobileNet CNN with state-of-art CNN models

Model name	TAcc.	Sensi. @Speci.	Speci. @Sensi.	Prec.	VLoss	Vacc.
Proposed fine-tune MobileNet	0.9959	1.00	0.9995	0.9959	0.308	0.9860
Modified GoogLeNet (Steinbrener <i>et al.</i> , 2019)	0.8815	0.8593	0.9223	0.8816	0.670	0.8691
AlexNet (Zhu <i>et al.</i> , 2018)	0.9212	0.9023	0.9243	0.9244	0.587	0.9210
Incep + ResNet (Tran <i>et al.</i> , 2019)	0.8727	0.9023	0.8730	0.8976	0.604	0.8928
Modified AlexNet (Wu <i>et al.</i> , 2020)	0.9342	0.9023	0.9383	0.9314	0.483	0.9320

Figure 5 shows the confusion matrix of the classifier. The confusion matrix summarizes the performance of the proposed fine-tuned MobileNet classifier. The confusion matrix shows 6 classes of strawberry and 6 classes of cherry fruit. The average classification accuracy is about 98.60%, with an average F1-score of 96%. The proposed model learned features specific to our dataset and froze a few layers to force weights to be tuned to dataset-specific. The fully connected layer is omitted to prevent flatten and modifying outputs to six classes. Comparing the results of the proposed fine-tuned MobileNet with GoogLeNet, it is observed that as the depth of the model increases curve of over-fitting linearly increases.

The validation accuracy of the proposed classifier is shown in Fig. 6. In the proposed classifier, the base layers are frozen to retain the previous weights and few layers at the top are trained. The fine-tuned MobileNet architecture adds depthwise layer, pointwise layer, ReLU and Batch normalization layer, global average pooling layer to reduce computation cost and enhance accuracy. The validation accuracy is about 98.60%. The proposed classifier classifies strawberry and cherry fruit images into 12 different classes.

Validation loss and training loss are depicted in Fig. 7. The proposed classifier has lesser validation loss than training loss at epoch 1, 2. The difference between validation loss and training loss diminished as the number of epochs are increased. A good classifier always has a smaller validation loss and it represents the classification accuracy. At 15 epoch, the fine-tuned MobileNet classifier can classify all classes of strawberry and cherry fruits accurately. The proposed classifier and other state-art-CNN models are trained for 15 epochs. In the proposed fine-tuned MobileNet classifier, the loss values almost constant and near to 0 at 15th-epoch.

Figure 8 illustrates the result of the classification of strawberry and cherry fruit types. The proposed classifier accurately classify the images of strawberry and cherry fruits into Blueberry, Huckleberry, Mulberry, Raspberry, strawberry, strawberry wedge, Cherry Brown, Cherry Red, Cherry Rainier, Cherry wax Black, Cherry wax Red, Cherry wax Yellow.

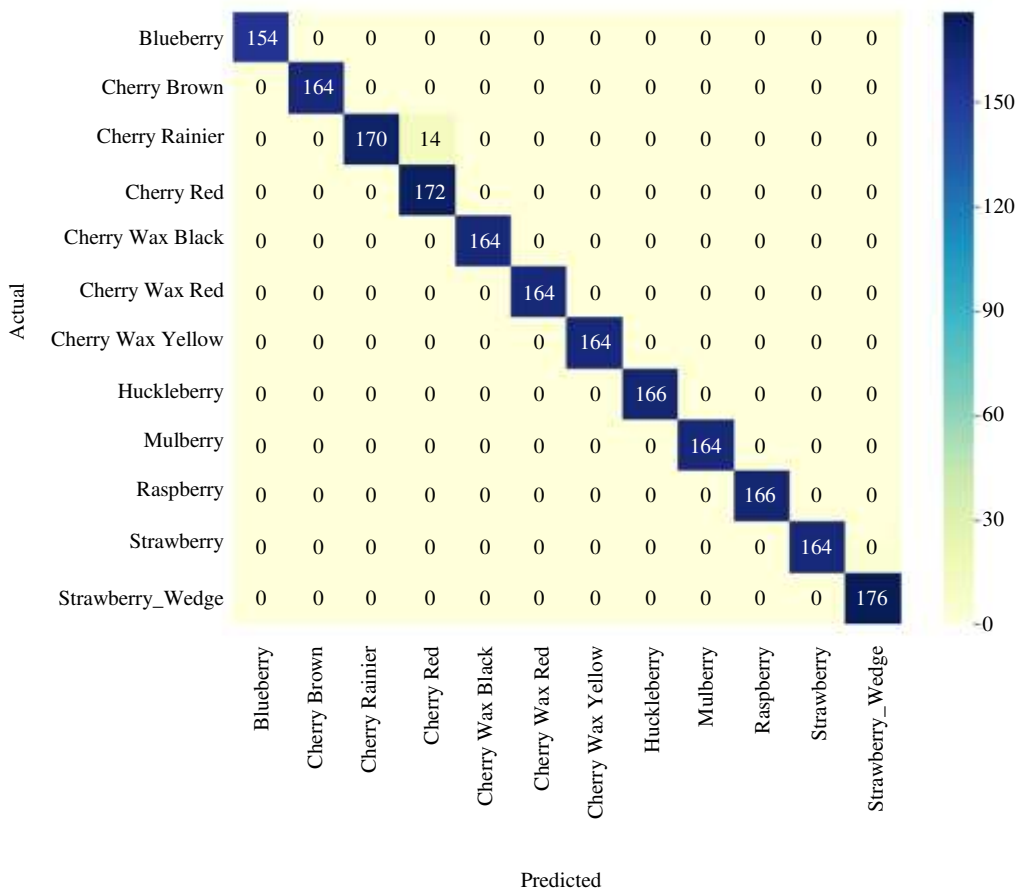


Fig. 5: Confusion Matrix of proposed fine-tune MobileNet model

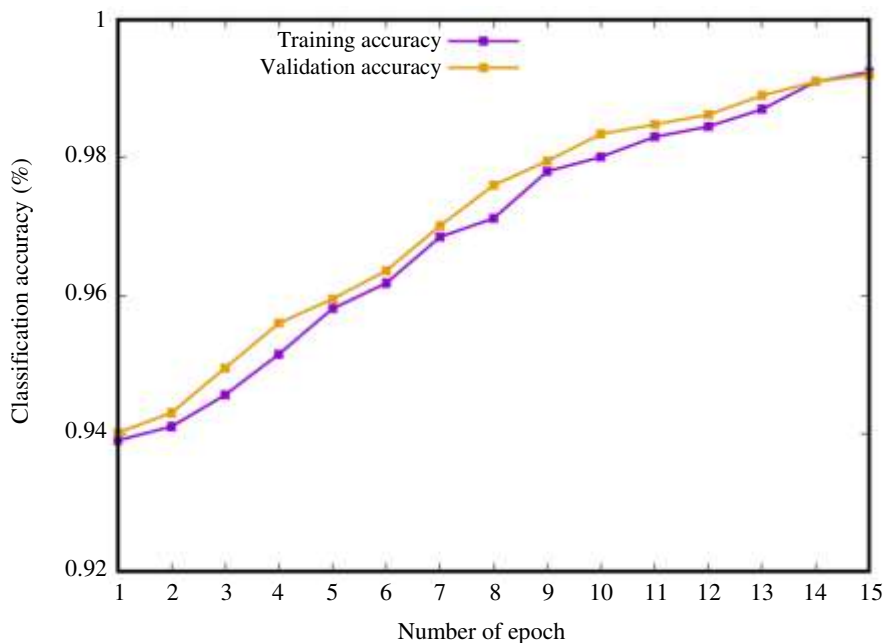


Fig. 6: Proposed fine-tune MobileNet model's validation and training accuracy

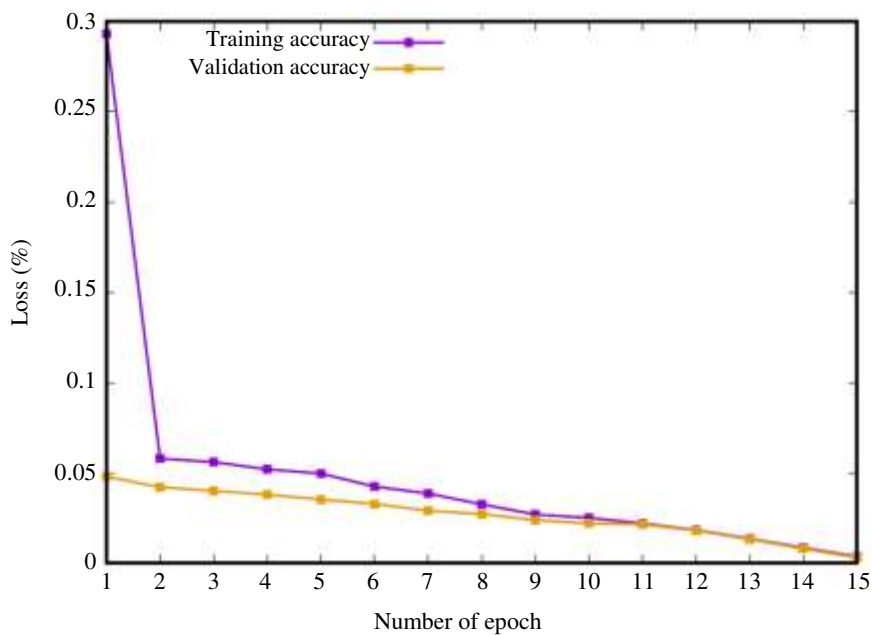


Fig. 7: Proposed fine-tune MobileNet model' validation and training loss



Fig. 8: Input Image of strawberry and cherry fruit and classification of strawberry and cherry fruit types

Conclusion and Future Work

The cultivator, fruit-trader and the consumer can distinguish strawberry and cherry fruits according to its physical appearance if the number of strawberry and cherry fruit types are one or two. But, when there are more than one strawberry, cherry fruit types and their physical appearances almost the same with change in fruit's color. The goal of this paper is to develop a classifier that distinctly detects and classifies strawberry, cherry fruit types. A Fine-tuned MobileNet classifier is designed and used to classify fruits in this study. Top layers of the original MobileNet are frozen and few layers are replaced with depthwise layer, pointwise layer, ReLU and Batch normalization layer, global average pooling layer in the proposed fine-tuned MobileNet model. Few layers of original MobileNet are replaced to train the model with features that are more specific to the dataset on which the model was trained, therefore, more accuracy in the classification of fruit types. The ReLU and dropout are employed in a training model to avert over-fitting and enhance convergence computation. The publicly available dataset is used in this study. The proposed classifier distinctly classify the strawberry, cherry fruits type as Blueberry, Huckleberry, Mulberry, Raspberry, strawberry, strawberry wedge, Cherry Brown, Cherry Red, Cherry Rainier, Cherry wax Black, Cherry wax Red, Cherry wax Yellow. The proposed classifier categorizes strawberry and cherry fruit types 98.60% accurately in the uneven background, complex cultivation field.

In future work, we design a CNN-based model and deploy in handled devices such as mobile for classification of all fruit types at super bazaar market, wholesalers and retailers. In addition to this, a CNN-based model shall be used to identify damages on fruits and within fruits, softness, maturity and nutrients of the fruit. In the future, develop a CNN-based model that can be used in orchards for harvesting fruits and control diseases on fruits.

Acknowledgment

The authors would like to thank all my colleagues and research scholars of the department and reviewers who have take their valuable time for giving useful comments.

Author's Contributions

Venkatesh: Conceptual design, model design, acquisition of data, analysis of the experiment results and writing a manuscript.

Nagaraju Yallappa: Has participated in all experiments, coordinated data analysis, dataset collection and coordinated image pre-processing.

Siddhanth Udayashankar Hegde: Design and execution of the model.

Stalin Sangeetha Raj: Data acquisition, analysis and interpretation of data and proofreading.

The authors have followed all research ethics.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Bock, C. H., Poole, G. H., Parker, P. E., & Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis and by hyperspectral imaging. *Critical Reviews in Plant Sciences*, 29(2), 59-107.
- Cao, C., Huang, Y., Yang, Y., Wang, L., Wang, Z., & Tan, T. (2018). Feedback convolutional neural network for visual localization and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7), 1627-1640.
- Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., & Traore, D. (2019). Deep neural networks with transfer learning in millet crop images. *Computers in Industry*, 108, 115-120.
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3-4), 197-387.
- Elavarasan, D., & Vincent, P. D. (2020). Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. *IEEE Access*, 8, 86886-86901.
- Femling, F., Olsson, A., & Alonso-Fernandez, F. (2018, November). Fruit and vegetable identification using machine learning for retail applications. In 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) (pp. 9-15). IEEE.
- Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, 17(9), 2022.
- Ge, Y., Xiong, Y., Tenorio, G. L., & From, P. J. (2019). Fruit localization and environment perception for strawberry harvesting robots. *IEEE Access*, 7, 147642-147652.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

- Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access*, 7, 59069-59080.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Lin, Z., Mu, S., Huang, F., Mateen, K. A., Wang, M., Gao, W., & Jia, J. (2019). A unified matrix-based convolutional neural network for fine-grained image classification of wheat leaf diseases. *IEEE Access*, 7, 11570-11590.
- Liu, B., Ding, Z., Tian, L., He, D., Li, S., & Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in Plant Science*, 11, 1082.
- Lu, J., Hu, J., Zhao, G., Mei, F., & Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. *Computers and electronics in agriculture*, 142, 369-379.
- Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., & Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture*, 161, 280-290.
- Ramesh, S., & Vydeki, D. (2020). Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Information processing in agriculture*, 7(2), 249-260.
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- Sifre, L., & Mallat, S. (2014). Rigid-motion scattering for image classification. Ph. D. thesis.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, 21(2), 110-124.
- Somasundaram, J., Sinha, N. K., Dalal, R. C., Lal, R., Mohanty, M., Naorem, A. K., ... & Chaudhari, S. K. (2020). No-till farming and conservation agriculture in South Asia—issues, challenges, prospects and benefits. *Critical Reviews in Plant Sciences*, 39(3), 236-279.
- Steinbrener, J., Posch, K., & Leitner, R. (2019). Hyperspectral fruit and vegetable classification using convolutional neural networks. *Computers and Electronics in Agriculture*, 162, 364-372.
- Sustika, R., Subekti, A., Pardede, H. F., Suryawati, E., Mahendra, O., & Yuwana, S. (2018). Evaluation of deep convolutional neural network architectures for strawberry quality inspection. *Int. J. Eng. Technol*, 7(4), 75-80.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- Tran, T. T., Choi, J. W., Le, T. T. H., & Kim, J. W. (2019). A comparative study of deep CNN in forecasting and classifying the macronutrient deficiencies on development of tomato plant. *Applied Sciences*, 9(8), 1601.
- Ünal, Z. (2020). Smart Farming Becomes Even Smarter With Deep Learning-A Bibliographical Analysis. *IEEE Access*, 8, 105587-105609.
- Wu, A., Zhu, J., & Ren, T. (2020). Detection of apple defect using laser-induced light backscattering imaging and convolutional neural network. *Computers & Electrical Engineering*, 81, 106454.
- Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. *IEEE Access*, 8, 116556-116568.
- Yuan, L., Huang, Y., Loraamm, R. W., Nie, C., Wang, J., & Zhang, J. (2014). Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Research*, 156, 199-207.
- Zhang, Y., Song, C., & Zhang, D. (2020). Deep learning-based object detection improvement for tomato disease. *IEEE Access*, 8, 56607-56614.
- Zhao, Y., Lin, F., Liu, S., Hu, Z., Li, H., & Bai, Y. (2019). Constrained-Focal-Loss Based Deep Learning for Segmentation of Spores. *IEEE Access*, 7, 165029-165038.
- Zhu, L., Li, Z., Li, C., Wu, J., & Yue, J. (2018). High performance vegetable classification from images based on alexnet deep learning model. *International Journal of Agricultural and Biological Engineering*, 11(4), 217-223.