**Project Report on**



**Identification Of Ayurvedic Leaves And Their Benefits Using CNN**

Submitted in partial fulfillment of the course of

**PG- Diploma in Big Data Analytics (PG-DBDA)**

from

**C-DAC ACTS (Bangalore)**

# Guided by: Mr.Abhay Mane

Presented by:

Mr. Abhishek Prataprao Solanke PRN: 210950125067

Mr. Aayush Sunil Dani PRN : 210950125002

Mr. Abhinandan Bhardwaj PRN : 210950125003

Mr. Sujit Sunil Pulujkar PRN : 210950125069

Mr. Shubham Bhimrao Hole PRN : 210950125065

**Centre of Development of Advanced Computing (C-DAC), Bangalore**



CERTIFICATE

TO WHOMSOEVER IT MAY CONCERN

### This is to certify that

### Mr. Abhishek Prataprao Solanke

**Mr. Aayush Sunil Dani**

**Mr. Abhinandan Bhardwaj**

### Mr. Sujit Sunil Pulujkar

### Mr. Shubham Bhimrao Hole

have successfully completed their project on

# Identification Of Ayurvedic Leaves And Their Benefits Using CNN

# under the guidance of

# Mr. Abhay Mane

Mr. Abhay Mane Ms.Uma Prasad

**(Project Guide) (Course Co-ordinator)**



**ACKNOWLEDGEMENT**

This project “**Identification Of Ayurvedic Leaves And Their Benefits Using CNN”** was a great learning experience for us and we are submitting this work to Advanced Computing Training School (CDAC ACTS).

We all are very glad to mention the name of **Mr.Abhay Mane , Project Engineer, (C-DAC))**for his valuable guidance to work on this project. His guidance and support helped us to overcome various obstacles and intricacies during the course of project work.

We are highly grateful to **Mr. Ramesh Naidu** (ACTS training Centre, C-DAC), for his guidance and support whenever necessary while doing this course Post Graduate Diploma in **Big Data Analytics(PG-- DBDA)**through C-DAC ACTS, Bangalore.

Our most heartfelt thanks goes to **Ms. Uma Prasad**(Course Coordinator, **PG-DBDA**) who gave all the required support and kind coordination to provide all the necessities to complete the project and throughout the course up to the last day here in C-DAC ACTS, Bangalore.

**From:**

### Mr. Abhishek Prataprao Solanke

**Mr. Aayush Sunil Dani**

**Mr. Abhinandan Bhardwaj**

### Mr. Sujit Sunil Pulujkar

### Mr. Shubham Bhimrao Hole

**TABLE OF CONTENT**



1. **Abstract**
2. **Introduction And Overview Of Project**

**2.1 Objective Of The Project**

**2.2 Models Used For Prediction**

**2.3 Design / Working Of Mode**

**2.4 Neural Network**

**2.5 Convolutional Neural Network**

* 1. **Hyper Parameter Tuning**

1. **Evaluation Metrics**
2. **Data Preprocessing**
   1. **Exploratory Data Analysis**
      1. **Various Libraries Used For Our Model**
      2. **Loading Dataset**
      3. **Creating The Dataframe**
      4. **Details Of DataSet**
   2. **EDA**
   3. **Data Preprocessing**
3. **Model Building**

**5.1 Hyper Parameter Tuning Of Model**

**5.2 Convolution Neural Network (CNN)**

**5.2.1 Visualizing The Model**

**5.2.2 Model Summary**

**5.2.3 Compile/ Optimizing The Model**

**5.2.4 Training The Model**

**5.2.5 Training Model Visualization**

**5.2.6 Saving The Model And Evaluating /Testing The Model**

**5.2.7 Classification Report**

1. **Predicting The Scientific Name And Its Benefits**

**6.1** **User Interface**

**6.2 Predicting The Output**

1. **Conclusion**
2. **Future Scope**
3. **Bibliography**



# Abstract

.

Ayurvedic leaves are herbs that are known to have certain compounds which are nutritious for health. In India there are 30,000 types of plants and 7000 of them are classified as medicinal plants (herbs). The human body is complex and organic, while chemical medicines contain chemicals that are inorganic and pure. Therefore, chemical medicines are considered not very suitable for consumption by the human body, which if consumed continuously can even be bad for human health. However, some chemical drugs are actually symptomatic (temporary) so they must be taken for life by patients with certain diseases. Therefore a system is needed to be able to help the community to recognize medicinal plants better, in this case the medicinal plants are focused on the introduction of medicinal leaves. In this study identification of medicinal plant leaves is carried out using the Convolutional Neural Network method and benefits are displayed on StreamLit UI. This project will build a system of identification of medicinal plant leaves by using Convolutional Neural Networks. Using training data that is carried out in a computer set and then implemented in DL-based model , for displaying the scientific name and benefits of medicinal plant leaves we used StreamLit UI.

# Introduction and Overview of Project



# In this project, identification of medicinal plant leaves was carried out using the Convolutional Neural Network method. CNN is one of the algorithms from the branch of Machine Learning that is based on Artificial Neural Networks (ANN) or its development, namely Deep Learning which is a development of the Multilayer Perceptron (MPL) to process two-dimensional data, one of them is image. CNN is used in image data to detect and recognize objects in an image, with Back propagation type training. The way CNN works is similar to MLP, but in CNN each neuron propagated on the network has a two-dimensional shape, so that the weight and linear operating parameters on CNN are different. This project will build a system of identifying 32 types of leaves of hypertensive medicinal plants using Convolutional Neural Networks. In order to facilitate the user in using the system, the system built will be implemented on a web based UI using StreamLit library of python to recognize the types and benefits of medicinal plant leaves that are identified.

## Objectives Of The Project:

* + 1. Detecting the Ayurvedic plant leaves in CNN
    2. Identifying scientific name and their benefits in StreamLit UI.

## Models Used For Prediction:

1. Convolutional Neural Network



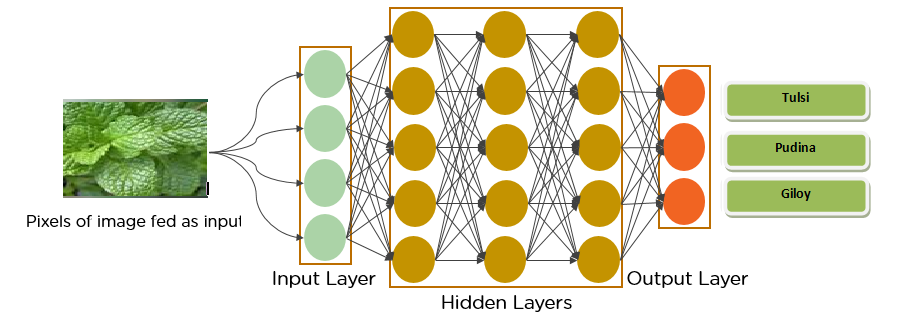
## Design / Working Of Model:



## 

## Neural Networks:

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so the network generates the best possible result without needing to redesign the output



criteria. The concept of neural networks, which has its roots in artificial intelligent, is swiftly gaining popularity in the development of trending system.



Neural networks are multi-layer networks of neurons (the yellow,red and green nodes in the chart above) that we use to classify things, make predictions, etc.

The arrows that connect the dots shows how all the neurons are interconnected and how data travels from the input layer all the way through to the output layer.

### Advantages of Neural Network:

* Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them.
* The input is stored in its own networks instead of a data base, hence the loss of data does not affect its working.
* These networks can learn from examples and apply them when a similar event arises, making them able to work through real-time events.
* Even if a neuron is not responding or a piece of information is missing, the network can detect the fault and still produce the output.
* They can perform multiple tasks in parallel without affecting the system performance.

# Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep mastering algorithm that may soak up an input image, assign significance (learnable weights and biases) to diverse components/objects inside the image and have the ability to distinguish one from the opposite. The pre-processing required in a ConvNet is tons decrease in comparison to different class algorithms. at the same time as in primitive techniques filters are hand-engineered, with enough training, ConvNets have the capacity to research those filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity sample of Neurons in the Human mind and became inspired by using the business enterprise of the visual Cortex. person neurons respond to stimuli simplest in a confined place of the visual field referred to as the Receptive area. a collection of such fields overlap to cover the complete visual vicinity.

# Hyper Parameter Tuning



# Hyper- Parameters:

# Model parameters are learned from data and hyper-parameters are tuned to get the best fit. Searching for the best hyper-parameter can be tedious, hence search algorithms like grid search and random search are used.

# D:\CDAC SOFTWARE\abhinandan\CDAC\Project\LeafProject\Leaf Project Contents\Hypertunning.jpg

* + 1. **Optimal Hyper parameters:**

Hyper-parameters control the over-fitting and under-fitting of the model. Optimal hyper-parameters often differ for different datasets. To get the best hyper-parameters the following steps are followed:

1. For each proposed hyper-parameter setting the model is evaluated.
2. The hyper-parameters that give the best model are selected.
   * 1. **Hyper Parameters Search :**

Grid search picks out a grid of hyper parameter values and evaluates all of them. Guesswork is necessary to specify the min and max values for each hyper-parameter.

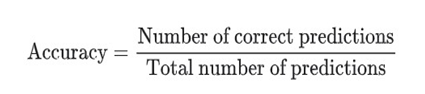
Random search randomly values a random sample of points on the grid. It is more efficient than grid search. In this model we used Random search which predicts precise CNN network with respect to given Datasets and initial hyper parameters.



# Evaluation Metrics

## Accuracy:

Accuracy is one metric for evaluating classification models. Informally, **accuracy** is the fraction of predictions our model got right. Formally, accuracy has the following definition:



For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

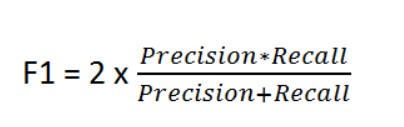
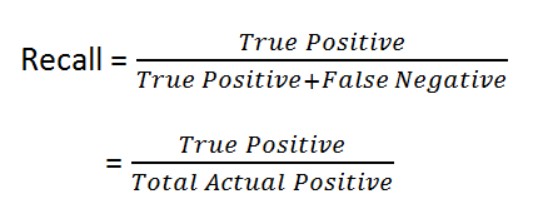
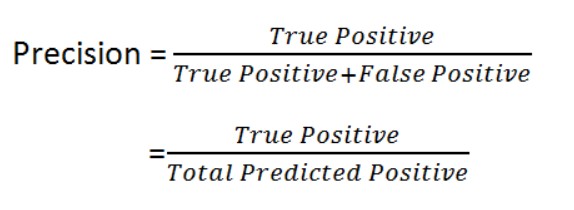
Accuracy=TP+TNTP+TN+FP+FN

Where *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

# Precision And Recall

Precision is a good measure to determine, when the costs of False Positive is high. For instance, email spam detection. In email spam detection, a false positive means that an email that is non-spam (actual negative) has been identified as spam (predicted spam). The email user might lose important emails if the precision is not high for the spam detection model.

In the field of information retrieval precision is the fraction of retrieved documents that are relevant to the query:



### Recall

Recall calculates how many of the Actual Positives our model capture through labeling it as Positive (True Positive). Applying the same understanding, we know that Recall shall be the model metric we use to select our best model when there is a high cost associated with False Negative.

* 1. **F1 Score**

F1 is a function of Precision and Recall.

F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall AND there is an uneven class distribution (large number of Actual Negatives).

## Support

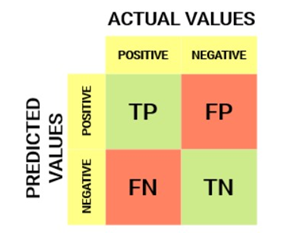
Support is the number of actual occurrences of the class in the specified data set. Imbalanced support in the training data may indicate structural weaknesses in the reported scores of the classifier and could indicate the need for stratified sampling or re-balancing.

## Confusion Matrix



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 those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making.

For a binary classification problem, we would have a 2 x 2 matrix as shown below with 4 values:



Let’s decipher the matrix:

* The target variable has two values: **Positive** or **Negative**
* The **columns** represent the **actual values** of the target variable
* The **rows** represent the **predicted values** of the target variable



# Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

Before going into pre-processing and data exploration we will explain some of the concepts that allowed us to select our features.

### Exploratory Data Analysis :-

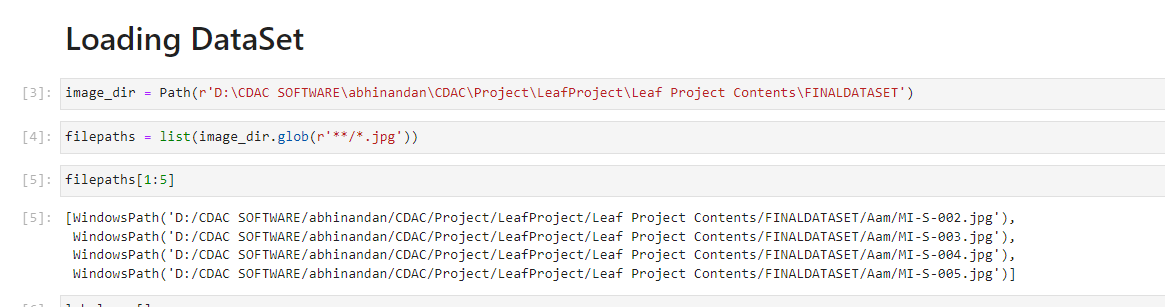
Refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

### Various Libraries Used For Our Model:

### 



* + 1. **Loading Dataset**

****

### Creating The DataFrame

### 

### Details Of DataSet

****

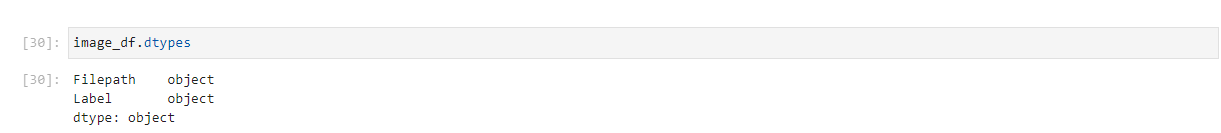
* 1. **EDA**

It is a way of visualizing, summarizing and interpreting the information that is hidden in rows and column format. EDA is one of the crucial step in data science that allows us to achieve certain insights and statistical measure that is essential for the business continuity, stockholders and data scientists. It performs to define and refine our important features.

1. Handle Missing value
2. Removing duplicates
3. Outlier Treatment
4. Normalizing and Scaling( Numerical Variables)
5. Encoding Categorical variables( Dummy Variables)

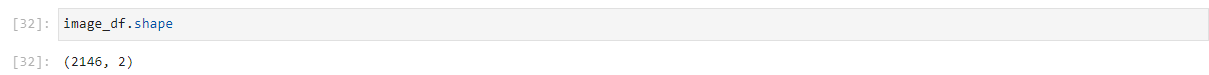
### Variable identification and data types

The very first step in exploratory data analysis is to identify the type of variables in the dataset. Variables are of two types Numerical and Categorical..dtypes method to identify the data type of the variables in the dataset .



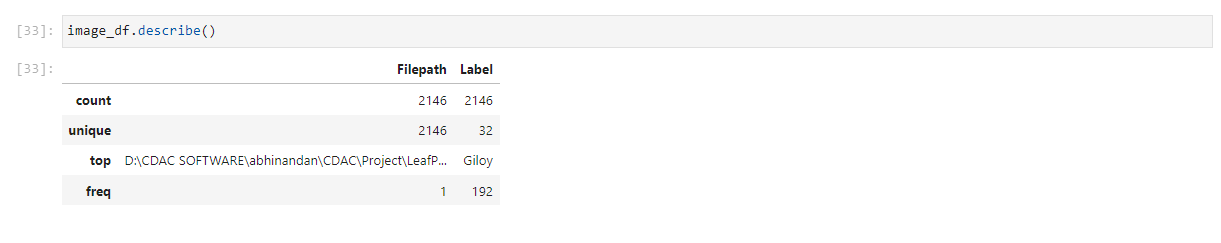
### Size of the dataset

We can get the size of the dataset using the shape method.



### Describe the dataset

Describe() function to get various summary statistics that exclude NaN values.this function returns the count, mean, standard deviation, minimum and maximum values and the quantiles of the data.



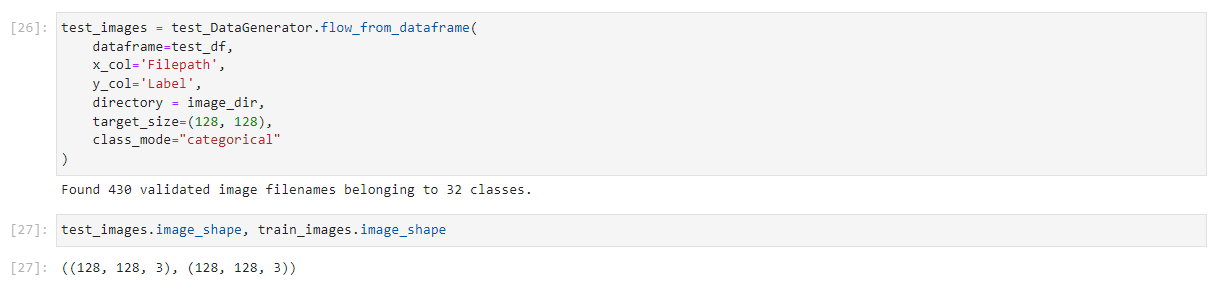
### Data preprocessing

* + 1. **Augmentation**

After execution of this code, the independent variable *X* and dependent variable *Y* will transform into the following.







# 5. Model Building



The modeling process was divided into two main parts: traditional machine learning models and convolution neural networks. We followed hyper parameter tuning, initializing CNN & Convolution layers, Pooling operation, Flattening operation, Fully connected layer and Output layer.

### Hyper Parameter Tuning Of Model

### 

### 

### Summary Of Hyper parameter Tuning

### D:\CDAC SOFTWARE\abhinandan\CDAC\Project\LeafProject\Leaf Project Contents\tuner.PNG

### Best Hyper Parameters

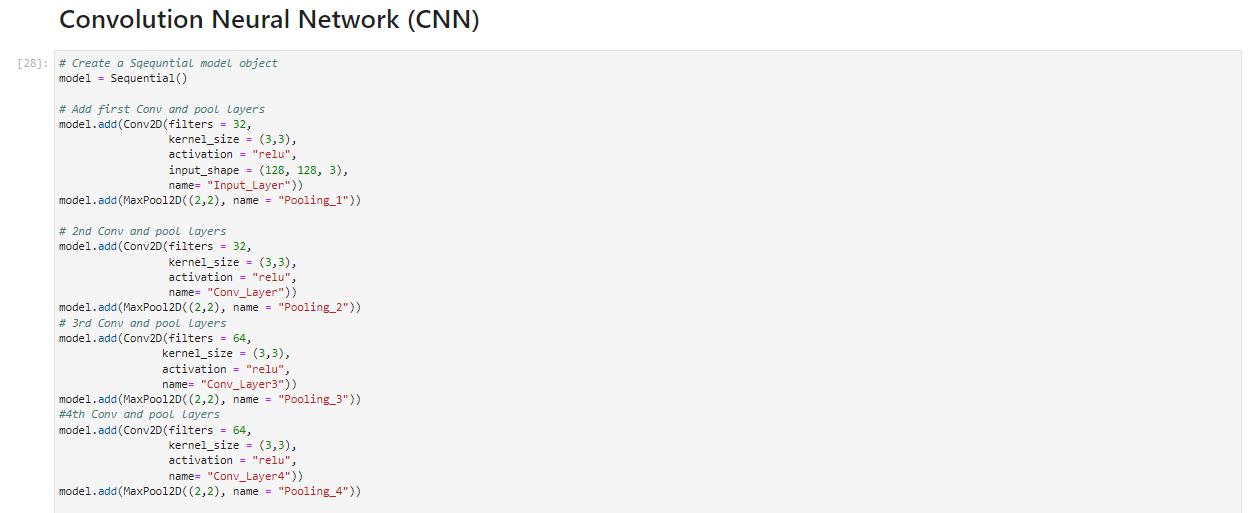


### D:\CDAC SOFTWARE\abhinandan\CDAC\Project\LeafProject\Leaf Project Contents\tuner2.PNG

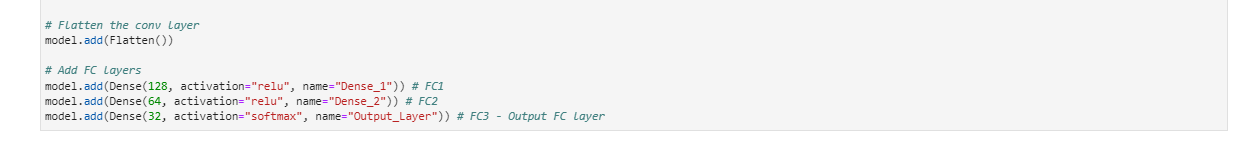
### Convolution Neural Network (CNN)

We will build the Convolution Neural Network

1. Conv2D : This layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs. If use\_bias is True, a bias vector is created and added to the outputs. Finally, if activation is not None, it is applied to the outputs as well.
2. MaxPooling2D: Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.
3. Fully conned and output layer: Fully Connected layers in a neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. In most popular machine learning models, the last few layers are full connected layers which compiles the data extracted by previous layers to form the final output.





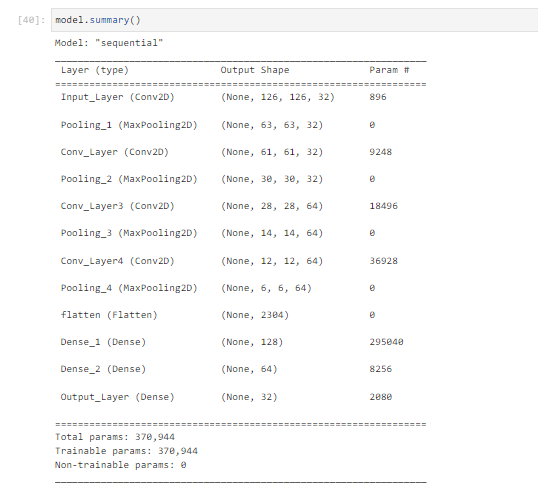


### Visualizing the Model

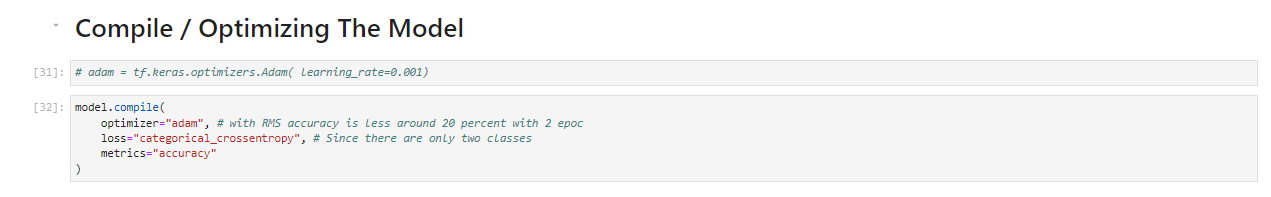
### D:\CDAC SOFTWARE\abhinandan\CDAC\Project\LeafProject\Leaf Project Contents\Training Model Visulization.PNG

* + 1. **Model Summary**

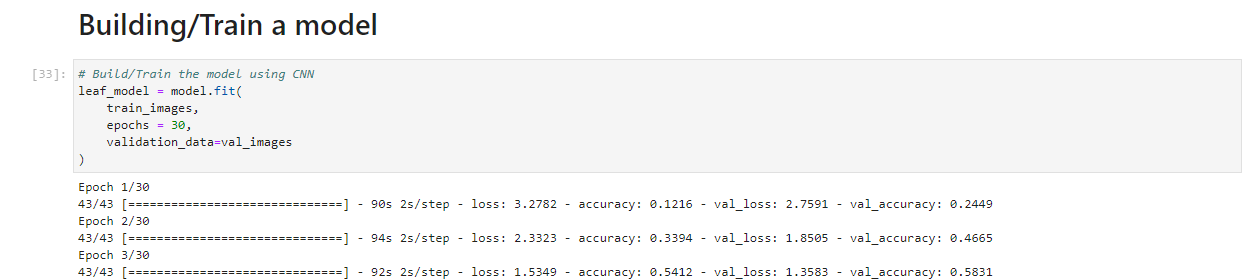


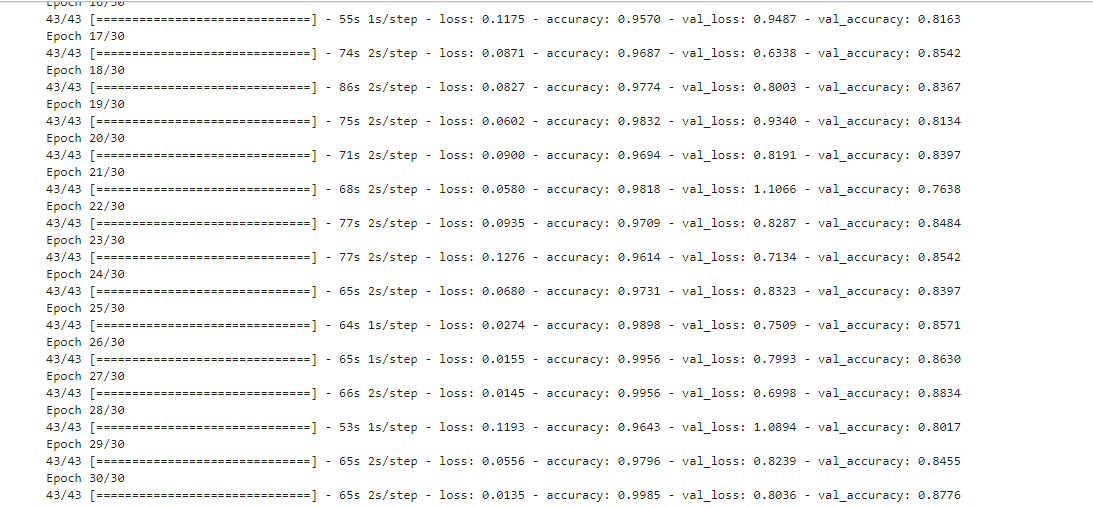
****

* + 1. **Compile /Optimizing The Model**

****

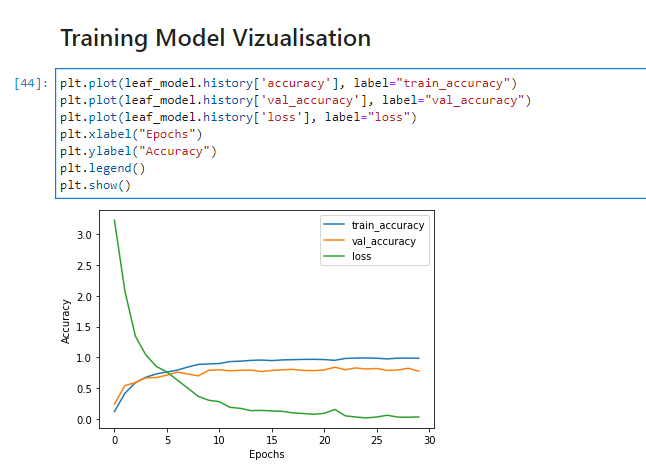
* + 1. **Training The Model**

****

****

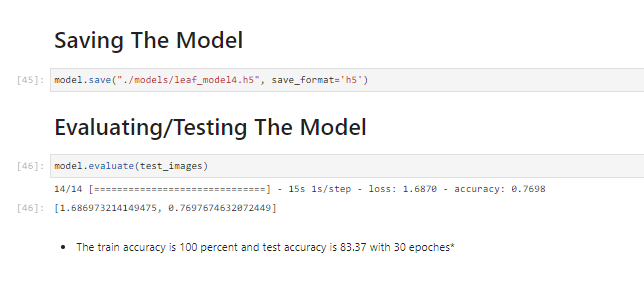


* + 1. **Training Model Visualization**

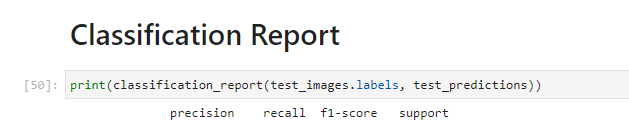
****

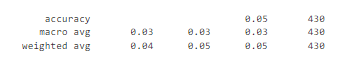
* + 1. **Saving The Model and Evaluating /Testing The Model**



****

* + 1. **Classification Report**

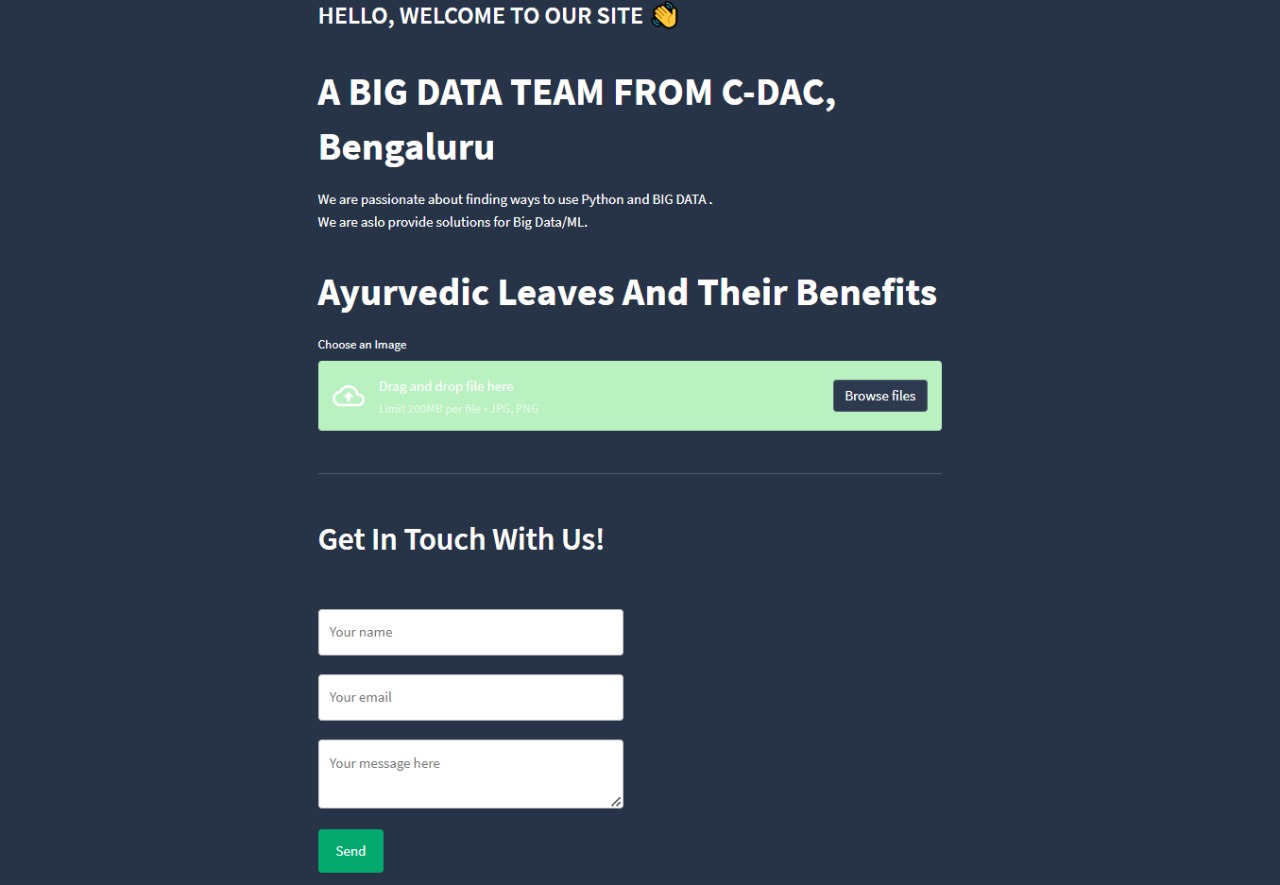
****

****

1. **Predicting The Scientific Name And Its Benefits.**

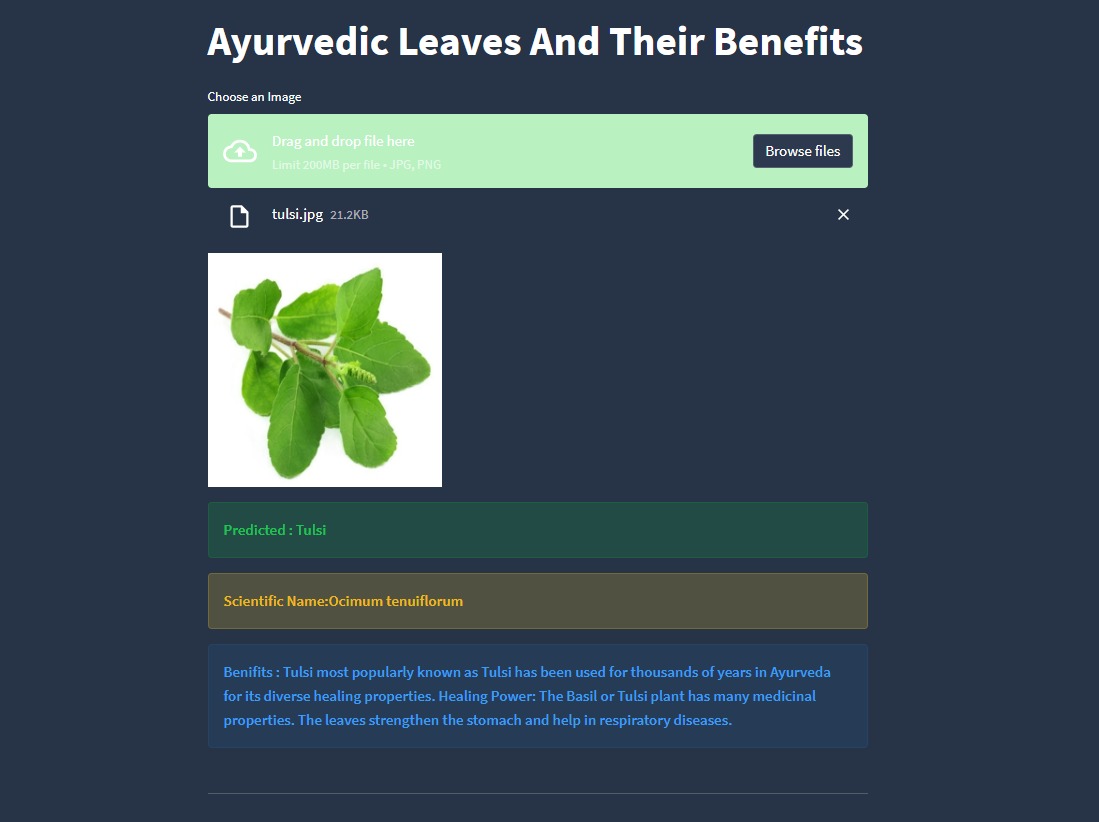


**6.1 User Interface .**

****

**6.2 Predicting The Output.**



****

****

# Conclusion



This project has successfully implemented the Convolutional Neural Network method to extract features on medicinal plant leaves and identify them into 32 classes of medicinal plant leaves based on the closest value between the training data and test data.

This project analysis the Identification of Medicinal Plant Leaves using Convolution Neural Network with Accuracy(87.8%). The accuracy of the model can be increased by including more images of the Medicinal Plant Leaves.

# Future scope



* To differentiate a healthy and unhealthy leaves.
* Farmers can track the growth of their crops by just scanning the picture of leaf.
* People can use the application for the means of identification of medicinal plants.

### Bibliography



1. Dataset
   * <https://data.mendeley.com/datasets/nnytj2v3n5/1>
   * https://www.kaggle.com/datasets/ahilaprem/mepco-tropic-leaf
2. Documentation

* <https://github.com/AbhishekSolanke/newrepo1>

1. Resources
   * <https://pythonprogramming.net/keras-tuner-optimizing-neural-network-tutorial/>
   * <https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/>
   * <https://www.analyticsvidhya.com/blog/2021/05/tuning-the-hyperparameters-and-layers-of-neural-network-deep-learning/>
   * <https://www.tensorflow.org/>
   * <https://streamlit.io/>



1. Algorithm
   * [https://dataaspirant.com/ensemble-methods-bagging-vs-boosting-](https://dataaspirant.com/ensemble-methods-bagging-vs-boosting-difference/) [difference/](https://dataaspirant.com/ensemble-methods-bagging-vs-boosting-difference/)
   * [https://towardsdatascience.com/a-comprehensive-guide-to-convol](https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53) [utional-neural-networks-the-eli5-way-3bd2b1164a53](https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53)
   * [https://towardsdatascience.com/a-complete-guide-to-principal-com](https://towardsdatascience.com/a-complete-guide-to-principal-component-analysis-pca-in-machine-learning-664f34fc3e5a) ponent-analysis-pca-in-machine-learning-664f34fc3e5