

Thesis Proposal for the Degree of Master of Computer Information System

**MEDICINAL PLANT CLASSIFICATION THROUGH
LEAVES BASED ON CONVOLUTIONAL NEURAL
NETWORK USING ResNet50**



Rupesh Kuinkel

(2019-1-46-0023)

**Faculty of Management
Pokhara University, Nepal**

December 2022

Abstract

Medicinal plants (herbs) are plants that are known to have certain compounds which are nutritious for health. The human body is complex and organic. Therefore a system is needed to be able to help the community to recognize medicinal plants better.

In this study, identification of medicinal plant leaves was carried out using the Convolutional Neural Network method. This research will build a system of identification of medicinal plant leaves by using Convolutional Neural Networks. Automatic plant species classification has always been a great challenge. Classical machine learning methods have been used to classify leaves using handcrafted features from the morphology of plant leaves which has given promising results. However, we focus on using non-handcrafted features of plant leaves for classification. So, to achieve it, we utilize a deep learning approach for feature extraction and classification of features. Recently Deep Convolution Neural Networks have shown remarkable results in image classification and object detection-based problems. With the help of the transfer learning approach, we explore and analyse a pre-trained network i.e. "ResNet50" loaded with ImageNet weights.

The model is trained on the plant leaf image data set, consisting of leaf images from six different unique plant species. Automatic plant species classification could be helpful for food engineers, people related to agriculture, researchers, and other plant enthusiasts people.

Keywords: Plant Leaf Recognition, Image of medicinal plant leaves Deep Learning, Transfer Learning, Convolutional Neural Network, ResNet50, Medicine

Table of Contents

Abstract	i
Table of Contents	ii
List of Figures	iii
List of Abbreviations	iii
1. Introduction	1
1.1 Background	2
1.1.1 Convolutional Neural Network (CNN)	2
• Pooling Layer	4
• Fully connected Layer	4
• Normalization Layer	5
• Activation Functions	5
1.1.2 Pre-trained Data Model	6
1.1.2.1 Very Deep CN for Large-Scale Image Recognition (VGG-16)	6
1.1.2.2 ResNet50	6
1.1.2.3 Inception	7
1.1.2.3 Efficient Net	7
1.2 Problem Statement	8
1.3 Research Objectives	8
2. Literature Review	9
3. Methodology	10
3.1 Data Collection	11
3.2 Data Preprocessing	11
3.3 Model Training and Fine Tuning	11
3.4 Classification and Testing	12
3.5 Evaluation	12
3.5.1 Confusion Matrix	12
Accuracy:	13
Precision:	13
Recall:	13
4. Expected Output	14
5. Tentative Time Schedule	15
References	16

List of Figures

Figure 1: Typical structure of CNN.....	5
Figure 2: Illustration of Convolutional operation.....	6
Figure 3: Illustration of Pooling operation.....	7
Figure 4: Architecture of pre-trained ResNet50 trained in ImageNet Dataset.....	9
Figure 5: Steps involved in methodology.....	13
Figure 6: Feature extraction and fine-tuning process.....	14

List of Abbreviations

CNN	Convolutional Neural Network
RoI	Region of Interest
YOLO	You Only Look Once
SSD	Single Stage Object Detector
R-CNN	Region-based Convolutional Neural Network
FPN	Feature Pyramid Network
FC	Fully Connected Layer
ReLU	Rectified Linear Unit

1. Introduction

Image classification is a big problem in computer vision for the decades. In case of humans the image understanding and classification is done easily but in case of computers it is very expensive task. In general, each image is composed of set of pixels and each pixel is represented with different values. Henceforth to store an image the computer must need more spaces for store data. To classify images, it must perform higher number of calculations. For this it requires systems with higher configuration and more computing power.

In recent times, the industrial revolution makes use of computer vision for their work. Automation industries, robotics, medical field, and surveillance sectors make extensive use of deep learning [1]. Deep learning has become the most talked-about technology owing to its results which are mainly acquired in applications involving language processing, object detection and image classification. Nowadays, Deep learning algorithms are providing successful results in the areas like computer vision. The CNN, a machine learning algorithm is being used for the image classification. CNN is a type of feed-forward artificial neural network that has been successfully applied to analyze visual images. It is inspired by the biological processes and the neurons are connected as in animal visual cortex. CNN is supervised deep learning approach which requires large data for training on the network. After training the model will learn the weights and the accuracy of the classifier is improved [2].

Image classification and detection are the most important pillars of object detection. Object detection has been witnessing a rapid revolutionary change in the field of computer vision. Its involvement in the combination of object classification as well as object localization makes it one of the most challenging topics in the domain of computer vision.

In simple words, the goal of this detection technique is to determine where objects are located in a given image called object localization and which category each image belongs to, which is called object classification.

1.1 Background

1.1.1 Convolutional Neural Network (CNN)

A CNN consists of an input layer, output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers (ReLU). Additional layers can be used for more complex models. Example of a typical CNN is depicted in figure below [6]:

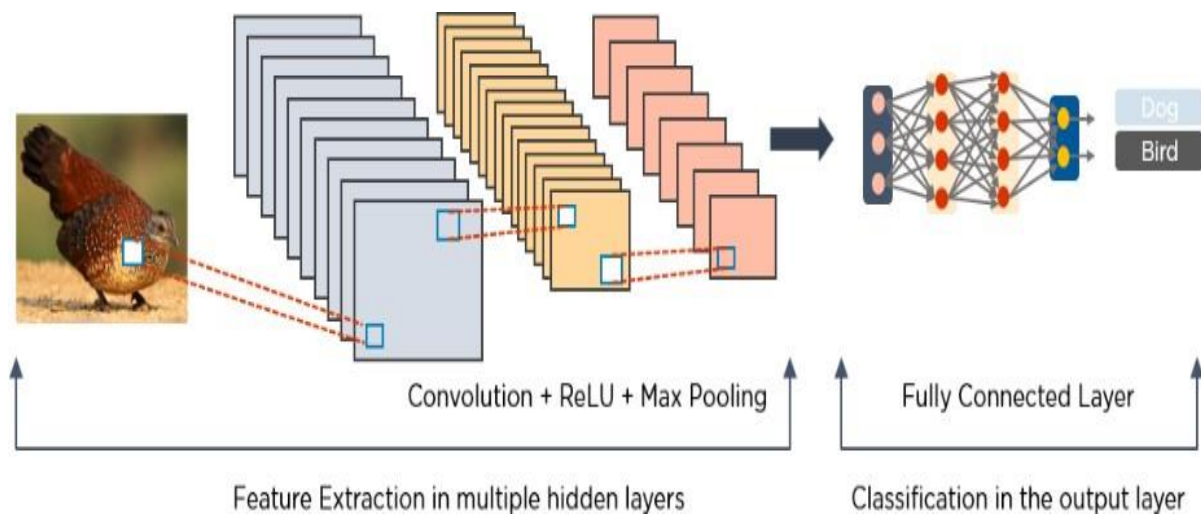


Figure 1:- Typical structure of CNN[6]

- Convolutional Layer

This layer is the first layer of CNN which is used to extract the various features from the input images. The mathematical operation of convolution is performed between the input image and a filter of a particular size $M \times M$. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter ($M \times M$). The output is termed as the Feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image. The convolution layer in CNN passes the result to the next layer once applying the convolution operation in the input. Convolutional layers in CNN benefit a lot as they ensure the spatial relationship between the pixels is intact [7]. Suppose the input image is 3×4 and the convolution kernel size is 2×2 , as illustrated in given fig;

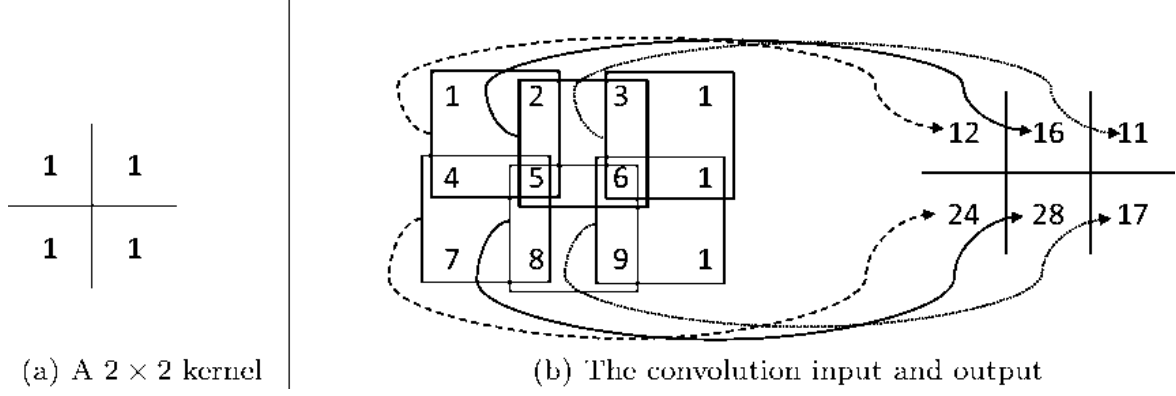


Figure 2: Illustration of Convolutional operation [8]

If we overlap the convolution kernel on top of the input image, we can compute the product between the numbers at the same location in the kernel and the input, and we get a single number by summing these products together. For example, if we overlap the kernel with the top left region in the input, the convolution result at that spatial location is: $1 \times 1 + 1 \times 4 + 1 \times 2 + 1 \times 5 = 12$. We then move the kernel down by one pixel and get the next convolution result as $1 \times 4 + 1 \times 7 + 1 \times 5 + 1 \times 8 = 24$. We keep move the kernel down till it reaches the bottom border of the input matrix (image). Then, we return the kernel to the top, and move the kernel to its right by one element (pixel). We repeat the convolution for every possible pixel location until we have moved the kernel to the bottom right corner of the input image, as shown in Figure 3.[8]

In precise mathematics, the convolution procedure can be expressed as an equation [8]:

$$y_{i^{l+1}, j^{l+1}, d} = \sum_{i=0}^H \sum_{j=0}^W \sum_{d^l=0}^{D^l} f_{i,j,d^l} \times x_{i^{l+1}+i, j^{l+1}+j, d^l}$$

equation is repeated for all $0 \leq d \leq D = D^{l+1}$, and for any spatial location (i^{l+1}, j^{l+1}) satisfying $0 \leq i^{l+1} < H^l - H + 1 = H^{l+1}$, $0 \leq j^{l+1} < W^l - W + 1 = W^{l+1}$. In this equation, $x_{i^{l+1}+i, j^{l+1}+j, d^l}$ refers to the element of x^l indexed by the triplet $(i^{l+1} + i, j^{l+1} + j, d^l)$

- Pooling Layer

In CNN, Pooling layer is used to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of Pooling operations. It basically summarizes the features generated by a convolution layer. In Max Pooling, the largest element is taken from feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer. This CNN model generalizes the features extracted by the convolution layer, and helps the networks to recognize the features independently. With the help of this, the computations are also reduced in a network [7].

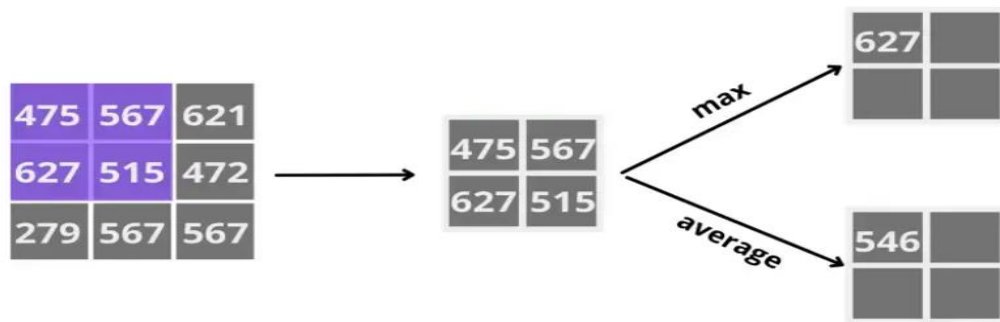


Figure 3: Illustration of Pooling operation

- Fully connected Layer

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture. In this, the input image from the previous layers are flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place. The reason two layers are connected is that two fully connected layers will perform better than a single connected layer. These layers in CNN reduce the human supervision [7].

- Normalization Layer

This layer is also known as the process of ReLU which involves changing all negative values within the filtered image to 0. A units using sigmoid or tanh functions can suffer from the vanishing gradient problem which cause slow optimization convergence, and in some cases the final trained network may fall off to a poor local minimum station. ReLU doesn't face gradient vanishing problem as with sigmoid and tanh function. The purpose of ReLU is to increase the nonlinearity of the CNN.

- Activation Functions

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network. It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions. Each of these functions have a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred and for a multi-class classification, generally softmax us used. In simple terms, activation functions in a CNN model determine whether a neuron should be activated or not. It decides whether the input to the work is important or not to predict using mathematical operations.[7]

1.1.2 Pre-trained Data Model

Pre-Trained Models for Image Classification [9]

- VGG-16
- ResNet50
- Inceptionv3
- Efficient Net

1.1.2.1 Very Deep CN for Large-Scale Image Recognition (VGG-16)

The VGG-16 is one of the most popular pre-trained models for image classification. Introduced in the famous ILSVRC 2014 Conference, it was Developed at the Visual GraphicsGroup at the University of Oxford, VGG-16 beat the standard of AlexNet and was quickly adopted by researchers and the industry for their image Classification Tasks[9].

1.1.2.2 ResNet50

Just like Inceptionv3, ResNet50 is not the first model coming from the ResNet family. The original model was called the Residual net or ResNet and was another milestone in the CV domain back in 2015.

The main motivation behind this model was to avoid poor accuracy as the model went on to become deeper.

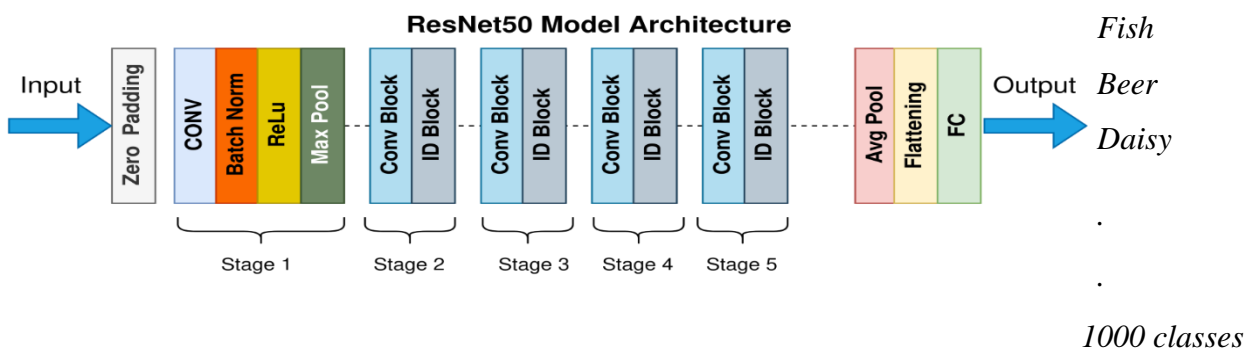


Figure 4: Architecture of pre-trained ResNet50 trained in ImageNet Dataset

1.1.2.3 Inception

Inception Module just performs convolutions with different filter sizes on the input, performs Max Pooling, and concatenates the result for the next Inception module. The introduction of the 1×1 convolution operation reduces the parameters drastically.

1.1.2.3 Efficient Net

It is the latest model used by Google. In Efficient Net, the authors propose a new Scaling method called Compound Scaling. The long and short of it is this: The earlier models like ResNet follow the conventional approach of scaling the dimensions arbitrarily and by adding up more and more layers.

ResNet follow the conventional approach of scaling the dimensions arbitrarily and by adding up more and more layers.

1.2 Problem Statement

Nepal is ranked as 9th among the Asian countries for its floral wealth with an estimated 9,000 species of flowering plants [3]. Medicinal herbs were the main ingredients of traditional therapies, and they were considered a main lifeline and frequently were the first choice. Drugs like codeine, quinine, and morphine all contain plant-derived ingredients. Nepal is rich in plant diversity. So, a large number of plant species having useful chemical constituents are found in Nepal, which has been very useful to cure the diseases in various places of the country. But Globalization of herbal medicine along with uncontrolled exploitative practices and lack of concerted conservation efforts, have pushed many of Nepal's medicinal plants to the verge of extinction.

In fact, automatic plant species classification has always been a great challenge. Classical machine learning methods have been used to classify leaves using handcrafted features from the morphology of plant leaves which has given promising results. However, we focus on using non-handcrafted features of plant leaves for classification. So, to achieve it, we utilize a deep learning approach for feature extraction and classification of features. Deep Convolution Neural Networks have shown remarkable results in image classification. ResNet-50 is one of the popular pre-train models and specifically used for the classification of plant leaves. Hence, we have used this model to train the image datasets for the classification of medicinal plants present in the dataset.

1.3 Research Objectives

The general purpose of object detection is to identify and locate one or more effective targets from still image. It comprehensively includes a variety of important techniques, such as image processing, pattern recognition, artificial intelligence and machine learning[5]. The main objectives of this thesis are

- Classification of the medicinal plants through leaves found in mountain region of Nepal using CNN based on ResNet50
- To analyze the performance of ResNet50 , for image classification that can be useful for plant researchers, pharmaceutical field, herbal medicines manufacturing and other related sectors.

2. Literature Review

In [10], Mohanty, S. P., Hughes, D. P., & Salathé, M. uses the deep convolutional neural network architecture and trained a model on images on plant leaves with the goal of classifying both crop species and the presence and identity of disease on images. They compared performance of the two CNN architectures i.e. AlexNet and GoogleNet on PlantVillage dataset of leaf diseases. The performance measures considered were precision, F1 score, recall and accuracy of the model. They have done implementation on three scenario i.e. color, grayscale and segmented images for measuring the CNN performance where they found that GoogleNet outperformed AlexNet.

In [11], S. Raina and A. Gupta have studied various techniques for plant leaf disease detection using leaf image. In this work, they presented the basis of plant disease detection techniques used by various researchers. Experimental analysis has been demonstrated with different models of CNN like multi layered convolutional neural network.

In [12], Ganatra and Patel presented Imagebased plant disease classification by fine-tuning different convolutional neural networks. Different frameworks have been tested and contrasted, including VGG16, Inception V4, ResNet 50 and ResNet101. The evaluation precision of ResNet50 and ResNet101 was 99.70 percent and 99.73 percent respectively. In this work, Overall ResNet 101 was able to achieve highest accuracy among other models while offers lowest log loss. Whereas VGG 16 offered low accuracy compared with other deep convolutional architectures.

In [13], Thushara and Rasool provided one of a kind system of applying deep learning strategy to quickly recognize and analyze plant infections from leaves pictures. The Authors' proposed strategy had distinguished between healthy and four different infected leaves. Dataset consisted real time captured images (infected and healthy leaf) and the 96% of accuracy is achieved.

3. Methodology

This research will be implemented in three phases data collection, model implementation and performance evaluation. In first phase, data will be collected and processed using pre-trained models ie ResNet50. In second phase, feature extraction and fine tuning will be done for medicinal plant to train the dataset. In the third phase, the trained model will be evaluated on our real time data for analyzing performance of pre-trained model.

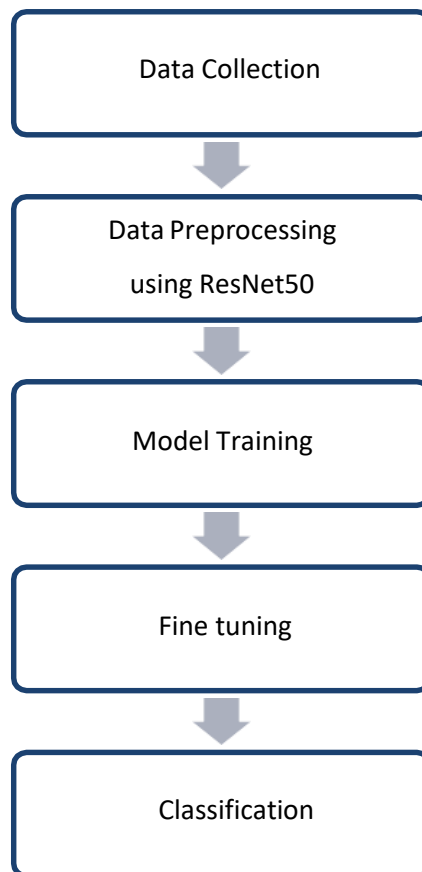


Figure 5: Steps involved in Image Classification

3.1 Data Collection

Leaf images of required plants to train were collected from different standard open data repositories like ‘kagle’, ‘mendely’[14] and few images were also taken through mobile phone. Various sample image of medicinal plants of six classes found in mountain region of Nepal were taken for study i.e; Tulsi "*Ocimum tenuiflorum*", Paan "*Piper betle*", Aloe Vera "*Aloe barbadensis miller*", Black Til "*Sesamum radiatum*", Baramasi Flower "*Catharanthus roseus*", Pudina "*Mentha spicata*". Images will be taken with different orientation, perspective, and different lighting conditions for variations.

3.2 Data Preprocessing

The data (images) collected will be preprocessed using the pretrained models i.e. ResNet50. Following steps will also be applied to improve the image data(features) by suppressing unwanted distortions data preprocessing:-

- Reading the image
- Resizing image
- Remove Noise(denoise)
- Morphology (smoothing edges)

3.3 Model Training and Fine Tuning

After preprocessing the images using Pre-trained Model(ResNet50), the images will pass through this phase. Here, the feature extraction of the leaves of medicinal plants will take place. Eventually, after this fine tuning of the images will take place.

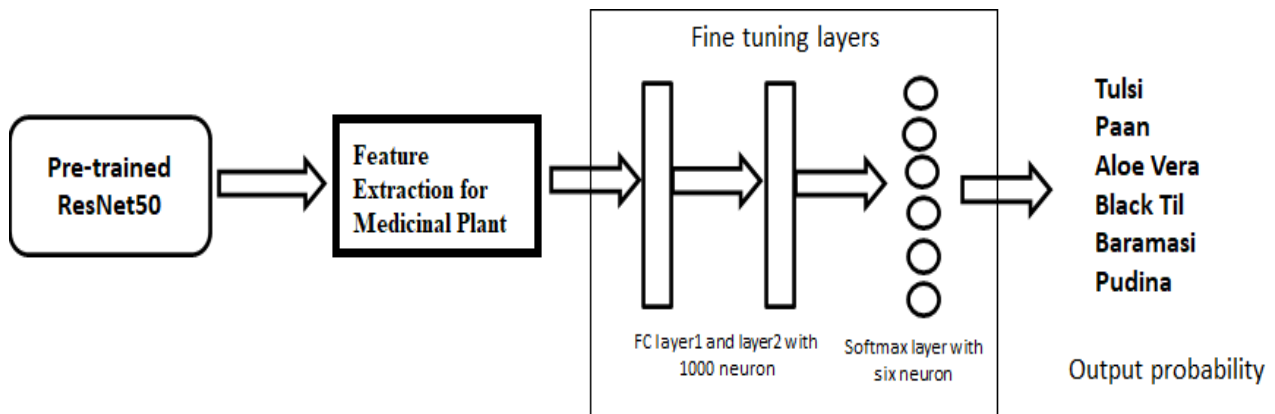


Figure 6: Feature Extraction and Fine-Tuning Process

3.4 Classification and Testing

After the implementation of the algorithms and selection of proper data set, the images will go for classification and will be tested using PYTHON.

3.5 Evaluation

The algorithm implemented will be evaluated using different scientifically proven metrics such as: accuracy, precision, recall and F1 Score.

3.5.1 Confusion Matrix

The Pre-trained networks will be analyzed and compared using three primary metrics. They are precision, recall, and f1-score. These metrics will be calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) obtained from the confusion matrix produced by each pre-trained network. A confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with predicted by the machine learning model.

		Actual Values	
		Positive (1)	Negative(0)
Predicted Values	Positive (1)	TP	FP
	Negative(0)	FN	TN

Where;

TP(True Positive) → actual class of data is True and the predicted is also True.

FP (False Positive) → actual class of data is False and the predicted is also False.

FN (False Negative) → actual class of data is False and the predicted is True.

TN (True Negative) → actual class of data is True and the predicted is False.

Accuracy:

Accuracy is the number of correct prediction divided by the total number of prediction. It gives better result when the target classes is the data are nearly balanced.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision:

Precision is the ratio of correct positive predictions to the overall number of positive predictions.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall:

Recall gives the measure of proportion of the true positives that were correctly identified by the model. It is the ratio of correct positive predictions to the overall number of positive examples in test set.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 Score:

It serves as a harmonic mean of precision and recall i.e.,

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. Expected Output

The result that obtained during this study among the proposed model is expected to classify the proposed classes of each medicinal plants through pre-trained models. Similarly, the evaluation comprising of accuracy, precision, recall and F1-score will be carried out for performance measure.

5. Tentative Time Schedule

The Gantt chart below depicts the working schedule:

Table 1 Tentative Time Schedule

ID	Task name	Dec 2022		Jan 2023		Feb 2023		March 2023	
		1w	1w	1w	1w	1w	1w	1w	1w
1	Literature Review								
2	Familiarization of tools								
3	Algorithm Study								
4	Mid Term								
5	Data collection & testing								
6	Output Analysis								
7	Documentation								

References

- [1] Pathak AR, Pandey M, Rautaray S., "Application of deep learning for object detection," *Procedia Comput Sci.* 2018;132:1706–17.
- [2] M. Ramprasath, M. Anand and S. Hariharan, "Image Classification using Convolutional Neural Networks," *International Journal of Pure and Applied Mathematics*, vol. 119, 2018.
- [3] Bhattarai KR, Maren IE, Chaudhary RP. (2011) Medicinal plant knowledge of the Panchase region in the middle hills of the Nepalese Himalaya. *Banko Janakari*.
- [4] Deering, S. (2019, February 28). Nature's 9 Most Powerful Medicinal Plants and the Science Behind Them. *Healthline*. Retrieved September 22, 2022, from <https://www.healthline.com/health/most-powerful-medicinal-plants#evening-primrose-oil>
- [5] Cao, D., Chen, Z. & Gao, L. An improved object detection algorithm based on multi- scaled and deformable convolutional neural networks. *Hum. Cent. Comput. Inf. Sci.* 10, 14 (2020).
- [6] M. Hussain, J. J. Bird and D. R. Faria, " A study on CNN Transfer Learning for Image Classification," *School of Engineering and Applied Science Aston University, UK*.
- [7] Basic CNN architecture: Explaining 5 layers of Convolutional Neural Network. *upGrad blog*. (2022, October 3). Retrieved November 4, 2022, from <https://www.upgrad.com/blog/basic-cnn-architecture/>
- [8] Wu J. Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*. 2017 May 1;5(23):495.
- [9] Huilgol, P. Top 4 pre-trained models for image classification: With python code. *Analytics Vidhya*. Retrieved November 4, 2022, from <https://www.analyticsvidhya.com/blog/2020/08/top-4-pre-trained-models-for-image-classification-with-python-code/> (2022, June 15).

- [10] Mohanty, S. P., Hughes, D. P., & Salathé, M., "Using deep learning for image-based plant disease detection." *Frontiers in plant science*, 7, 1419(2016).
- [11] S. Raina and A. Gupta, "A Study on Various Techniques for Plant Leaf Disease Detection Using Leaf Image," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 900-905, doi: 10.1109/ICAIS50930.2021.9396023.
- [12] Ganatra N, Patel A. Performance analysis of fine-tuned convolutional neural network models for plant disease classification. *International Journal of Control and Automation*. 2020 Jan;13(3):293-305.
- [13] Thushara BL, Rasool TM. Analysis of plant diseases using expectation maximization detection with BP-ANN classification. *XIII (VIII)*. 2020 Aug.
- [14] Arun Pandian, J.; Geetharamani, G. Data for: Identification of Plant Leaf Diseases Using a 9-Layer Deep Convolutional Neural Network. *Mendeley Data*. 2019. Available online: <https://data.mendeley.com/datasets/tywbtsjrjv/1> (accessed on 29March 2020).