Fruit Classification and Plant Disease Identification Using One Shot Learning

A Report submitted

in partial fulfillment for the Degree of

Bachelor Of Technology

in

Computer Science

by

AHAMBARISH SAIKIA (181210005)

JAPMAN SINGH MONGA (181210024)

VARTIKA CHATURVEDI (181210056)

pursued in

Department of Computer Science and Engineering

NATIONAL INSTITUTE OF TECHNOLOGY

DELHI

2022

CERTIFICATE

This is to certify that the project report entitled **Fruit Classification and Plant Disease Identification Using One Shot Learning** submitted by **Ahambarish Saikia, Japman Singh Monga and Vartika Chaturvedi** to the National Institute of Technology, Delhi, in partial fulfilment for the award of the degree of **Bachelor Of Technology in (Computer Science and Engineering)** is a *bona fide* record of project work carried out by him/her under my/our supervision. The contents of this report, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

<Signature>

Dr Chandra Prakash

Department of Computer Science and Engineering

DECLARATION

We declare that this project report titled “Fruit Classification and Plant Disease Identification Using One Shot Learning” submitted in partial fulfilment of the degree of B. Tech in (Computer Science and Engineering) is a record of original work carried out by us under the supervision of Dr. Chandra Prakash, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

NIT Delhi

Ahambarish Saikia
181210005
Japman Singh Monga
181210024
Vartika Chaturvedi
181210056

ACKNOWLEDGMENTS

We would like to express our special thanks to our mentor, **Dr Chandra Prakash Sir** as well as our Director, **Dr Ajay Sharma Sir** for providing us the golden opportunity to work on this project. We are truly delighted on the completion of our project as it was during the course of this project, when we explored the domain of applying Machine Learning techniques into the agricultural field. We would also like to express our heartfelt gratitude to everyone who has helped us write this article, both directly and indirectly.

Last but not the least, we would like to thank our classmates who provided us with their valuable feedback and supported us throughout the course of this project.

Ahambarish Saikia
181210005
Japman Singh Monga
181210024
Vartika Chaturvedi
181210056

ABSTRACT

With the developments taking place in artificial intelligence and machine learning in the neoteric times, machine learning algorithms have proved to be beneficial in the real-world industries. Variegated ML techniques have also been developed to assist humans in the field of agriculture. However, one of the major problems encountered by various studies around the globe has been scarcity of publicly available data topical to this field. As a result, the studies cannot be generalized and are only specific to certain crops/fruits. In this study, this issue has been highlighted with the help of transfer learning. Transfer learning has been applied for 2 tasks; Fruit Classification and Plant Disease Identification. As a result, it has been observed that as the training dataset decreases, accuracy also drastically decreases, which brings us to the issue that if there are very less training images available then the model cannot learn well. Subsequently, to address this problem, in this study, the technique of one-shot learning has been employed for 2 tasks: Fruit Classification and Plant Disease Identification. Siamese network with the proposed sequential CNN has been trained on the prepared datasets. Subsequently, the performance of the model for the classification tasks is evaluated using the N-Way one shot learning for and keeping accuracy as a metric. For the task of plant disease identification, as the images were available in 3 formats, i.e., coloured, segmented and grayscale, hence the process was repeated for each of the available datasets and the accuracies were compared.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ NOTATIONS

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ML	Machine Learning

CHAPTER 1

INTRODUCTION

In recent times, with the rise of artificial intelligence, machine learning, deep learning and neural networks have gained a lot of popularity in both research fields and industry [1]. Earlier in the days when ANN was introduced, there were several dilemmas and hindrances like overfitting problems, vanishing gradients and less data available, which prevented ANNs from being applied in the real-world scenarios. But as the time progressed, with increasing storage capacity, and development of new algorithms and advancement in computational power, neural networks have become a blessing for today's world. There have been numerous fields in which neural networks are being used to perform variegated tasks [2].

CNNs have been used to perform very complex image classification tasks and have proved to be very substantial. Such advances in image processing and classification techniques allow us to perform research and apply the image processing techniques in the field of agriculture [1]. Significant development has been made in the past few years and now, machine learning and deep learning are almost everywhere in this industry. Convolutional Networks are very similar to that of normal neural networks, they are made up of neurons and have their assigned learnable weights and biases. Each neuron performs some dot product (and sometimes also operates non-linearity) on the received input. This network also results in a single differentiable score from the raw image pixels on one end and class scores on the other. They have some loss function too like on the fully-connected layer.

ML has been used for almost every aspect such as yield prediction and crop quality prediction for crop management, soil management and water management for field conditions management, and species breeding and species recognition for species management [3]. It also has been used for prediction of plant diseases on an early basis so that we're able to prevent the disease from spreading. Food security has been a major concern with the expected world population growth of more than 9.7 billion by 2050[1]. Plant diseases are a major threat to the crops and hence, an automated method is needed to predict the disease and simultaneously remove the infected parts.

In this study, we aim to develop an approach for 2 tasks- fruit detection/identification based on fruit images, and plant disease identification based on leaf images. Fruit identification will certainly be useful for a variety of tasks and will help in reducing the manpower by significant factors. On the other hand, plant disease detection will certainly be useful for detection and prevention of diseases among crops and will be beneficial for food security.

As it is not always possible that a lot of data is available as the training examples to be fed into the transfer learning model, we need to address the problem of scarcity of data. This can be done using the technique of One Shot Learning. This technique helps us to train the model by using only a single training example and then classify the input image into the desired category. For such a model to work, we use the Siamese Network. It is a twin network which consists of the same architecture as well as the hyperparameters involved with it.

One Shot learning methodology uses a pair of images that need to be fed into the Siamese Network. Based on the output that we receive from the network, we are able to predict whether the pair of images are similar or not. This helps us to solve the problem of image verification, that is to verify whether the two images belong to the same category or not. But we also need to solve the problem of image classification. We do this by using N way One Shot Learning. In this, we use N number of pairs of images as the testing dataset and out of these the model predicts the one pair that contains the images belonging to the same category.

There are other variants of one shot learning as well, like few shot and zero shot; in which we use a handful of images and zero image to train respectively. These techniques can come to be very handy when we have to deal with very less availability of images. This can be put to use to solve some very big real life problems as the scarcity of data is a huge one in today’s world.

The two tasks, that are,fruit detection/identification based on fruit images, and plant disease identification based on leaf images can sometimes have very less data available to be trained by using a transfer learning model. This technique of One Shot can prove to be very beneficial in such circumstances.

CHAPTER 2

LITERATURE REVIEW

With the rise of machine learning and artificial intelligence in recent times, various research as well as industrial developments have taken place. But there is a major challenge that is faced by the researchers; scarcity of data. A number of scenarios occur in the real world wherein, there is not much information or data available. A general approach in the agricultural field has been through transfer learning, but with the scarcity of data being highly prevalent, the need to explore some new techniques increases. There are a few techniques using which this challenge could be addressed, one such technique is one shot learning. One shot learning not only improves the efficiency but also does not compromise with the computational complexity. It takes only one training image to classify the testing image into one of the categories. As this study aims to use one shot learning for a fruit classification task, it is imperative to look at numerous studies that have been conducted in the topical area.

2.1 Fruit Classification

H Muresan et.al. in their study [4], introduced a new Fruits 360 dataset which contained high quality images of 131 classes of fruits. There are 90000+ images belonging to 131 classes. In addition to this, they conducted numerical experiments by training a basic CNN from scratch for 30 epochs and various different configurations. A top accuracy of 98.66% was achieved by the the network, however, the aim to improve the accuracy by making changes in the network and using some different techniques, along with creating a mobile application for the same purpose.

HM Zawbaa et. al. in their work [5], have presented a fruit recognition system for classifying and identifying the fruit types. The parameters used for this purpose are shape and color. The process consists of three phases: pre-processing, feature extraction, and classification. For feature extraction, they used scale-invariant feature transform (SIFT) other than the shape and color of the fruits. For classification, they used the K-nearest neighbor and support vector machine. The model was run on only three fruits, namely strawberries, oranges, and apples. The highest accuracy achieved while using the shape and color in the SVM as the feature extraction method was 90.91% for apples and 78.8% for oranges. Whereas, when using the SIFT as the feature extraction, they got 96.97% for apples and 87.85% for strawberries. Although the accuracy is high, the number of fruits that they have used in order to conduct this process is very low. The number of varieties of fruits can be increased in order to get better results covering a vast range of fruits.

L Hou et. al. in their study [6] have provided a solution to the problem of identifying the type of fruits present in the supermarkets and have considered the factor that they might be lying on top of the other, hence overlapping some portions of the other fruits. For this, the image regions are extracted using a selective search algorithm, then the regions have been selected by means of entropy of fruit images and are finally fed as the input to the Convolutional Neural Network(CNN). Although, by using the CNN along with a selective search algorithm, the highest accuracy that has been achieved is 99.77%, but the number of different fruits that have been considered is significantly low. They have also not taken into account the other external factors that could affect the results, such as lighting.

For fruit recognition and detection, R Khan et al. developed an enhanced Convolutional Neural Network [7]. The authors had to distinguish things from the backdrop and distinguish between overlapping fruits. Fruit Detection and Fruit Recognition are two aspects of the suggested technique. The outline contour of and recognised as Region Of Interest is created using several pre-processing procedures. A collection of Region of Objects can be used to represent these single photos. They built two distinct neural network models based on the type of the dataset for recognition, keeping the problem of fruit viewpoint in mind. A Convolutional layer performs a set of mathematical operations on each input image and produces a single value output feature map. After that, these layers are usually applied.

Using EfficientNet and MixNet as the classification engine, LT Duong et al. provided a feasible solution [8] for fruit recognition. In addition, two learning strategies - randomization and transfer learning - were presented. This method was tested on a fruit dataset that included 48905 training photos and 16421 testing images. This paper advances to the field of fruit categorization by utilising two new deep neural networks and comparing the proposed dataset results to those of a standard Convolutional Neural Network. A kernel or filter is used by CNN to collect specific features from the input image. They make an attempt to depict spatial and temporal connections. Each filter is convolved with the input feature, resulting in outputs that map to both. A big kernel aids in the retention of high-resolution patterns, whereas a small kernel aids in the extraction of low-resolution patterns. EfficientNet is a method for improving CNN performance by scaling the breadth, depth, and resolution of the CNNs using a set of predetermined scaling components. The input image has three colour channels, R, G, and B, each of which is 224x224 pixels in size. The next layers are reduced in resolution to lower the size of the feature map, but increased in width to improve accuracy. The goal of the network family is to reduce the number of parameters. EfficientNet and MixNet configurations achieve more than 99 percent prediction accuracy, while ConvNet achieves 94.52 percent accuracy. They calculated accuracies for 95 categories (using the Fruit360 Dataset).

WC Seng et.al. [2] have used three different parameters to classify and recognize the different fruits. The different parameters are color-based, shape-based and size-based. This method classifies and recognizes the fruit images based on obtained feature values by using nearest neighbor classification. The accuracy obtained through this method is 90%. Although the accuracy is high, the method is performed on a very small number of datasets. More variety of fruit images could be taken into account to improve the versatility of the model. Furthermore, the texture could also be included as one of the feature analysis in order to gain better discerning of different fruit images.

2.2 Plant Disease Identification

EC Too et.al. in their study [1], evaluated and compared the performances of 6 state-of-art deep convolutional neural network architectures for the image-based plant disease classification task. They classified 38 classes of the Plant Village Dataset, one of the most common publicly available datasets for this classification task. As a result of empirical comparison, they recommended DenseNet 121 as the best network with 99.75 % accuracy for this task with no overfitting and performance deterioration. It also is fast and accurate which is desirable for the mobile applications. However, more research needs to be done to improve the computational time for the model.

Mohanty et.al [9]., in this paper, using Deep Neural Networks for training purposes of the model training is used. The authors have fine-tuned network parameters to improve the training process. The whole dataset is set upon three different versions, first colored, next monochrome and finally segmented of the leaves, in case there was bias on the other two sets. To keep a check on how the approach would work on unseen data and also to keep a track if any approaches were overfitting they ran a test case from 80-20 split till 20-80 split(into 10 step increment). The authors trained two important architectures AlexNet and GoogleNet. They analyzed the performances both with the dataset once training them from scratch and once by adapting to already trained models (trained on the ImageNet dataset) by using transfer learning, without limiting the learning of any layers. With a total of 60 experimental configurations each for 30 epochs and the same hyper-parameters. The overall accuracy varied from 85.53% (AlexNet-Scratch) to 99.34%(GoogleNet-Transfer learning). In this study, Arnal Barbedo et.al. worked on an image based dataset containing 12 plant species with a total of about 56 diseases, each presenting very different characteristics in terms of number of samples, number of diseases and variety of conditions.

In the study[13], transfer learning was applied to a pre-trained model, GoogleNet. Checking the influence of the backgrounds, two seperate CNNs were trained. The first one was trained using raw images, and the second one had segmented images. The results in this paper sought out three different inferences. In some results the background removal brought positive results whereas, in some negative results were observed and in some mixed or no changes were to be seen in the results. A few examples can be observed in terms of common beans, in which the accuracy improved from 65% to 93%. In soybean on the other hand, the accuracy reduced from 86% to 90% and in grapevines, the accuracy remained the same. The use of a limited number of plant species, resulted in a less robust model. And this in turn prevents effective propagation of the advancements made till now like extracting visual information, possibility of multiple disorders in a single plant, and other interference factors. However, defining how many images would be enough is not an easy task. In the future, this study could be extended by introducing a number of different species of plants so as to make the model more robust and hence, more effective in classification of the plant diseases.

Guan Wang et. al. in their study[14], researched about accurate estimation of the disease severity, disease management and yield loss prediction. Deep learning averts the need of using intensive feature engineering and segmentation of the leaves based on threshold levels. A part of the PlantVillage dataset is used in this study. Using an automated method, the authors wanted to highlight the severity of the plant disease by using the leaf images. In all these image-based methods the methodology followed is pretty similar, the image is first segmented from its background and then the required tissue is used for processing. These images are arbitrarily sized, hence they are first normalized and adjusted according to the required resolution. The authors have used two models, one by building the network from scratch and the other by using the conventional transfer learning methodologies. The former consists of only a few convolutional layers with a few filters per layer. This is in turn connected to two fully-connected networks ending with a softmax layer. Initially, as the depth of the model increases the performance also increases. The best accuracy is achieved at 79.3% while the network had 8 convolutional layers. But the result soon depletes as soon as the network's depth crosses the threshold of 8 layers, due to the insufficient training data for a model with a huge number of parameters. For the transfer learning model, they have compared three pre-trained models, namely, VGGNet, Inception-v3 , and ResNet50.The VGG-16 pre-trained model outshines the others and performs the classification task in minimum time. The best accuracy achieved by it is 90.4%. The major drawback of this model is the lack of variety in the dataset that has been used. For the healthy images of the leaves, the researchers have used only healthy images of tomatoes and apples. Similarly, for the infected portion of the dataset, black rot leaves of tomatoes and apples have only been used. This study can be further extended to other crops which will make the model more robust.

Chen Bingcai et.al. in their study [15], have presented a study of amalgamating the idea of smartphones and computer vision using Deep Learning. They have used CNN's architecture in the transfer learning and have used deep feature extraction. For the transfer learning part, they have used VGG16, Google net and Resnet models. Out of the 38 different classes that are available in the Plant Village dataset, they have used only 14 classes. They have used the pre-trained transfer learning models to extract the features for the classification of diseases using KNN (K nearest neighbor) and SVM (Support Vector Machine). However, due to the different parameters of the pre-trained models, they could not specify any particular model as a standard for the classification of diseases in plants. However, for different parameters, they have suggested some models based on the results that they have obtained. The highest accuracy produced is by VGG16 of 97.82% followed by ResNet 50 with an accuracy of 95.38% and Google Net with an accuracy of 95.3%. Although this is a good approach, still we may not be able to use the pre-trained transfer learning models in case of scarcity of data. For the scarcity of data, we need a new approach that could take very less training images and provide us with the necessary results. They have also used very few classes of different plants as their dataset. Using a higher number of classes could have made the model more robust.

Aniirudh Ramesh et.al. have used the Plant Village dataset in their study. In this study, they have focused on one crop, that is, tomato. They have used two pre-trained transfer learning models, that is, AlexNet and VGG16 net and have discussed their results based on the different hyperparameters. In the initial phases of training, the learning rate is

kept extremely small due to the fact that the model used is a pre-trained transfer learning model. However the learning rate is increased to some extent in the later phases of the training process. The final performance can be calculated by using a combination of different hyperparameters. The overall accuracy of AlexNet and VGG16 net is 97.49% and 97.29% respectively. By altering the different hyperparameters, the authors observed that the highest accuracy was observed when the number of images is 373.
However, as this paper focuses on only one crop, that is, tomato, we can say that this model is not robust enough. Also, in case of scarcity of data, this model might not be able to function well as it uses pre-trained transfer learning models and they require a very high number of images for the training process.

2.3 One Shot Learning

Li Fie Fie et.al. in their work[10], realized the shortage of dataset and started to work towards bridging the gap between the scarcity of dataset and the models. They have coined the concept that, by using a few images also, we can train our model so as to produce decent results. This concept works on the prior knowledge that is previously learned. The authors have correlated this new concept with that of probability density function. Using the probability density function along with the Bayesian implementation of it, they were able to test a classification problem based on 101 different categories of data. This concept is a revolutionary concept because before this paper, lakhs of images were required for the training purposes itself. The results produced in this paper are not that satisfactory as this was the first time when the concept of using only a few images for the training purposes were used. Although, the results in this paper might not be that appealing but they give us a great hope that there is definitely a way using which the training can be reduced to a few images only. In this paper as well, the authors ran the tests on 101 categories of images using just a few training images (one to five) to produce models that are able to achieve detection of around 70% to 95% .

Brenden M. Lake et.al. in their study[11], for the first time, used only one training image to classify which character the stroke refers to. They have used only one image for training purposes. For the testing portion, they have used 20-way classification. In this, there are 20 images out of which only one image is similar to the testing image. They have used the MNIST dataset to work on this research. It is the first time when somebody has actually used only a single image so as to classify which category of character the stroke belongs to. The testing was done on three different models, the stroke model (that is the one described in this paper), the Deep Boltzmann Machine (described in DBM, Salakhutdinov & HInton in 2009), and the Nearest Neighbour Model. After a number of rounds of testing, they were able to clearly depict that the stroke model outshined the others significantly. This model gives us a hope that using only a single image, the training for a model can be completed and can produce satisfactory results. This model outshines the previous models that were used for this category.

Gregory Koch et.al. in their work [12]introduced the concept of Siamese Network in One Shot Learning. There have been a number of concepts to implement One shot learning. However, the Siamese Network seems to be the most widely used concept. In this paper, the authors have taken the Omiglot dataset to achieve this task. The work is divided into two broad sections. In the first section, they perform the verification task based on the image pairs, stating whether the two images(pair of images) are the same or not. The second part is testing that the entered image is similar to which of the images from the given batch. The Siamese network uses the same weight while working with a pair of images to generate the feature vectors of the given image. In the given paper, they are using the Siamese Network for character recognition.
They have also compared their work with the already defined and trained working models and have outshined all of them. They have used a number of layers for their deep CNNs' architecture. For future work, we can try to vary the layers that have been used. Along with that we can also extend our future work to improve the quality of the Siamese Network.

CHAPTER 3

METHODOLOGY

3.1 Dataset Description

In this study, the proposed datasets are Fruits-360 and Plant Village Dataset. Both the datasets are publicly available on Kaggle

.

In the **Fruits-360 Dataset [4]**, there are 131 classes of fruits and vegetables, with a total of 90483 images. These images are divided into 67692 training and 22688 testing images. The images are scaled to 100x100 pixels. It is done, so as to avoid the confusion between different fruit sizes due to the smaller images. Also, the background of all the images was also removed in order to remove the bias. This dataset helps us to recognize the fruit and its variety.

(a) (b) (c)

Fig 3.1: Images of (a) Apple Braeburn, (b) Avocado Ripe, and (c) Banana from Fruit-360 Dataset.

The **Plant Village Dataset** contains over 54306 images of healthy and infected leaves of crops. The dataset has overall images of leaves covering 26 different diseases. These images are divided in the ratio of 80:20, that is 80 percent of the images (43445) are training images and 20 percent of the images are testing images (10861). These images available in three formats- RGB, Grayscale and Segmented. The RGB images consist of colourful images whereas the grayscale images consist of black and white images of the leaves. The segmented images mean that the background of those images has been removed so as to reduce the bias. This dataset helps us to identify the disease with which the plants are infected.

(a) (b) (c)

Fig 3.2: Images of Apple Scab in (a) RGB, (b)Grayscale, (c) Segmented form.

3.2 Description of Transfer Learning and One Shot Learning

3.2.1 Deep convolutional neural networks/transfer learning

One of the most mainstream approaches is to perform a classification task using the deep convolutional neural networks. CNNs are mainly associated with image classification or computer vision tasks and have eased these tasks to a great extent. CNNs make use of the concept of convolution to perform mathematical operations. CNNs are composed of convolutional, pooling and fully connected layers.

The combination of CNNs with the concept of transfer learning has proven to be a blessing to humanity. Transfer learning is a strategy for taking knowledge obtained from one problem and applying it to a different but related problem. Many pre-trained models (like VGG 16, VGG 19, Dense Net etc) which are trained on a large ImageNet data-set are available as open source and can be used to develop more advanced models by using the pre-trained weights.

Some advantages of CNNs are: CNNs can make use of spatial data with the use of convolutions. The pre-processing required in CNNs are pretty lower as compared to other classification algorithms. Also, less computing power required for the training and testing phases.

CNNs are more useful in picture classification because they can successfully capture the Pixel dependencies, as well as their Spatial and Temporal dependencies, by adding filters to the images/dataset. The design also performs well because the number of hyperparameters is reduced and weights are reused. Traditional Neural Networks may lose image features and hence hinder the classification process. This is an important point to note because with large datasets our architecture should also be prepared to handle such feature extraction without losing any detail.Some famous architectures of the CNNs publicly available are VGGNet, AlexNet, ResNet, GoogleNet etc.

The Kernel is the most important element in carrying out the convolutional operation of the neural network. It is generally fitted to the image according to pixels. Let’s say I have an image with 5x5x4 input, then usually our kernel might be anything between 2x2x2 to 4x4x4. This kernel moves around the image and convolves with the image portion by making strides. The objective of the Convolution enables us to extract high-level features from the image like pixels, edge discovery etc.

The Pooling Layer shrinks the convolved feature's spatial size. As a result, dimensionality reduction reduces the computational potential of the data. It can also be used to extract prominent features that are both positional and rotationally invariant. There are different kinds of Polling layer and also, they have each of their unique functionality like Max Pooling returns maximum value, Average returns the average of all the values from the position of the image covered by the kernel.

We'll flatten the image into a column vector now that we've turned it into a format suited for our Multi-level Perceptron. This flattened output is fed into a feed-forward NN, and backpropagation is applied to each iteration of the training. Over some epochs the difference in dominating and low-level features is developed and they are fed into a SoftMax layer so that the classification algorithm can run through those test images.

Hence, in this study, we aim to propose a deep learning model with the help of transfer learning to perform the image classification tasks on the aforementioned datasets.

3.2.2 One shot learning

One shot learning is the approach which aims to remove the problem of scarcity of labelled dataset. Scarcity of labelled dataset is one of the major problems of classification tasks that are performed using deep neural networks. Less dataset leads to underfitting of the model and the model not being able to generalise well. Some of the aforementioned research articles in the literature review section have also mentioned the problem of less dataset available for some categories of fruits.

To solve the problem of scarcity of data, many methods have been proposed. One such approach is termed as one-shot learning, in which the model is fed with only one training example so as to classify the test image to a particular class. There have been a number of attempts to reduce the number of training dataset as presented in the work Fei-Fei et al[10] In the mentioned paper, the authors reduced the number of training images to a handful of images and then tested them so as to show that a model can learn using a smaller number of images as well.

One shot learning was proposed first by Lake et al.,[11]. The authors first introduced to the world that it is possible to use only a single image to extract the required information so as to classify the test images to a particular category. One shot learning is widely used today in the field of machine learning. Few applications of the given concept include face recognition, passport checks, object image classification, low data drug discovery etc.

The concept of one-shot learning is done using the Siamese network, as shown in Fig 3.3. This network, sometimes known as the twin network, is a neural network that consists of the same architecture along with the same weights (hence the name, twin network). The twin network consists of two sequential convolutional neural networks (CNNs), which are fed with a pair of images. These sequential CNNs then give us the feature vectors as output. By analysing these feature vectors so obtained, we can determine that the images that were fed into the twin network were similar or not. This kind of network was first introduced by authors named Bromely and LeCun in the 1990s so as to address the problem of image matching, in the case of signature verification tasks.

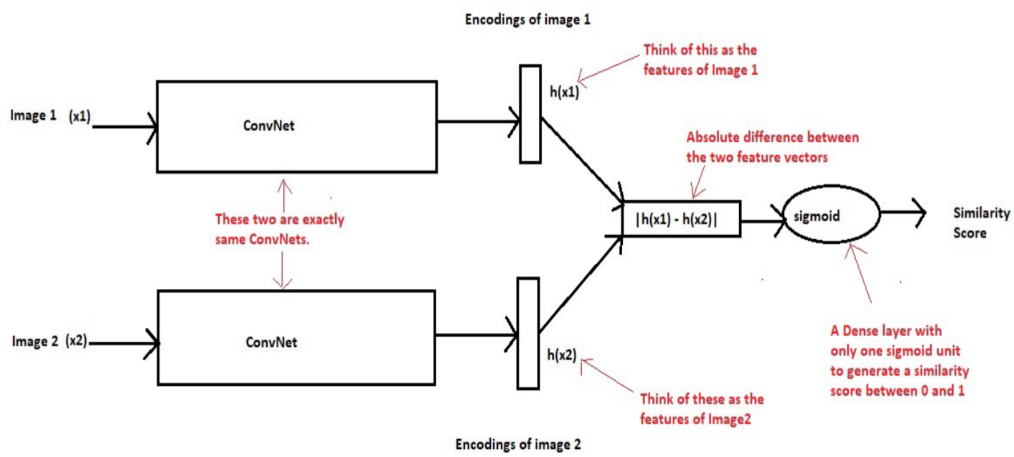


Fig 3.3 Siamese Network for One Shot Learning

In a task of one-shot classification, we require only one training example of each class. 2 images are fed into a model, and the model learns a similarity function between the images. The similarity score thus generated is scaled between 0 and 1 using a sigmoid function, 0 being the low similarity score, and 1 being the highest similarity score. Hence, in order to classify the new image, only one reference image (from the training set) is needed. This solves the aforementioned problem of the scarcity of dataset.

The major difference between one shot learning and the traditional transfer learning is that transfer learning requires a lot of images as the training dataset and the former requires only a single image for the same. Although transfer learning has been a traditional and widely used approach, it cannot always be the optimised approach as it requires a lot of images. This can be seen especially in the scenario when there are thousands of categories to which the model can classify the image into. In such a case, for transfer learning to hold, we will be needing to train the model or to give the model a lot of images for each of the thousands of categories so mentioned. But, in one shot learning, we only need to provide the model with one training image per category so as to classify the testing image into one of the categories.

For the tasks that have to be completed in this project, one shot learning can be very useful, as with the smaller number of images of the fruits also, we can train our model so as to recognise the type of fruit. This can also be beneficial when rare fruits are to be recognised. In such cases, even with the scarcity of the dataset, we will be able to recognise the fruit, using one shot learning to train our model. We can further extend one shot learning for the second task as well, that is for plant disease detection using the leaves of the infected plants. This methodology will not only help us to detect rare diseases, but will also help us to use fewer training dataset in order to complete the task.

There are a few variants of this approach, called few shot learning and zero shot learning. In few shot learning, we use a handful of images to train the model as can be seen in Fei-Fei et al.[10]**.** This type of approach is helpful when we have more than one image but still very less number of images available for that particular category. The application of such an approach can be seen while training the model for rare cases in different categories. Zero shot approach is also an emerging approach that can be seen nowadays. It uses the metadata of the image, that is the properties, appearance and functionality to classify the image into a particular category. This kind of approach cannot be carried out without prior knowledge of the image that needs to be identified or classified. One of the applications of such an approach can be seen when we are aware of the image, for example if we know that the constellation Cassiopeia looks like a distorted ‘W’, then we can identify it in the night sky.

3.3 Fruit Classification

3.3.1 Transfer Learning for Fruit Classification

In order to establish the baseline accuracy and observe the trend for the performance metrics in accordance with size of training dataset/number of training images, we've employed the concept of transfer learning to train a basic deep CNN model. VGG 16 model has been trained on 6 different train-validation-test split ratios in order to observe the variation of performance metrics with the size of training data. Fig 3.4 gives the diagrammatic representation of the architecture used for training. The process was repeated for each of the split ratios.

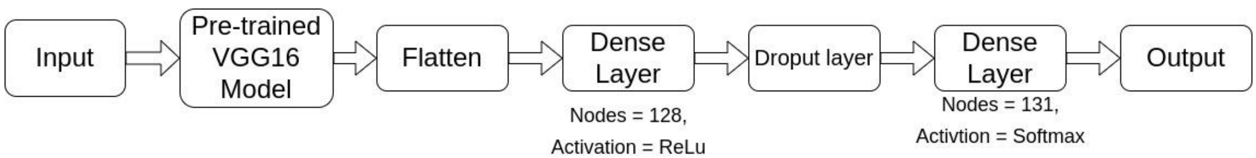


Fig 3.4: Fruit Classification; Diagrammatic representation of the architecture used.

For training the model, first the input images are fed on the pretrained VGG-16 network the weights of which have been trained on Imagenet dataset. The weights of this network were set to be non-trainable. The top layer of the network was excluded.

Karen Simonyan and Andrew Zisserman of the Oxford Robotics Institute created the Visual Geometry Group Network (VGG) based on the convolutional neural network architecture. At the 2014 Large Scale Visual Recognition Challenge, it was addressed (ILSVRC2014). Instead of a huge number of hyper-parameters, VGG networks concentrated on having 3x3 filter convolution layers with stride 1 and always used the same padding and maxpool layer of a 2x2 filter with stride 2. Throughout the architecture, the convolution and max pool layers are arranged in a regular manner. VGG 16 has 16 layers and hence a smaller number of trainable parameters than VGG 19. The images are augmented in terms of rotation of 15 degrees with respect to the original orientation of the image. This is to prevent overfitting of the image.

The output tensor from the pretrained network were flattened out to give a 1-D array which serves as an input for the next layer, which is a fully connected layer. Like ANN is an information passing technique, it includes a large number of connected processing units that work together to process information. This generates meaningful insights that can be used further in the architecture down the line. Similarly, to ANN this Fully-Connected Layer works. It passes the previously output information as input, and passes the linear calculated function of inputs through an activation function to gather the output for that node. Again, this output serves as the input for the upcoming next layer in-line. For this Fully-Connected layer, 128 nodes have been used and the Re-Lu function has been used for operation. This activation function helps in learning about some more complex features with adding non-linearity in the network.

Further, we've used a dropout of 0.5. This dropout layer is quite famous given by Geoffrey Hinton published whitepaper 'A Simple Way to Prevent Neural Networks from Overfitting'. It keeps the model from being overfit. At each update of the training phase, Dropout works by setting the outgoing edges of hidden nodes in the network to 0 at random. The dropout rate in our scenario is 0.5. This means that only half of the randomly chosen neurons will travel through the network for the following layer's input (of the next layer). It is a regularisation method for neural networks that aids in decreasing the interdependence of the neurons' learning. As a result, overfitting can be avoided.

Because our classification is based on 131 classes, the output of the fully connected layer with dropout was then fed into a softmax classification layer with 131 nodes. Because this is a softmax layer, the softmax activation function is used. Softmax expands the idea of producing probability from logistic regression. In a multi-class situation like ours, it assigns decimal probability to each class. The model's output is calculated in this layer.

Table 3.1: Fruit Classification; Name of layers and corresponding hyperparameters used in our approach.

Name of Layer	Hyperparameters
Flatten Layer	
Hidden Neural Dense layer	Number of nodes = 128, Activation function = ReLu
Dropout Layer	Dropout value = 0.5
Output Dense Layer	Number of nodes = 131, Activation function = Softmax

Table 3.2: Fruit Classification; Hyperparameters used for training in Transfer Learning Approach

Hyperparameter Name	Value
Train and Validation Batch Size	32
Number of Epochs	15
Learning Rate	10e4
Data Scaling Size	100*100
Optimizer	Adam Optimizer
Loss Function	Binary Cross Entropy Loss

Through the results, it was evident that accuracy was directly proportional to the training dataset size. As the number of training images decreased, the accuracy decreased. Hence, it was evident that in a task like this, the training images in real world are scarce, and a model cannot learn with such a less amount of dataset available. Hence, we've proposed to implement one shot learning to rectify this issue.

3.3.2 One shot learning

To address the problem of scarcity of dataset, we propose to implement one shot learning. Fig 3.5 gives the diagrammatic representation of the steps involved in implementing one shot learning. Initially, as the need of dataset for one shot learning task is low, the dataset is preprocessed and prepared according to the needs. Thereafter, the dataset is split in train and test set. The prepared dataset is fed in the Siamese network for training. Subsequently, the performance of the model is evaluated using N-way one-short learning.

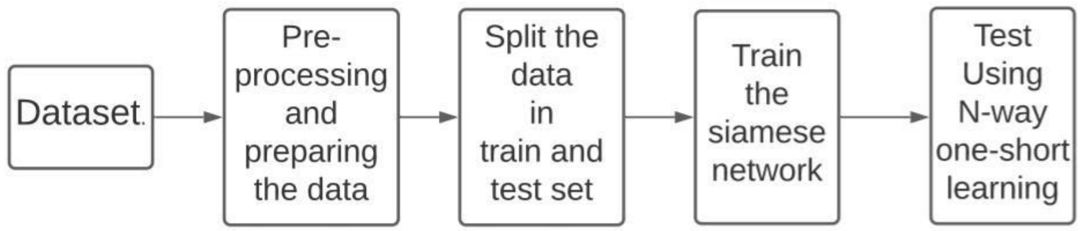


Fig 3.5 : Fruit Classification; Diagrammatic representation of the proposed methodology for one shot learning

Fig 3.6 gives the diagrammatic representation of the Siamese network that we've prepared for the Fruit Detection task. Siamese network takes 2 images as input and feeds them to a sequential convolutional neural network model. Sequential model gives the feature vectors as the output for each of the images, which are then subtracted and passed on to the final dense layer to generate the output. We calculate the Euclidean Distance between the two feature vectors so as to understand whether the images are similar or not . The model learns a similarity function between the images. The similarity score/Euclidean distance thus generated tell us about the similarity of images. Higher the distance, less similar are the images and vice versa.

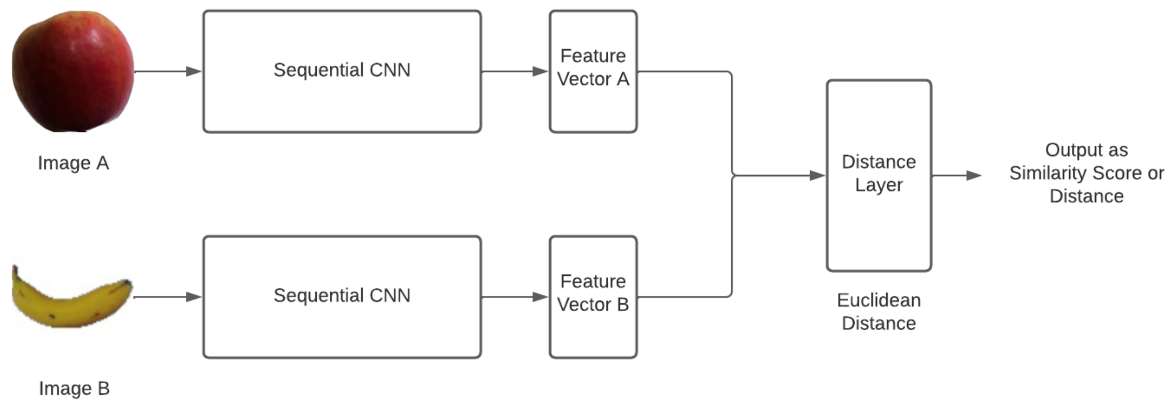


Fig3.6: Fruit Classification; Siamese Network used in our proposed methodology.

For the Siamese network, we need the dataset to be in pairs, and as the dataset that we have is not in the form of what is required, so to implement the above-mentioned methodology, we first have to prepare the dataset according to our needs. That is, to perform the one-shot methodology, we need to have a pair of images to be fed into the Siamese network. The dataset can be classified into two types, genuine pair and imposite pair.

The genuine pairs are those pairs which have the images of the fruits belonging to the same class, whereas, the imposite pairs will have the images of the fruits belonging to different classes. The genuine pairs are labelled as one, which denotes that they belong to the same category. Whereas the imposite pairs are labelled as zero to determine otherwise. In total, we have 20,000 pairs of images, which include 10,000 images from each genuine and imposite pair. To maintain uniformity, we have taken 20 images each from the 131 classes of fruits that we had initially in the dataset. So, we have a total of 2620 images of fruits, out of which we have made 20,000 pairs of images, randomly. The Fig 3.7 shows the genuine and imposite pairs of images that we have created (the dataset to be fed into the Siamese network).

(a) Genuine Pair

(b) Imposite Pair

Fig 3.7: Dataset Preparation for Fruit Classification; Two types of pairs, (a)Genuine Pair, (b) Imposite Pairs

Now, as we know that Siamese network takes 2 images as input and feeds them to a sequential convolutional neural network model which in turn gives us the feature vector. So, Fig 3.8 below gives the architecture of sequential CNN used in our project.

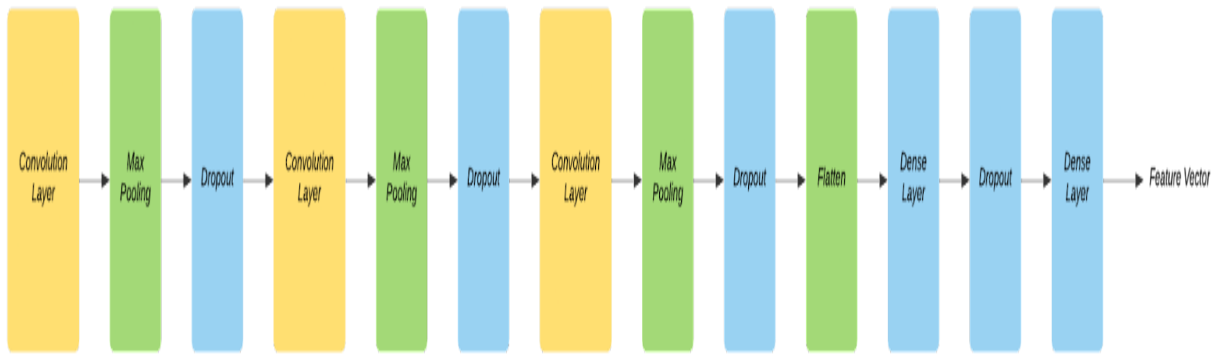


Fig 3.8: Fruit Classification; Architecture of the sequential CNN that we have used.

The Sequential CNN consists of a convolutional layer, followed by the max pooling layer and then the dropout layer. This architecture is repeated thrice. After that, the flatten layer is applied, after which again we have the dense layer, dropout layer and then again, the dense layer. This is the architecture for the sequential CNN used in our architecture.

Table 3.3: Fruit Classification; hyperparameters used in our network.

Hyperparameter Name	**Value**
Optimizer Used	RMS prop
Epochs	20
Loss	Contrastive Loss

We have used contrastive loss as our loss function. The Siamese network can be trained using many types of loss functions. But contrastive loss and triplet loss works better with one shot learning as compared with binary cross entropy loss.

The equation for contrastive loss is given below:

$$Contrastive\ Loss = \frac{1}{n}\sum_{n}(1 - Y_{true}) * {D_w}^2 + Y_{true} * \{\max(0, margin - D_w)\}^2$$

Where N is the number of training examples, Y_{true} is the Actual label (1 for genuine pair, 0 for imposite and D_w is the Predicted Euclidean Distance between the two feature vectors by the network. We've kept Margin as 1 (typical).

In order to evaluate our Siamese network, we employ the concept of N-Way one shot learning. To understand what exactly is N-Way one shot learning, we take an example of 4-Way One shot learning. Fig 3.9 shows the 4 pairs that we've generated. Out of these pairs, only one pair (pair A) is a genuine pair, i.e. belonging to same class, while all of the other pairs (Pair B, Pair C, and Pair D) are imposite pairs, i.e. belonging to different classes.

Now, all of the pairs are fed one by one into our trained Siamese Network. For each of the pairs, our Siamese network generates a similarity score, say S1 for pair A, S2 for pair B, S3 for pair C, and S4 for pair D. Now, in an ideal case, pair belonging to same class should return highest similarity score or the least distance.

(a) Pair A (b)Pair B

(c) Pair C (d)Pair D

Fig 3.9: Fruit Classification; Support set for 4-Way N-Shot Learning with 1 genuine pair(Pair A) and 3 imposite pairs.

So, from amongst the similarity scores generated, pair with highest similarity score or the least distance is marked as a prediction and is predicted as genuine. Further, the prediction is compared with actual label. This process is repeated for a number of iterations. Accuracy is calculated by dividing correct predictions by total iterations.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Iterations} * 100\ \%$$

Similarly, for N-Way One Shot Learning, for one iteration, to generate the support set, one random class is chosen. Then, an image is randomly chosen from that class, say Image A. Now to create a genuine pair, one more random image is chosen from that class and a pair is made with Image A. Now, (N-1) imposite pairs are generated by randomly selecting (N-1) classes and further selecting a random image from each of (N-1) classes. Hence, a support set of N pairs, with 1 genuine pair and N-1 imposite pairs is generated. All the pairs are fed into the network, similarity scores or the Euclidean distance is generated and class with highest similarity score with the reference class is predicted. The predicted class is compared with the actual label. This process is repeated for a number of steps/ iterations and accuracy is calculated using the above formula.

In our methodology, we implemented N-Way one shot learning for N=5,10,15,20,25,30,35,40. Also, for each of the mentioned N, the process was repeated for 100,150,200 iterations. Subsequently, the accuracies were noted and the line plots were plotted to visualize the variation of accuracy with N.

3.4 Plant Disease Identification

3.4.1 Transfer learning

Similar to the task of fruit classification, transfer learning has been used in order to establish the baseline accuracies. VGG 16 model has been trained and evaluated for 6 train-test split ratios. As the plant village dataset has 3 types of images of plant leaves, i.e.: grayscale, color and segmented. So, as the trend will be common for all the three images, so we base our study only on the colored images.

Fig 3.10 gives the diagrammatic representation of the architecture used for training. The process was repeated for each of the split ratios.

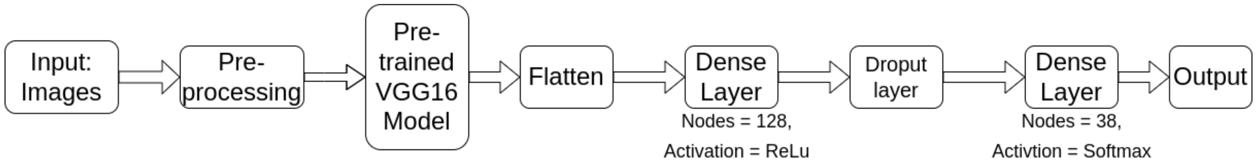


Fig 3.10 Diagrammatic representation of the architecture used for transfer learning in plant disease identification.

For training the model, first the input images are fed on the pretrained VGG-16 network the weights of which have been trained on Imagenet dataset. The weights of this network were set to be non-trainable. The top layer of the network was excluded.

The output tensor from the pretrained network were flattened out and a array of one dimension was obtained and was fed into next layer; fully connected layer. This output acts as the input for the next layer in-line once more. 128 nodes were used in this Fully-Connected layer, and the Re-Lu function was used to operate it. With the addition of non-linearity to the network, this activation function aids in the learning of more complicated features. A 0.5 dropout has been used. This means that only half of the randomly chosen neurons will travel through the network for the following layer's input (of the next layer). It is a regularisation method for neural networks that aids in decreasing the interdependence of the neurons' learning. Thus, helping to prevent overfit. The output of dense layer with dropout was subsequently fed into a softmax classification layer with 38 nodes because our classification is based upon 38 classes. Because this is a softmax layer, the softmax activation function is used. Softmax extends the logistic regression idea of producing probabilities. In a multi-class situation like ours, it assigns decimal probability to each class. This layer derives the output of the model.

Table 3.4 Plant Disease Identification; Name of layers and corresponding hyperparameters used in our approach.

Name of Layer	Hyperparameters
Flatten Layer	
Hidden Neural Dense layer	Number of nodes = 128, Activation function = ReLu
Dropout Layer	Dropout value = 0.5
Output Dense Layer	Number of nodes = 38, Activation function = Softmax

Table 3.5: Hyperparameters used for training in Transfer Learning Approach

Hyperparameter Name	Value
Train and Validation Batch Size	64
Number of Epochs	30
Learning Rate	10e4
Data Scaling Size	128*128
Optimizer	Adam Optimizer
Loss Function	Binary Cross Entropy Loss

Through the results, it was evident that accuracy was directly proportional to the training dataset size. As the number of training images decreased, the accuracy decreased. Hence, it was evident that in a task like this, the training images in real world are scarce, and a model cannot learn with such a less amount of dataset available. Hence, we've proposed to implement one shot learning to rectify this issue.

3.4.2 One shot learning

As described in section 3.2.2, one shot learning has been implemented for fruit classification. Similarly, we implement one shot learning for the task of plant disease identification to address the problem of scarcity of dataset. Fig 3.5 gives the diagrammatic representation of the steps involved in implementing one shot learning. Initially, as the need of dataset for one shot learning task is low, the dataset is preprocessed and prepared according to the needs. Thereafter, the dataset is split in train and test set. The prepared dataset is fed in the Siamese network for training. Subsequently, the performance of the model is evaluated using N-way one-short learning.

As we know that Plantvillage dataset consists of images in 3 forms; color, segmented and grayscale. In this study, we perform one shot learning on all the 3 datasets separately. Fig xx gives the diagrammatic representation of the steps involved in the methodology.

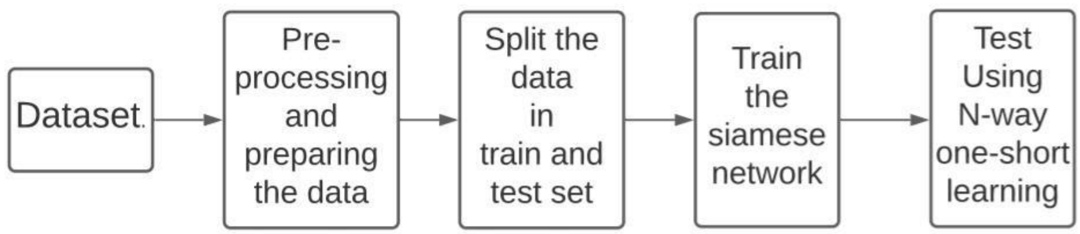


Fig 3.11 Plant Disease Identification; Diagrammatic representation of the proposed methodology for one shot learning

Siamese network takes 2 images of plant leaves as input and feeds them to a sequential convolutional neural network mode which gives feature vectors as the output for each of the images, which are then subtracted and passed on to the final dense layer to generate the similarity score as output so as to understand whether the images are similar or not . The model learns a similarity function between the images. The similarity score/Euclidean distance thus generated tell us about the similarity of images. Higher the distance, less similar are the images and vice versa.

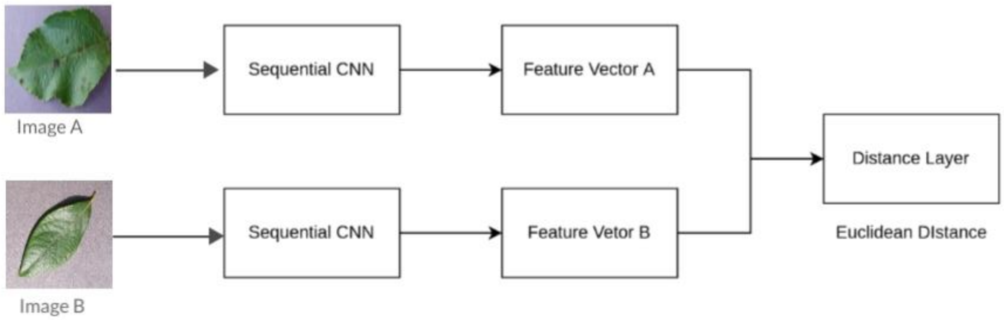


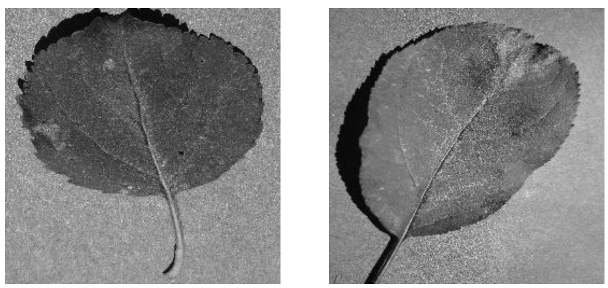
Fig 3.12: Plant Disease Identification; Diagrammatic representation of the proposed methodology for one shot learning

As we know that data needs to be in form of pairs to feed in the Siamese netwoks, so the first ste in the pipeline is to prepare the data in pairs. The genuine pairs are those pairs which have the images of the fruits belonging to the same class, whereas, the imposite pairs will have the images of the fruits belonging to different classes. The genuine pairs are labelled as one, which denotes that they belong to the same category. Whereas the imposite pairs are labelled as zero to determine otherwise. For 1 dataset, we have 18,000 pairs of images, which include 9,000 images from each genuine and imposite pair. To maintain uniformity, we have taken 18 images each from the 38 classes of plant leaves that we had initially in the dataset. So, we have a total of 684 images of plant leaves, out of which we have made 18000 pairs of images, randomly. The process is repeated for all the 3 dataset; color, grayscale and segmented. So, we obtain 3 datasets with 18000 pairs of images in each of the dataset.

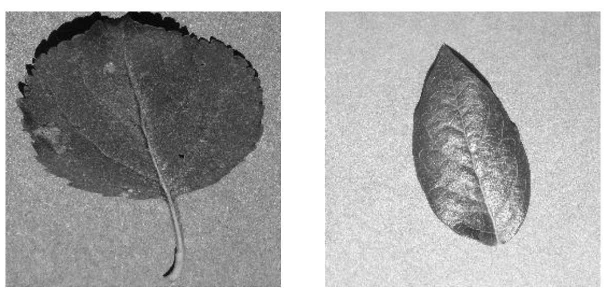
(a)Genuine Pair

(b) Imposite Pair

Fig 3.13: Two types of pairs for colored images, (a)Genuine Pair, (b) Imposite Pairs



(a)Genuine Pair



(b)Imposite Pair

Fig 3.14: Two types of pairs for grayscale images, (a)Genuine Pair, (b) Imposite Pairs

(a)Genuine Pair

(b) Imposite Pair

Fig 3.15: Two types of pairs for segmented images, (a)Genuine Pair, (b) Imposite Pairs

As Siamese network takes 2 images as input and feeds them to a sequential convolutional neural network model which in turn gives us the feature vector. So, Fig 3.8 below gives the architecture of sequential CNN used in our project.

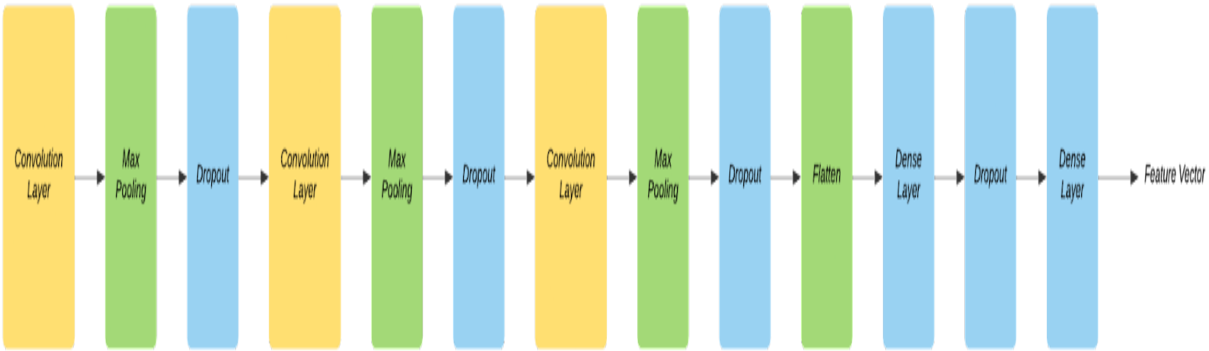


Fig 3.16: Plant Disease Identification; Architecture of the sequential CNN that we have used.

Table 3.6 : Plant Disease Identification; hyperparameters used in our network.

Hyperparameter Name	**Value**
Optimizer Used	RMS prop
Epochs	25
Loss	Contrastive Loss

In order to evaluate our Siamese network, we employ the concept of N-Way one shot learning.For N-Way One Shot Learning, for one iteration, to generate the support set, one random class is chosen. Then, an image is randomly chosen from that class, say Image A. Now to create a genuine pair, one more random image is chosen from that class and a pair is made with Image A. Now, (N-1) imposite pairs are generated by randomly selecting (N-1) classes and further selecting a random image from each of (N-1) classes. Hence, a support set of N pairs, with 1 genuine pair and N-1 imposite pairs is generated. All the pairs are fed into the network, similarity scores or the Euclidean distance is generated and class with highest similarity score with the reference class is predicted. The predicted class is compared with the actual label. This process is repeated for a number of steps/ iterations and accuracy is calculated using the above formula.

In our methodology, we implemented N-Way one shot learning for N=5,10,15,20,25,30,35. Also, for each of the mentioned N, the process was repeated for 100,150,200 iterations. The process is repeated for each of the 3 trained Siamese networks of each of the 3 datasets, i.e. grayscale, color and segmented.

CHAPTER 4

RESULTS

4.1 Results of Transfer Learning for Fruit Classification

The results obtained after training the model for 15 epochs are stipulated below. Table 1 shows the values of various performance metrics obtained for the training phase, for each of the six different splits. Table 2 shows the shows the values of various performance metrics obtained for the testing phase, for each of the six different splits.

Table 4.1: Fruit Classification; Values of performance metrics obtained for different split ratios.

Train-Test Ratio	Testing Accuracy	Testing AUC	Testing Precision	Testing Recall	Testing F1 score
10%-90%	77.46	0.984	0.962	0.655	0.779
50%-50%	90.79	0.9831	0.9335	0.8961	0.912
60%-40%	92.76	0.9864	0.9398	0.9215	0.931
67%-33%	94.13	0.9899	0.9508	0.9375	0.944
70%-30%	94.82	0.9903	0.956	0.9452	0.951
90%-10%	99.75	0.99	0.998	0.996	0.996

From the above tables, we can infer that the as expected, the split ratio of 90-10 yielded the highest testing accuracy of 99.75 %. With the increase of training data size, there is a gradual increase in each of the performance metrics. As evident from the training loss and accuracy graphs, there were no signs of overfitting and the graphs were smooth.

4.2 Results of One-Shot Learning for Fruit Classification

As we all know that, Siamese network takes 2 images as input and feeds them to a sequential convolutional neural network model. Sequential model gives the feature vectors as the output for each of the images, which are then subtracted and passed on to the final dense layer to generate the output. So, after training our network with 13600 pairs, and validating on 3400 pairs, following results were obtained. It is worthy to note that, till this step, these results show that how good our model is performing on the verification task, i.e., if 2 images are fed, then whether our model is able to tell if they belong to same category or not. This is basically a verification task.

Table 4.2:Fruit Classification; Results of One-Shot Learning after training.

Phase	**Accuracy**
Training	92.94%
Validation	92.97%
Testing	87.5%

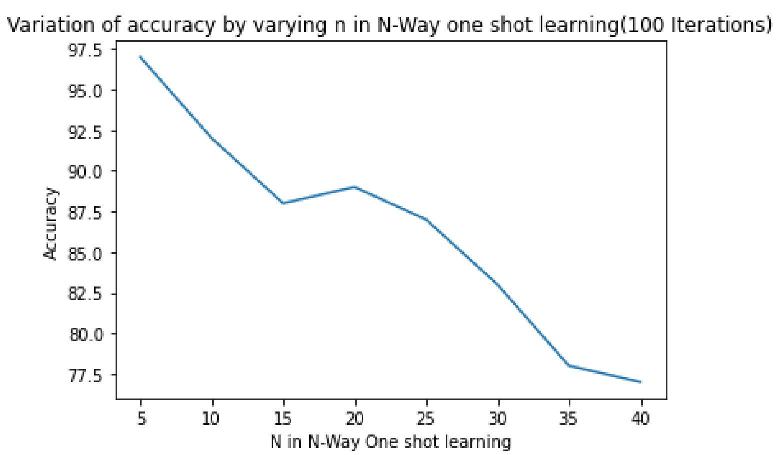
To evaluate our model for a classification task, we used the methodology of N-Way One shot learning. for N-Way One Shot Learning, for one iteration, to generate the support set, one random class is chosen. Then, an image is randomly chosen from that class, say Image A. Now to create a genuine pair, one more random image is chosen from that class and a pair is made with Image A. Now, (N-1) imposite pairs are generated by randomly selecting (N-1) classes and further selecting a random image from each of (N-1) classes. Hence, a support set of N pairs, with 1 genuine pair and N-1 imposite pairs is generated. All the pairs are fed into the network, similarity scores or the Euclidean distance is generated and class with highest similarity score with the reference class is predicted. The predicted class is compared with the actual label. This process is repeated for a number of steps/ iterations and accuracy is calculated using the above formula.

In our methodology, we implemented N-Way one shot learning for N=5,10,15,20,25,30,35,40. Also, for each of the mentioned N, the process was repeated for 100,150,200 iterations. Subsequently, the accuracies were noted and the line plots were plotted to visualize the variation of accuracy with N. Table 5 gives the variation of accuracies.

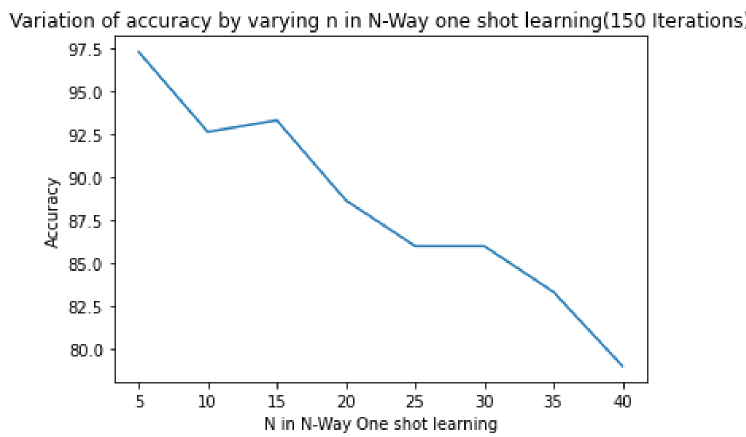
Table 4.3: Fruit Classification; Variation of accuracy(%) with number of pairs and number of iterations

	N=5	**N=10**	**N=15**	**N=20**	**N=25**	**N=30**	**N=35**	**N=40**
100 Iterations	97%	92%	88%	89%	87%	83%	78%	77%
150 Iterations	97.33%	92.66%	93.33%	88.66%	86%	86%	83.33%	79%
200 Iterations	96%	94%	90.5%	86%	85%	82%	81.5%	80%

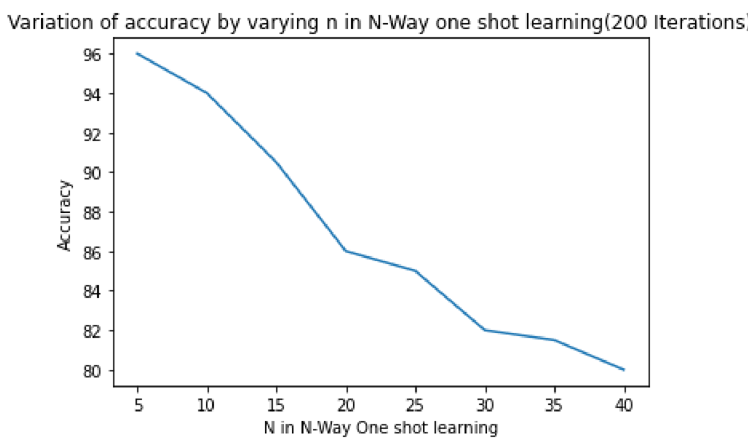
Fig 4.1 gives the plot of variation of accuracy with N, for each of the iterations.



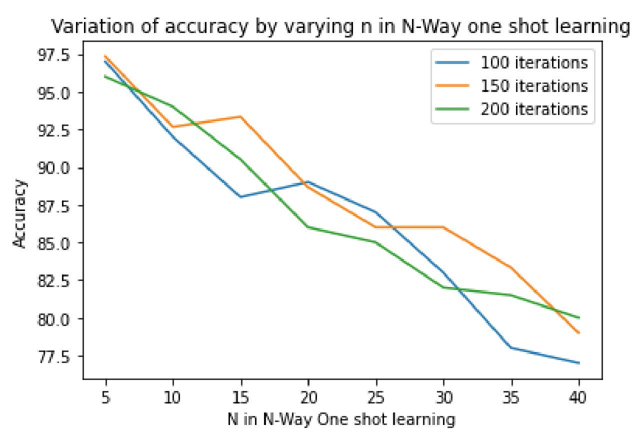
(a)



(b)



(c)



(d)

Fig 4.1: (a) Variation of accuracy by varying n in N-way One Shot Learning for (a) 100 iterations, (b)150 iterations, (c)200 iterations. Also, (d) gives the combined plot for all 3 iterations

By Table 5 and plots of fig 6, it has been observed that for all the iterations, there is an expected gradual decrease in accuracy with increase of N. When N-Way one shot learning is performed for 100 iterations, top accuracy of 97% is observed for N=5 and least accuracy of 77% is observed. As the number of iterations are increased, the testing becomes more stringent and robust. As the iterations are increased to 150, a top accuracy of 97.33 % is reported along with least accuracy of 79%. For 200 iterations, a top accuracy of 96% is noted for N=5, along with a least accuracy of 80%. Similar trends can be visualized in the plots described in Fig 6. It can be observed that even with a small number of training images, a least accuracy of 80% is obtained for a classification task with 40 classes (tested through 200 iterations). For N=5, the accuracy goes as high as 97.33 % in 150 iterations.

4.3 Results of Transfer Learning for Plant Disease Identification

For the task of lant disease identification, after training the pre-trained VGG 16 model for 30 epochs, following results are obtained.

Table 4.4: Plant disease identification; Values of performance metrics obtained for different split ratios.

Train-Test Ratio	Testing Accuracy	Testing AUC	Testing Precision	Testing Recall	Testing F1 score
10%-90%	77.99	0.9739	0.8668	0.7506	0.804526
50%-50%	86.66	0.9910	0.9098	0.8329	0.869653
60%-40%	87.82	0.9915	0.9123	0.8526	0.8801
67%-33%	89.72	0.9922	0.9186	0.8782	0.897946
70%-30%	90.15	0.9935	0.9262	0.8820	0.9025
90%-10%	94.05	0.9959	0.9525	0.9308	0.941525

It is evident from the obtained results that as the training set decrease, the accuracy also decreased. So, the accuracy is directly proportional to the training dataset available. We can observe that the 90-10 split yielded a maximum accuracy of 94.05%. It was as expected as 90% of training data was fed to train the model and only 10% was kept for the testing set. The 10-90 split yielded in the least accuracy of 77.99% . It was as expected as only 10% of the data was fed for training and 90% was used for testing. Similar trends can be observed for precision, recall and F1 score. As evident from the training loss and accuracy graphs, there were no signs of overfitting and the graphs were smooth.

4.4 Results of One-Shot Learning on Fruit 360 Dataset

To evaluate our model for a classification task, we used the methodology of N-Way One shot learning. In our methodology, we implemented N-Way one shot learning for N=5,10,15,20,25,30,35. Also, for each of the mentioned N, the process was repeated for 100,150,200 iterations. This process was carried out for each of the 3 datasets; color, segmented and grayscale.

Table 4.5: Plant Disease Identification; Variation of accuracy (%) with number of pairs (for 100 iterations)

	N=3	**N=6**	**N=9**	**N=12**	**N=15**	**N=18**	**N=21**
Color	90.5%	88.33%	85%	83.5%	81%	78.5%	77.5%
Grayscale	88%	87.5%	84%	81%	77.5%	74.66%	71%
Segmented	93%	90%	91%	86.5%	81%	78%	76%

Table 4.5 shows the testing accuracy for different values of N. These are iterated for 100 iterations. As, the dataset was divided into three sections, namely color, grayscale and segmented, the table 4.5 shows the results for all the three diversities of datasets available. There is a decrease in

accuracy with the increased value of N. The highest accuracy that can be observed is 90.5% for colored images, 88% for grayscale images and 93% for the segmented images.

Table 4.6: Plant Disease Identification; Variation of accuracy (%) with number of pairs (for 150 iterations)

	N=3	**N=6**	**N=9**	**N=12**	**N=15**	**N=18**	**N=21**
Color	90.5%	86.5%	86.5%	84.66%	80.33%	77%	75.5%
Grayscale	88.5%	86.66%	83.33%	81%	77.66%	75.5%	73%
Segmented	92.33%	89.33%	87.66%	83%	83%	80.5%	78.6%

Table 4.6 shows the testing accuracy for different values of N. These are iterated for a total of 150 iterations. Table 4.6 shows the results for all the three variations of the datasets, namely colored, grayscale and segmented. There is a decrease in accuracy with the increased value of N. The highest accuracy that can be observed is 90.5% for colored images, 88.5% for grayscale images and 92.33% for the segmented images.

Table 4.7: Plant Disease Identification; Variation of accuracy (%) with number of pairs (for 200 iterations)

	N=3	**N=6**	**N=9**	**N=12**	**N=15**	**N=18**	**N=21**
Color	90.5%	91%	89%	85.33%	81%	77.5%	76%
Grayscale	87%	85.66%	82.5%	79.66%	77.5%	76%	75.5%
Segmented	92%	89%	85.5%	83%	80%`	77.33%	77.5%

From table 4.7, we can observe the testing accuracy for different values of N for a total of 200 iterations. Table 4.7 shows the results for all the three variations of the datasets, namely colored, grayscale and segmented. A decrease in the testing accuracy can be observed with the increased value of N. The highest accuracy that can be observed is 90.5% for colored images, 87% for grayscale images and 92% for the segmented images.

CONCLUSION

With the advances in machine learning algorithms and techniques, various attempts have been made to address the problem of scarcity of data in the field of agriculture. In this study, one such attempt was made by employing the technique of one shot learning for fruit classification task and plant disease identification. For fruit classification, a mini dataset was prepared and a siamese network was trained on the prepared dataset. Siamese network takes a pair of images as an input and outputs a similarity score. The trained model was evaluated using N-Way one shot learning. N was varied from 5 to 40 with a step size of 5. Each time, the process was repeated for 100, 150 and 200 iterations. There was a gradual decrease in accuracy with increase in N. However, even for 40- Way classification, the model was able to generate accuracy of 80%. For the task of plant disease identification, as there were 3 types of available datasets; colored, segmented and grayscale, hence 3 mini datasets were prepared for each of the image type available. Siamese networks were trained on each of the dataset. Further N-Way one shot learning was performed for each of the trained networks, and N was varied from 3 to 21 with a step size of 3. Each of it was repeated for 100,150 and 200 iterations. For 21 Way classification, dataset prepared from segmented images was able to generate accuracy of 77.5%.

REFERENCES

1. Too, Edna Chebet, et al. "A comparative study of fine-tuning deep learning models for plant disease identification." *Computers and Electronics in Agriculture* 161 (2019): 272-279.
2. Seng, Woo Chaw, and Seyed Hadi Mirisaee. "A new method for fruits recognition system." *2009 International conference on electrical engineering and informatics*. Vol. 1. IEEE, 2009.
3. Huang, Ziliang, Yan Cao, and Tianbao Wang. "Transfer learning with efficient convolutional neural networks for fruit recognition." *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. IEEE, 2019.
4. Zawbaa, Hossam M., et al. "Automatic fruit image recognition system based on shape and color features." *International Conference on Advanced Machine Learning Technologies and Applications*. Springer, Cham, 2014.
5. Khan, Rafflesia, and Rameswar Debnath. "Multi class fruit classification using efficient object detection and recognition techniques." *International Journal of Image, Graphics and Signal Processing* 11.8 (2019): 1.
6. Hou, Lei, et al. "Fruit recognition based on convolution neural network." *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 2016.
7. Fei-Fei, Li, Rob Fergus, and Pietro Perona. "One-shot learning of object categories." *IEEE transactions on pattern analysis and machine intelligence* 28.4 (2006): 594-611.
8. Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." *ICML deep learning workshop*. Vol. 2. 2015.
9. Lake, Brenden, et al. "One shot learning of simple visual concepts." *Proceedings of the annual meeting of the cognitive science society*. Vol. 33. No. 33. 2011.
10. Mureşan, Horea, and Mihai Oltean. "Fruit recognition from images using deep learning." *arXiv preprint arXiv:1712.00580* (2017).
11. Duong, Linh T., et al. "Automated fruit recognition using EfficientNet and MixNet." *Computers and Electronics in Agriculture* 171 (2020): 105326.
12. Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." *Frontiers in plant science* 7 (2016): 1419.