DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative Study

Conference Paper · May 2021

DOI: 10.1109/ICCCSP52374.2021.9465499

3 authors:

M. Israk Ahmed
Memorial University of Newfoundland

Shahriyar Mahmud Mamun
Daffodil International University

Asif Uz Zaman Asif
Daffodil International University

DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative Study

M. Israk Ahmed
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
israk15-8919@diu.edu.bd

Shahriyar Mahmud Mamun
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
shahriyar15-8600@diu.edu.bd

Asif Uz Zaman Asif
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
asifuzzaman.cse@diu.edu.bd

***Abstract*—In this paper, an attempt is addressed towards accurate vegetable image classification. A dataset consisting of 21,000 images of 15 classes is used for this classification. Convolutional neural network, a deep learning algorithm is the most efficient tool in the machine learning field for classification problems. But CNN requires large datasets so that it performs well in natural image classification problems. Here, we conduct an experiment on the performance of CNN for vegetable image classification by developing a CNN model from the ground. Additionally, several pre-trained CNN architectures using transfer learning are employed to compare the accuracy with the typical CNN. This work proposes the study between such typical CNN and its architectures(VGG16, MobileNet, InceptionV3, ResNet etc.) to build up which technique would work best regarding accuracy and effectiveness with new image datasets. Experimental results are presented for all the proposed architectures of CNN. Besides, a comparative study is done between developed CNN models and pre-trained CNN architectures. And the study shows that by utilizing previous information gained from related large-scale work, the transfer learning technique can achieve better classification results over traditional CNN with a small dataset. And one more enrichment in this paper is that we build up a vegetable images dataset of 15 categories consisting of a total of 21,000 images.**

***Index Terms*—Vegetable image classification, deep learning, CNN, VGG16, MobileNet, Inception-V3, ResNet.**

I. Introduction

Vegetables are one of the most common food items in everyday meals worldwide. People around the world produce many kinds of vegetables. On our planet there exists almost hundreds of thousands of species of vegetables according to some survey [1]. And vegetables are important for human beings as a result of their nutrients, minerals, phytochemical mixtures, and dietary fiber content. There are similarities in many vegetable types in terms of color, texture, and shape. Even country-wise same vegetable has a different name. From vegetable production to delivery, several common steps are operated manually. Like picking, and sorting vegetables. And recognizing a vegetable is a hard task for the customer in the market, as there are similarities between different vegetables. As several steps of vegetable production to consumption is still depended on manual operation with a huge sum of labor constrain, it truly influences the advancement of commercialization of vegetable items. To solve this issue, automation in vegetable picking, sorting, labeling is required by introducing a vegetable image classifier so that time and money can be saved. In modern days, in the field of agriculture, fundamental research work are classification and detection. As there are various sorts of vegetables and numerous individuals don't have any idea about them. So, the plan of a vegetable classifier will likewise carry straightforwardness to individuals lives. Also, the sorting of vegetables is manually done in super shops and distribution centers. Therefore, to solve these problems this research is conducted. The purpose of this paper is to classify vegetable images with higher accuracy, with the help of CNN and pre-trained DCNN with transfer learning. Nowadays, convolutional neural network is often used for classification, segmentation, image recognition etc. Deep network architecture is the utmost power of CNN which enables CNN to automatically learn mid to high-level considerations from new data [2], [3]. In this research, we proposed a CNN model developed from scratch as well as four fine-tuned state-of-the-art CNN architectures (InceptionV3, VGG16, Resnet, MobileNet) for vegetable image classification. Also a comparative study is done based on the performance between CNN and its architectures.

II. Related Work

This research is about classifying vegetable images with higher accuracy and efficiency. And also finding out the efficient model along with a technique that performs well in terms of accuracy, time, and cost. Almost 3 years ago, Om Patil et al. [4] used InceptionV3(known as GoogLeNet) for vegetable classification tasks. By fine-tuning the inception network and applying the transfer learning technique, the proposed model can classify 4 types of vegetables- carrots, onions, cucumbers, and tomatoes. The accuracy of their fine-tuned inception-V3 is

978-1-6654-3277-1/21/$31.00 ©2021 IEEE

99% for a comparatively smaller dataset that contains around 1200 images. Yuki Sakai et al. [5], proposed deep neural network for classification of vegetable by extracting features and learning the object. Recognition rate of their DNN model was 97.58%. They worked with a very small dataset, having 8 types of vegetables and the dataset contains only 200 total images. For the learning process of vegetable recognition, they used 3 million iterations to get that accuracy, which is lengthy and expensive as well. Frida Femling et al. [6], deep convolutional neural network-based transfer learning model has been used. For image collection, they used Raspberry Pi. They worked with only inception-V3 and MobileNet and got 96% accuracy with inception-V3 and 97% accuracy with MobileNet respectively. By two properties their model is appreciated. One is propagation time, another is how much time it takes for classifying a fruit or vegetable image. But the dataset size is small as it contains 4,000 images from ImageNet and a total of 4,300 images of 10 classes. Zhu L et al. [7], they proposed AlexNet network for vegetable image classification. And also a comparative study is done by addressing the Support Vector Machine classifier and the traditional back propagation neural network. They worked with 5 types of vegetables- pumpkin, mushrooms, broccoli, cauliflowers, and cucumber. And images were obtained from the ImageNet dataset and were expanded by adopting the data expansion method to a total of 24,000 images so that overfitting is reduced. The highest accuracy they gained from their experiment is 92.1% with the AlexNet network. Guoxiang Zeng [8], proposed image saliency technique and VGG architecture for the classification task of fruit and vegetables. To reduce the unnecessary noise from the image, they extracted dense features from each image and filtered the complicated backgrounds of it. And for determining the significant area in an input image, they choose a bottom-up graph-based visual saliency (GBVS) model [9]. They used a total of 12,173 images spanning 26 categories, among them 13 categories were vegetables (3678 images)- broccoli, celery, cowpea, green onion, garlic, cucumber, mushroom, carrot, onion, pumpkin, chinese cabbage, tomato, and pepper. And the classification accuracy of their model is 95.6%. Li et al. [10] proposed an improved VGG model (VGG-M-BN) and got 96.5% accuracy. They worked with 10 categories of vegetables. Images were mostly collected from the ImageNet dataset and were expanded by adopting the data expansion method.

Though all these works are good, all of them have limitations in terms of accuracy, efficiency, and effectiveness. The common problem we found in those works is the size of the dataset, and dataset source. Secondly, the training time of most of those work is so long. And in terms of cost and time, long training time is inefficient. And very few researchers work with only vegetable classification [4], [5], [7], [10] but all of that work have limitations. The minimum type of vegetable they worked with was four [4], and the maximum was ten [10].

III. Proposed Methodology

This study of vegetable image classification proposed a developed CNN model and state-of-the-art CNN model with transfer learning technique. In 2012, a breakthrough occurred in the image recognition area through the ILSVRC(ImageNet Large-Scale Visual Recognition Challenge) competition on the ImageNet dataset [11]. As of now different types of deep convolutional neural network architectures are introduced that varies in numbers of layers as well as complexity and available for use by anyone. In the past few years, using transfer learning concepts like fine-tuning and layer freezing in CNN architectures beat the traditional machine learning models in terms of performance and efficiency for image classification problems. Here, we used a CNN model and four state-of-the-art CNN architectures - VGG16, InceptionV3, ResNet, and MobileNet. These four models are pre-trained on the ImageNet dataset, a large-scale dataset that contains 1.2 million training data mostly animals and daily objects. By applying transfer learning techniques, learned features of these DCNN models may help to make very deep network architecture effective for our dataset.

A. CNN

In the field of computer vision, convolutional neural network basically a normalized multilayer perceptron (MLP) has been the most influential innovation. In recent years, CNN has dominated the computer vision and image processing field for large-scale image recognition, classification, and segmentation task. A typical CNN starts with an input layer, ends with an output layer, where in between them there exist multiple hidden layers. Convolutional, pooling, normalization (ReLU), fully-connected layers are part of the hidden layer [12]. Fig. 1 shows typical CNN layers. Input layer takes the target image

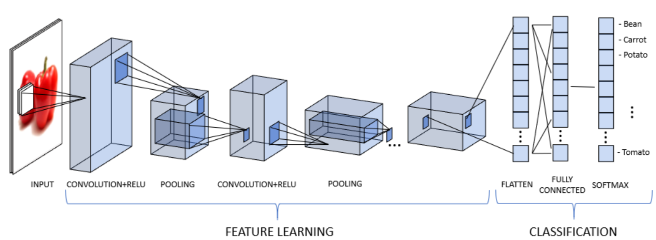

Fig. 1. CNN Layers [13].

data as input. The image is then reshaped to an optimal size and forwarded to the next layer which is a convolutional layer. There exists a number of kernel or filter that actually slides over the input and performs element-wise multiplications in order to extract features. Here, through an activation function, the negative weighted input will be replaced with zero otherwise it will go to output directly. The most widely used activation function is ReLU (Rectified-Linear Unit) and for many types of neural networks, it is the default activation function. It's a non-linear function and faster than the other activation function like- Sigmoid, Scaled Exponential Linear

Unit (SELU), Exponential Linear Unit (ELU), Gaussian Error Linear Unit (GELU) etc. Features that are extracted from the convolutional layer then sent to the pooling layer. This layer preserves only important features from a large image by reducing parameters. Then the fully connected layer translates these highly filtered images into categories. And another non-linear function named softmax finally gives the decimal probabilities ranged from 0 to 1 to each class. In this research, a 6-layer convolutional neural network is proposed that is completely built from scratch. The input image size is selected to 32×32 for reducing overall computational time so that a good model can be created in terms of efficiency. Data augmentation techniques like rotation, rescale, shear, zoom, and horizontal flip are also applied to the 32×32 size 3-channel training image data. ReLU was used as the activation function with each convolutional layer. And for improving generalization error a dropout rate of 0.25 is used so that it can overcome overfitting issues. Finally, softmax is used in order to find the probabilities of each class in decimal numbers. Fig. 2 shows the developed CNN model architecture.

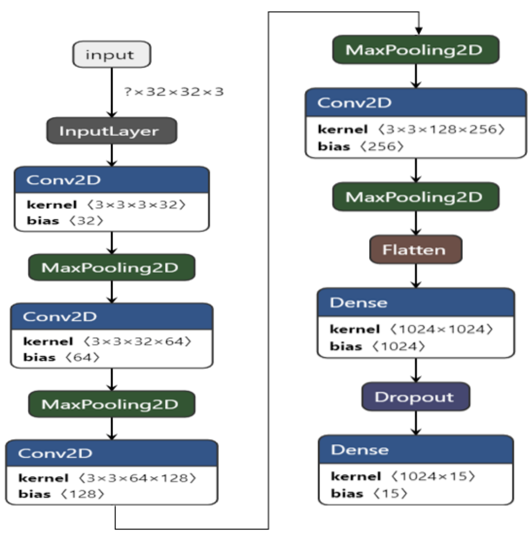

Fig. 2. Developed CNN Model.

B. Transfer learning

Building and training a deep convolutional neural network from scratch is time-consuming, costly, and hard. A deep network means it contains multiple layers, where it can also be multiple convolutional layers in exact order or sequence in order to classify the exact image. To learn feature mapping, when this kind of large and deep architecture is being trained from the ground, it needs a large-scale dataset. Transfer learning is a technique of re-using a previously developed model on a second related task. The utilization of knowledge that is learned from a previous domain for the improvement and optimization of a new domain is the core idea of it [12]. The concept of transfer learning in the field of ML was first presented in a NIPS-95 workshop named "Learning to Learn", agenda of which was a lifelong ML technique that is able to hold and reuse already learned information [14]. It can quickly transfer learned features from one(source) domain to the another(target) domain using a smaller dataset in the fastest manner using the easiest way [15]. Hence, the concept of transfer learning is adopted in this research and in this case target domain is vegetable image classification. Two approaches are available for implementing transfer learning- one is "Pre-trained Model" approach and another is "Develop Model Approach". In deep learning, the most commonly used approach is the "Pre-trained Model Approach" and it was selected for this research. In this approach, a pre-trained source model that is trained on large-scale data is selected, and then the whole model or parts of the model are used as the starting point for a model of another task. Where fine-tuning of the model may be required on the input-output pair of the target domain. Fine-tuning refers that, keeping the weights and biases of some layers unfrozen and using them for training so that the pre-trained model can perform well on the training data. This paper proposes transfer learning techniques on four pre-trained state-of-the-art CNN architectures - VGG16, InceptionV3, ResNet50, and MobileNet with fine-tuning.

C. Fine-tuning CNN Architectures for Transfer Learning

Fine-tuning refers to the technique of using learned features or weights and biases from a pre-trained deep CNN as the initialization of a target CNN model so that the target CNN can be trained on target data in a supervised manner [2]. As the interrelation between our target dataset and the source domain dataset is notable, for each architecture we used layer-wise fine-tuning. We fined-tuned our four state-of-the-art CNN architectures by freezing the convolutional base so that previously learned weights and biases can be repurposed in our task. We used the full architecture as it is, except the output layer which is basically the last fully connected layer. Here, convolutional base is the fixed feature extractor, and the extracted feature will be used for classifying the input image. For retraining these transferred networks we set the number of classes in the output layer to 15 referring to our multi-class classification task. Finally, the final layer was retrained.

1) VGG16: Visual Geometry Group(VGG) network contains VGG16 and VGG19. VGG16 was the first runner-up in ILSVRC and consists of 16 convolutional layers, three fully-connected, and five max-pooling layers [3]. It has over 138million parameters. It uses the ReLU activation function and dropout for improving generalization error with all fully connected layers, and also uses the softmax function in the output. In this model several 3*3 filters are used in order to replace the large-size filters or kernels. And these small-size kernels give the opportunity of complex feature extraction at a low cost.

2) InceptionV3: InceptionV3 is a state-of-the-art CNN architecture developed by google also known as GoogleNet. It has 48 layers and it replaced the last fully connected layer with average pooling right after the last convolutional layer. As a result total number of parameters(24 million) is reduced and makes this model more computationally efficient. In InceptionV3 it has eleven inception module, each module contains convolutional layers with ReLU activation function, convolutional filter for dimension reduction, max-pooling layer, fully-connected layer, and an output layer along with a softmax activation function [11]. It uses the ReLU activation function and dropout for improving generalization error with the fully connected layer.

3) MobileNet: MobileNet is a DCNN architecture that uses depthwise separated convolutions to construct lightweight deep convolutional neural networks which are suitable for mobile and embedded vision applications [16]. Here, depthwise separable convolutions mean, it does the typical combining and filtering tasks in different separate layers. Except for the first layer, it uses depthwise separable convolutions in order to reduce the model size rather than typical convolutions so that computational efficiency gets increased. It has total of 28 layers where all are followed by batch normalization. It uses ReLU as an activation function and uses the softmax function in the output.

4) ResNet50: ResNet elaboration is Residual Network, which is a particular sort of neural network that was presented in 2015 by Microsoft [17]. ResNet50 is an example network from ResNet network architecture. ResNet introduces a revolutionary technique to overcome the degradation problem named "residual mapping". It has 50 layers and over 23 millions of trainable parameters. It uses global average pooling instead of fully connected layers than other standard DCNN architecture. Though it's much deeper than other used architectures in this research, it's considerably light-weight.

IV. Dataset

This research is conducted to find out the highest accuracy for vegetable image classification. The initial experiment is done with 15 types of common vegetables that are found throughout the world. The vegetables that are chosen for the experimentation are- bean, bitter gourd, bottle gourd, brinjal, broccoli, cabbage, capsicum, carrot, cauliflower, cucumber, papaya, potato, pumpkin, radish and tomato. Fig. 3 shows a random example from each class. A total of 21000 images from 15 classes are used where each class contains 1400 images of size 224×224 and in *.jpg format. All of the experiment in this paper is done by using own dataset, and the dataset split 70% for training, 15% for validation, and 15% for testing purpose.

V. Results and Discussion

A. Experimental Results

In the developed 6-layers CNN model, the 3-channel input image size is 32×32 and the total number of parameters is over 1.4 million. In the 6-layer CNN model, there are four conv2D layers and 2 fully connected layers. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 50 epochs. With this combination of training, we have got the best accuracy from the developed CNN model. Fig. 4 shows the graph of training-validation loss and accuracy. The validation accuracy for our developed CNN model is 97.6% and testing accuracy is 97.5%. Also Fig. 5 shows the confusion matrix of the 6-Layer CNN model.

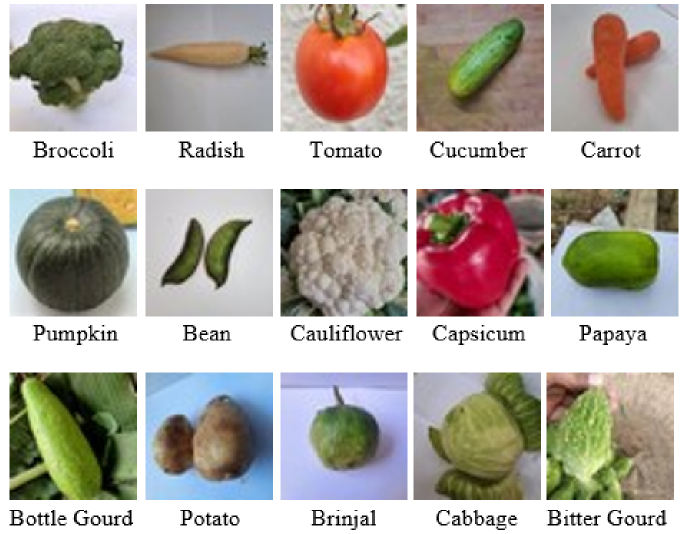

Fig. 3. Example From Each Class.

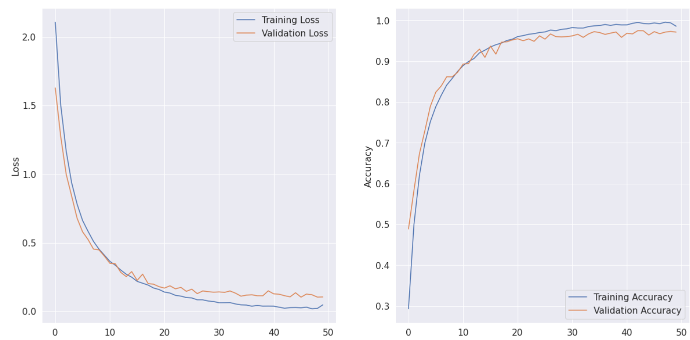

Fig. 4. 6-Layer CNN Training-Validation Loss and Accuracy.

VGG16 is a 16-layer CNN architecture that has over 138 million parameters. Input image size is 224×224 and in-built image pre-processing of VGG architecture is used before they get passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 6 shows the training-validation loss and accuracy graph for VGG16. With fine-tuning the architecture, we have got 99.8% validation accuracy and 99.7% testing accuracy from the VGG16 model. And Fig. 7 shows the confusion matrix of the model.

InceptionV3 has 48 layers and 24 million parameters. The input image size is 299×299 and in-built image pre-processing of Inception architecture is used before they get passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 8 shows training vs validation loss and accuracy graph for InceptionV3. With fine-tuning the architecture, we have got 99.6% validation accuracy and 99.7% testing accu-

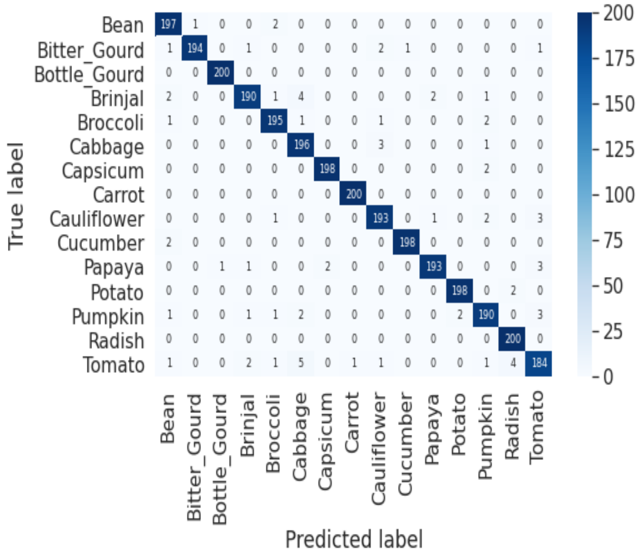

Fig. 5. Confusion Matrix of Proposed 6-Layer CNN.

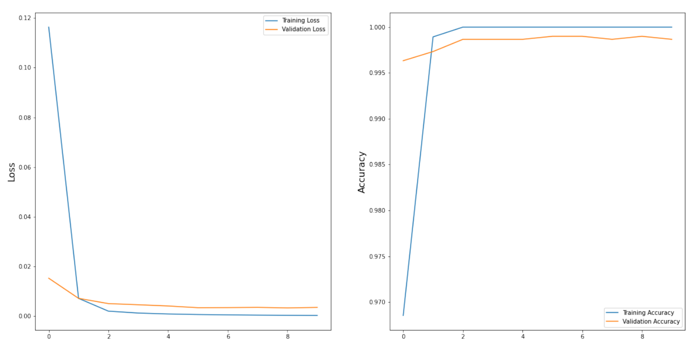

Fig. 6. VGG16 Training-Validation Loss and Accuracy.

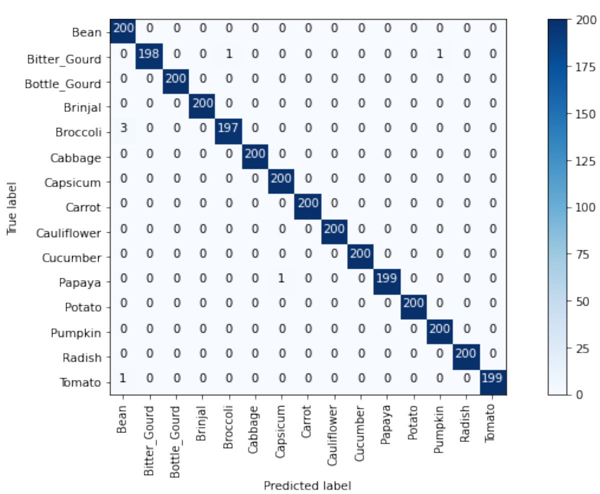

Fig. 7. Confusion Matrix of Fine-Tuned VGG16.

racy. Fig. 9 shows the confusion matrix for the fine-tuned InceptionV3.

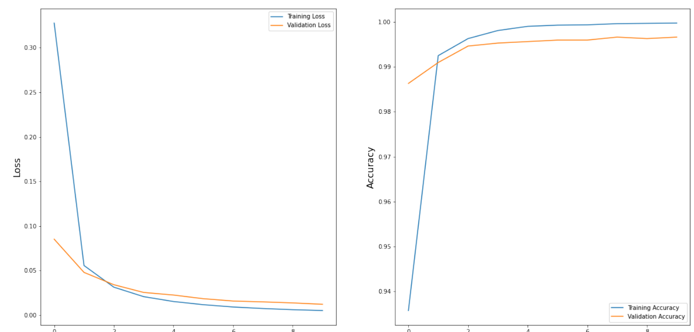

Fig. 8. InceptionV3 Training-Validation Loss and Accuracy.

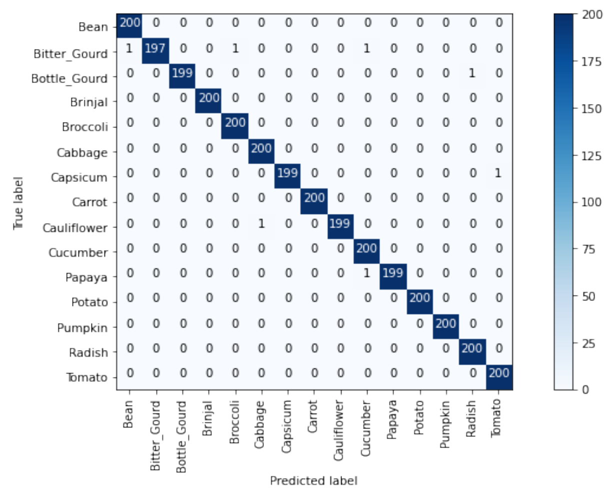

Fig. 9. Confusion Matrix of Fine-Tuned InceptionV3.

MobileNet is a very light-weight CNN architecture comparing to other architectures used in this research has only 28 layers and the input image size is 224×224. In-built image pre-processing of MobileNet architecture is used before each image gets passed to the network. MobileNet V1 is used for the experiment. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. With fine-tuning the architecture, we have got 99.8% validation accuracy and 99.9% testing accuracy. Fig. 10 shows training vs validation loss and accuracy graph for MobileNet. And Fig. 11 shows the confusion matrix for the model.

ResNet is also used in this experiment specifically ResNet50 that has only 28 layers and the input image size is 224×224. In-built image pre-processing of ResNet architecture is used before each image gets passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 12 shows the training-validation loss and accuracy graph for ResNet50. With fine-tuning the architecture, we have got 99.9% validation accuracy and 99.9% testing accuracy. Fig. 13 shows the confusion matrix for the fine-tuned ResNet50.

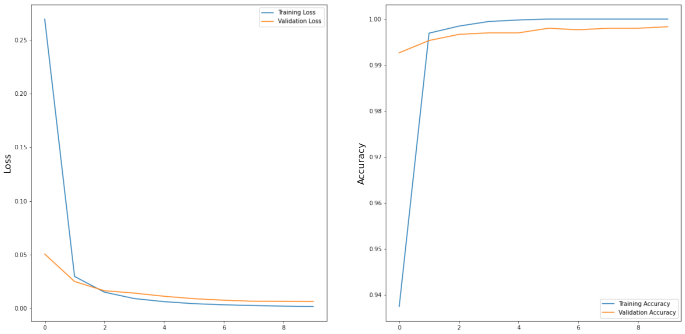

Fig. 10. MobileNet Training-Validation Loss and Accuracy.

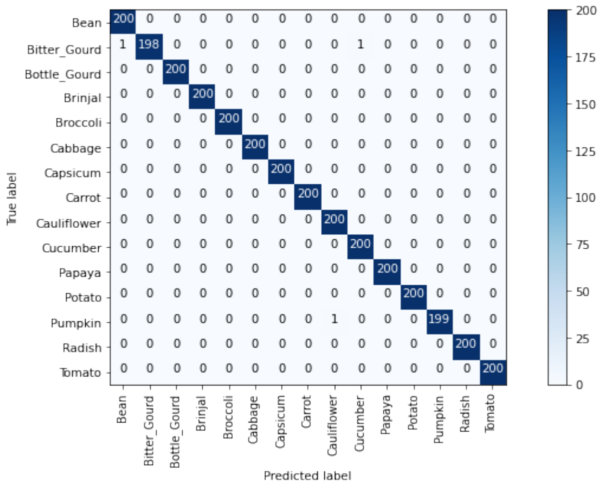

Fig. 11. Confusion Matrix of Fine-Tuned MobileNet.

B. Comparative Analysis

Building a CNN model from scratch is not a very easy task and with a small dataset it's harder to find out the best accuracy from a developed CNN model. For bringing out the best possible accuracy, tweaking the CNN models such as adding more layers, dropout, changing activation function, trying with different optimizer along with learning rate is important. Proposed 6-layer CNN is tweaked and optimized for the working vegetable dataset and gives an accuracy of

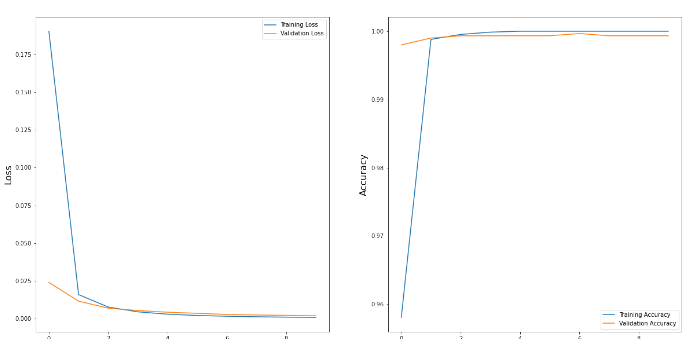

Fig. 12. ResNet50 Training-Validation Loss and Accuracy.

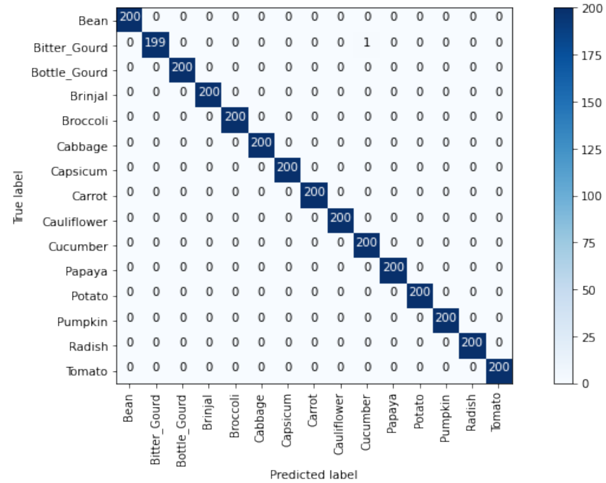

Fig. 13. Confusion Matrix of Fine-Tuned ResNet50.

97.5%, which is the highest compared with all the previous work conducted by building a model from scratch. Table I shows the result summary along with the applied technique. All the previous works done by using state-of-the-art CNN model has no major impact, the accuracy of those models is not significant as most of the experiment was done on a smaller dataset or collected from ImageNet. Table II shows previous methods, dataset size and results. With the proposed fine-tuning approach in state-of-the-art CNN architecture, the output is significantly impressive. All the four DCNN architecture used with transfer learning technique gives the accuracy over 99% each. Maximum accuracy is achieved from MobileNet and ResNet which is 99.9%.

TABLE I
RESULT SUMMARY

Technique	Method/ Algorithm	Epochs	Accuracy	Training Time
Build from scratch	CNN	50	97.5%	>1 hour
Transfer learning	VGG16	10	99.7%	<40 min.
	InceptionV3	10	99.7%	<1 hour
	MobileNet	10	99.9%	<30 min.
	ResNet	10	99.9%	<30 min.

TABLE II
EXISTING METHODS, DATASETS, AND RESULTS

Author	Method/ Algorithm	Dataset Size	Dataset Source	Accuracy
Om Patil et al. [4]	Inception V3	1200	Self collected	99%
Yuki Sakai et al. [5]	DNN	200	Self collected	97.38%
Frida Femling et al. [6]	MobileNet	4300	ImageNet	96%
	Inception V3			97%
Zhu L et al. [7]	AlexNet	24000	ImageNet	92%
Guoxiang Zeng [8]	VGG	3678	Self collected	95.6%

VI. Conclusion and Future Research

Agriculture is the most important sector but this sector is less focused on digitalization than others. There were some works done previously on vegetable classification but those are in limited scope with a very small dataset and less accuracy. By taking account of those problems this research is conducted to resolve those issues. In this study, vegetable image classification is done using a typical CNN model and CNN-based pre-trained VGG16, InceptionV3, MobileNet, and Resnet50 with two techniques. Proposed typical CNN model has six-layer and was build from scratch. On the other hand, pre-trained state-of-the-art CNN architectures are fine-tuned and applied using transfer learning techniques. From various species of vegetables, only 15 types of vegetables are selected for primary research of vegetable image classification task. A dataset consists of 21000 images from 15 classes is created locally and used for training and testing. A comparative study is also done on the performance between typical CNN and pre-trained CNN, to check which one is better, efficient, less time-consuming. Also, experimental results for different models and techniques are discussed and the overall accuracy achieved 99.9%. From the experimental result, it's clear that pre-trained CNN architectures are the future of machine vision. And as of now, it's the highest possible accuracy for the vegetable classification task, which seems quite promising. Also there are some options for future implications. Various types of actual devices can be made by utilizing this work. The sorting and labeling process of vegetables can be automated to save both time and human resources in super shops, warehouses. Extending this work by continuing the study with more classes and types by contributing to the existing dataset, to make it more robust.

References

[1] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[2] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*. IEEE, 2017, pp. 1–11.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] P. Om, G. Vijay *et al.*, "Classification of vegetables using tensorflow." *International Journal for Research in Applied Science and Engineering Technology*, vol. 6, no. 4, pp. 2926–2934, 2018.

[5] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, "A vegetable category recognition system using deep neural network," in *2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*. IEEE, 2016, pp. 189–192.

[6] F. Femling, A. Olsson, and F. Alonso-Fernandez, "Fruit and vegetable identification using machine learning for retail applications," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2018, pp. 9–15.

[7] L. Zhu, Z. Li, C. Li, J. Wu, and J. Yue, "High performance vegetable classification from images based on alexnet deep learning model," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 217–223, 2018.

[8] G. Zeng, "Fruit and vegetables classification system using image saliency and convolutional neural network," in *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE, 2017, pp. 613–617.

[9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the 19th International Conference on Neural Information Processing Systems*, ser. NIPS'06. Cambridge, MA, USA: MIT Press, 2006, p. 545–552.

[10] Z. Li, F. Li, L. Zhu, and J. Yue, "Vegetable recognition and classification based on improved vgg deep learning network model," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 559–564, 2020.

[11] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 757–775, 2020.

[12] M. Hussain, J. J. Bird, and D. R. Faria, "A study on cnn transfer learning for image classification," in *UK Workshop on computational Intelligence*. Springer, 2018, pp. 191–202.

[13] Prabhu, "Understanding of convolutional neural network (cnn) — deep learning," https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148, Mar. 2021.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[15] V. Chauhan, K. D. Joshi, and B. Surgenor, "Image classification using deep neural networks: Transfer learning and the handling of unknown images," in *International Conference on Engineering Applications of Neural Networks*. Springer, 2019, pp. 274–285.

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition (2015)," *arXiv preprint arXiv:1512.03385*, 2016.