International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 7 ǁ July 2022 ǁ PP. 929-941

Crop Plants Leaf Image Classification using Machine Learning Approach

Abstract: *The leaf is the main part of the plant that identifies its range and varieties, but leaf identification is a difficult and complex task due to the nature of the plant, which presents a vast pattern variation. Crop plants directly or indirectly provide food to humans and all other non-photosynthetic organisms, so we should explore more diversity of crop plants to meet the food shortage. But it is a very difficult task for humans to identify the leaves of the crop. Because human is not capable of the naked eye. Identify different types and species of leaves of similar-looking crop plants. Therefore, leaf detection of the crop plant is very essential, and it can be done using image processing techniques. In this research, the author himself collected 21 categories (i.e., 5931 RGB high-resolution images) of leaves of crop plants. A comparative study of the classification performance of classifiers has been done using Machine Learning (ML) models i.e., Logistic Regression (LR), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Naive Bayes (NB). The analysis of the performance has concluded that MLP has performed better than other classifier i.e., 87.00%, where the best performers have been successfully identified based on accuracy, precision, recall, F-measures and support performance measures.*
Keywords – Image Processing, Crop pant leaf, Pattern Variation, classification, Machine Learning

---

Date of Submission: 15-07-2022 Date of acceptance: 29-07-2022

I. INTRODUCTION

India is mostly a farming country. Farmers have a wide range of options when it comes to picking the right fruit and vegetable crop using photographs of diverse leaf patches. If sufficient care is not taken in this region, it can have major consequences for plants, affecting product quality, quantity, and productivity[1].Because of Crops become afflicted with numerous diseases as a result of changing seasonal conditions. These diseases first attack the plant's leaves then infect the entire plant, affecting the quality and amount of the crop grown[2].We know that The leaf is the main part of the plant that identifies its range and varieties, but leaf identification is a difficult and complex task due to the nature of the plant, which presents a vast pattern variation.Crop plants directly or indirectly provide food to humans and all other non-photosynthetic organisms, so we should explore more diversity of crop plants to meet the food shortage. But it is a very difficult task for humans to identify the leaves of the crop. Because human is not capable of the naked eye. Identify different types and species of leaves of similar-looking crop plants. Therefore, leaf detection of the crop plant is very essential, and it can be done using image processing techniques. The primary goal of this paper is the image classification of crop leaves,it is very important for us to recognize the category of plants leaves because most of the plant leaves look the same, which is very difficult to identify by human eyes, our paper is on crop leaf image classification in which we use machine learning techniques,and because these illnesses can spread, it's critical to diagnose each plant.Hence in this paper we are introducing crop plant leaf image classification using on machine learning techniques, but it can be used for identification and classification for other purposes as well.So, we can say that to identify many categories of crop and its types of leaves, we do image classification using machines so that its categories can be easily identified by the husbands. In this paper, we have introduced some categories of crops to perform image classification using machine learning of crop leaves.

Images are acquired with a digital camera on a cell phone and processed with image growing software, after which the leaf sport is utilized for train and test categorization. Image processing techniques and advanced computing skills have evolved into the system.

Image analysis can be used for a variety of purposes, including[1]:
1. To find sick leaves, stems, and fruits.
2. To determine the extent to which a disease has affected an area.
3. Determine the impacted area's limits.
4. Find out what color the damaged area is.
5. Determine the size and form of the leaf.
6. To appropriately identify the Object.

Today's time science is growing very fast. Due to continuous development, today computer has become a necessity in all our work.In the coming time, the machine age is about to begin or say that it has started. Where computers now have the ability to think and understand like humans. **So we will know about a very famous technique**, whose name is **machine learning**. So here first of all we will know that what is machine learning and how it works.

Machine learning(ML) is a form of artificial intelligence that enables a system to learn automatically.And if need be, he can also improve himself. ML can make systems learn automatically without being explicitly programmed.This work is an overview of this data analytics method which enables computers to learn and do what comes naturally to humans, i.e. learn from experience[3].In this, the system is made so efficient to work that the machine can complete that task on its own from next time on the basis of its previous experience, continuously improving it as humans do.Plant disease detection is a critical issue that must be prioritized for the sake of productive agriculture and the economy. Traditional methods for detecting plant disease are difficult to come by because they demand a lot of effort, time, and skill. The field of automated plant disease detection has recently gotten a lot of attention from academics, researchers, and physicians. Machine learning is detecting plant illness as soon as symptoms show on plant leaves. An in-depth examination is undertaken in this state-of-the-art review to evaluate the possibilities of employing machine learning models to identify plant diseases.

1.1 How Machine learning works

Machine learning is a form of artificial intelligence (AI) that teaches computers to think in a similar way to humans, such as learning from and improving upon past experiences, it works by identifying search patterns in data, and at least in human's interaction is involved. A big part of what makes machine learning so valuable is its ability to detect what has been left out of the human eye when reading or collecting data. Machine learning models can capture complex patterns that are overlooked during human analysis.
To understand the way machine learning works, it is very important to understand its types:

In general, following types of ML algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement

Machine learning employs two methods: supervised learning, which involves training a model on known input and output data in order to predict future outputs, and unsupervised learning, which involves uncovering hidden patterns or intrinsic structures in input data.

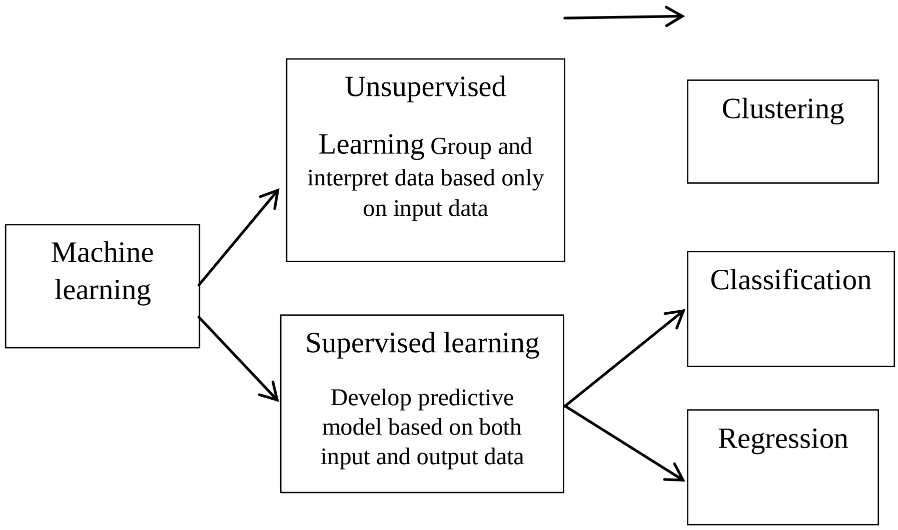

Fig-1: Unsupervised and supervised machine learning techniques

1.2Image Processing

It is the procedure of converting an image to a digital format and then executing operations on it to extract valuable information[4]. When implementing specific specified signal processing algorithms, the image processing system normally treats all images as 2D signals.Image processing, as the name implies, entails processing an image using a variety of approaches until we achieve our aim. **The basic definition of image**

processingrefers to processing of digital image, i.e., removing the noise and any kind of irregularities present in an image using the digital computer.
Some of the techniques used in image processing[4]

- Image representation
- image pre-processing
- image enhancement
- image restoration
- image analysis
- image reconstruction
- image data compression

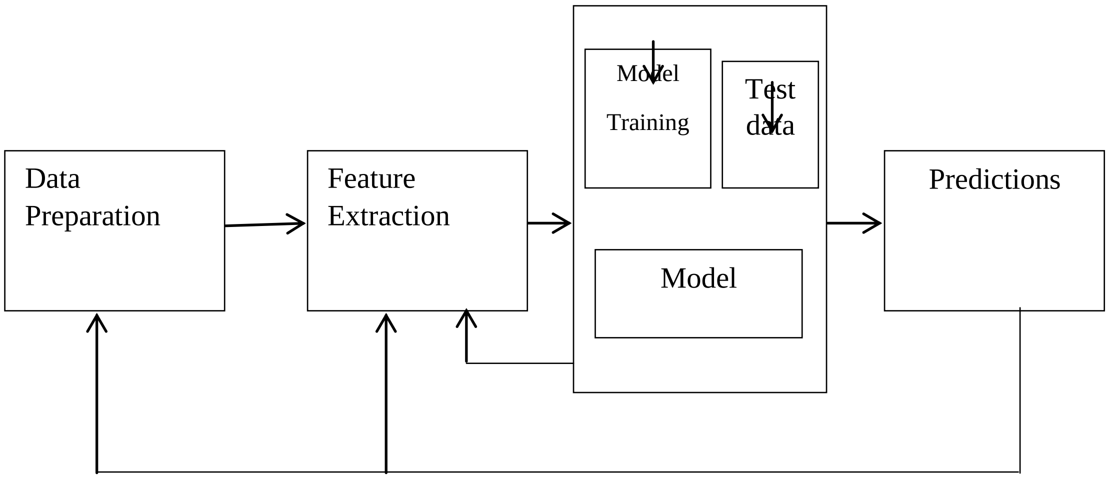

Fig-2: A traditional machine learning image processing approach is used.

Description of Problems

There are many categories and varieties of crop plants. Changes in the stage and nature of crop plants also depend on the climate of different places, in such a situation; their classification becomes very important for identification. Identifying so many categories of crop plant and its variety is very difficult by the human eyes. Technically Crop leaf prediction accuracy is a challenging problem, and many models have been proposed and validated so far. This problem requires the use of multiple datasets. The variety of crop harvested depends on various factors such as climate, season, soil used, etc.

The Proposed solution

Image classification is necessary to identify the category of the crop plant and its many varieties, which should be done by image processing and using machine learning. For this,images were captured with a digital camera on a cell phone and processed with image growing software, after which the leaf spot is utilized for training and test categorization. Image processing techniques and advanced computing skills have evolved into the system.Crop plants are identified by their leaves, so we use various type of classifier of machine learning to crop leaves identification and its accuracy. When collecting a dataset for a machine learning project, ensure that it is of high quality. It's criticalto make sure the data is of acceptable quality. Quality is important, but so is quantity. It's critical to have enough data to properly train youralgorithm. It's critical to pre-process the data by cleaning and finishing it, as well as annotating it, after it's beencollected.

Novelty of research:

- The main objective of this paper is to understand the classification of crop plants. Even though many researchers had worked in this segment, but in best of my knowledge, this is the first multiple crop plant leaves image dataset that has been created for 21 categories of crop plants with total 5931 RGB high-resolution images.
- The comparative study of classification performance of the classifiers has been done where the best performer has been identified successfully based on accuracy, precision, recall, F-measures, and support performance measures.
- Based on textual information, the highly misclassified leaf was examined, and it was discovered that the visual images of misclassified leaf categories were comparatively similar.

Organization report

The rest of this paper is categorized in 5 different sections as:

- Section-2 represents Literature review in which some of the existing works has been mentioned.
- Section-3 represents Problem definition and research objectives.
- Section-4 represents Experiment with data set description along with experimental design.
- Section-5 represents the experimental results and their detailed analysis.
- Section-6 represents conclusion with present and future research works.

II. LITERATURE REVIEW

In recent years, evolutionary neural networks have attracted much attention of the researchers because of their ability to give superior image classification accuracy. Several studies have found that image-based assessment methods produce more accurate and consistent results than human visual assessments[5]. They address any problem by combining neural networks with computation. In the 2012 Large Scaled Challenge for Visual Recognition, Krizhevsky et al. beat the second-best submission in ImageNet by 10.9 percent in classification accuracy. Advances in image processing have resulted in a variety of pre-processing techniques for picture extraction. The process of identifying discriminatory traits that serve as the foundation for classification is known as feature extraction. Multiple learning technologies, such as "Support Vector Machines" (SVM), "Nave Baye," "K-Nearest Neighbour" (KNN), and "Convolutional Neural Network" (CNN), can be used to solve the classification problem.

For the classification of plant leaves, Liu et al. suggested a ten-layer CNN model that attained an accuracy of 87.92 percent for the 32 classes. With the LeafSnap dataset, the ResNet model had a classification accuracy of 93.09 percent for plant identification. Photos acquired by Silva et al. using an Apple iPad tablet, plant leaf classification was performed by on the images captured by Silva et al. [6]. VGG16, VGG 19, and the Inception ResNetV2 model had classification accuracy of 91.5 percent, 92.4 percent, and 89.6 percent, respectively.The authors of [7] developed an algorithm that applies feature normalisation and dimensionality reduction to roughly 15 shape characteristics that are extracted. SVM has been applied for classification, and testing on the Flavia dataset yielded an overall accuracy of 87.40 percent.A technique that conducts classification by automatically extracting shape information was developed by Amlekar et al.[7]. With the help of a feed-forward back propagation neural network, classification has been carried out. The ICL dataset was used to evaluate this approach, and it achieved an accuracy of 96% for testing photos and 99% for training images.A method was created by Begue et al.[8]Using their own dataset, this contained images of the leaves of 24 different medicinal plants. From each leaf image, they retrieved shape-based characteristics. Random forest classifier achieved the greatest accuracy of 90.1 percent out of the several classifiers used (k-NN, naive Bayes, SVM, neural network, and random forest). In terms of multitemporal satellite imaging, the author claims that a lot of work has been done on crop classification using machine learning. In terms of crop health and yield surveys, some remote sensing researches have focused on each of the physical factors of cropping systems, such as nutritional value reference and water availability. Field surveys are used to create maps resulting from image interpretation and land cover and land use research. The images used depict a frame that will be covered. It is repeated in the study region or on other dates multiple times. Resolution and scale are frequently discussed. Agricultural systems, on the other hand, are frequently more endemic[9]. In a computational recognition system, leaf attributes such as form, size, and colour are critical. Contour-based or region-based extraction can be used to extract features. The descriptors for the contour-based extraction are length, width, aspect ratio, and leaf diameter. The main vein of the leaf is used to calculate the length descriptor, which spans from the main vein to the end tip. The width descriptor refers to the length of a leaf when viewed from one side to the other, from the leaves leftmost to rightmost points. The aspect ratio is calculated by dividing the length of a leaf by its width. The leaf diameter is the largest distance between two points inside the covered region of the leaf, whereas the perimeter is the maximum distance of the area covered by the leaf[10]. Once the classification of the plant is completed, work on the classification of disease can be further extended. Because ofCrops become afflicted with numerous diseases as a result of changing seasonal conditions. These diseases first attack the plant's leaves then infect the entire plant, affecting the quality and amount of the crop grown[11].

A collection of 58,200 crop leaf photos, spanning 14 different crops and 37 different classifications of healthy/diseased crops, was employed in the study. Different diseases of the same crop displayed striking commonalities among them. With an F1 score of up to 93.05 percent, the NASNetLarge fine-grained classification model based on the proposed attention mechanism had the best classification effect. The findings suggest that the proposed attention mechanism improves the fine-grained classification of crop disease photos significantly[12].

III. PROPOSED METHODOLOGY

In this work, we have collected more than 250 images of each 21 categories of crop plants' leaves. Figure 1 depicts the proposed system's operational flow. According to authors in [13], there are three types of images: scans, pseudo-scans, and pictures, depending on how the image was captured. In the scan and pseudo-scan categories, the leaf images are taken using the scanning method and photography, respectively; that is, they are taken indoors against a simple background. The third category includes pictures of plants that were taken in their natural habitats. Scan and pseudo-scan images are frequently used by researchers because they are simple to analyse[14].Then the pictures were cropped to remove the defects in the data images as much as possible. In this work, the crop plant leaf has been classified with the help of various machines learning (ML) model's algorithm to find out the performance of each classifier.The crop plant leaf is manually collected, and data processing followed by future extraction is done then various machine learning models were implemented then and their predictions were obtained these models have been evaluated based on their classification accuracy.

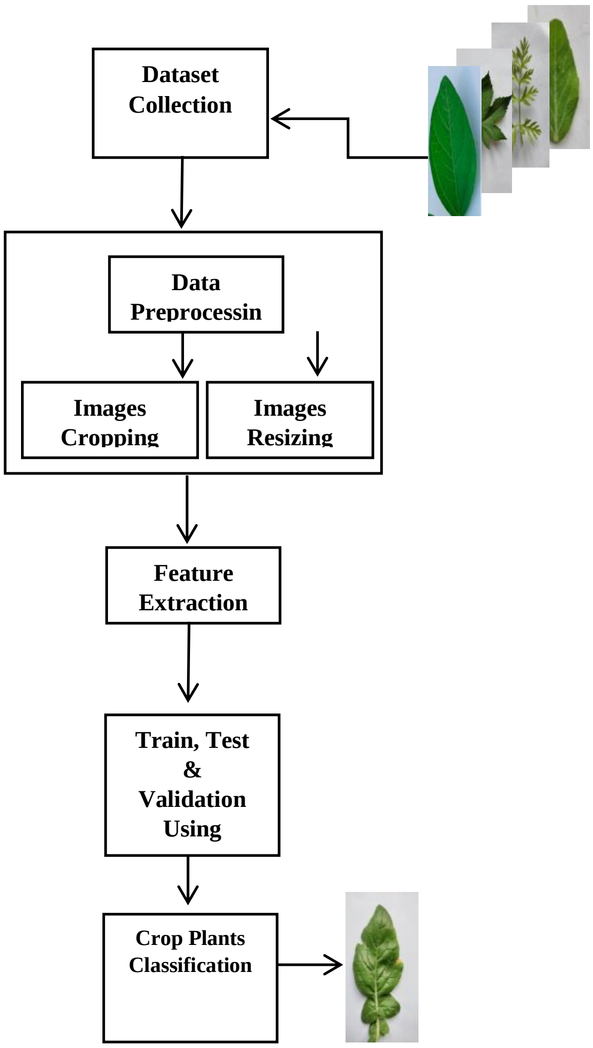

Fig-3: The Workflow of Crop Plants Leaf Classifications

The following subsections cover the specifics of each phase.

3.1 Data Preparation

In every computer vision-based system, data preparation is critical. Each image's preparation processes are depicted. As datasets for experimentation, we have taken images of leaves of 21 categories of crops. A total of 5931 RGB images have been taken from all categories combined. The image of a leaf is prepared using some operations. For instance, color space conversion from RGB (Red-Green-Blue) to another space is used for reducing the dependence on the device.[15]In RGB format image is captured and put in the machine for reading the dataset.It is represented adequately in a numerical format.This data set is organized in a tabular manner, with each row representing a different sample and each column a particular feature.Let's crop plant leaves as Dataset is D, which represents the collected crop plant leaves.

Say $D = \{x_i, y_i\}_{j=1}^{k}$
Where, $x_i \rightarrow$ associated label or images for each i^{th} image in the dataset.
$and \quad y_i \rightarrow$ its corresponding labeland, $y_i \rightarrow$ Element of Y ($y_i \in Y$)
i.e., Y=$\{y_i\}_{i=1}^{k}$;where |Y| is several classes of crop plant leaf
For j= {1, 2, 3,……., k} and k→The no. of samples in dataset for all ($y_i \in Y$)

3.2 Dataset pre-processing (Cropping, Resizing)
Pre-processing of an image is essential step since it improves the quality of the image for subsequent processing. This stage is essential since noise is present in images by nature, which could lead to less accurate classification.The image processing system normally treats all images as 2D signals, and each data point is represented by a two-dimensional vector. Each dimension corresponds to a certain aspect of a data point, which we indexed. As j = 1, 2, …k. so, the dataset is written as a set of data point-label pairs {($x_{1,}$ y_1) …(x_i, y_i), …, ($x_i y_j$)}. The table of data points $\{x_1, x_{2,\dots}, x_i\}$ can be written as $X \in R^{j*k}$.
In dataset pre-processing all images of dataset (D) have been cropped into a specific size and resized into a new dimension. The resize function f is used to change images into new dimensions. The new dimension of dataset is represented by D^r. The resized function $f()$ has shown in (1).

$$D^r = f(\{x_j^r, y_j\}_{j=1}^{k}, p \times q) \tag{1}$$

3.3 Feature Extraction
Principal Component Analysis, or PCA, is an algorithm for linear dimensionality-reduction approach for reducing the dimensionality of large data sets by transforming a large collection of variables into a smaller one that retains the majority of the information in the large set. PCA is related to singular value decomposition and to Eigen valueand Its allows to learn the important features from these high dimensional data that help compress, summarize, classify or rank them, without sharing the data directly.[16]Since $\boldsymbol{X}$ represents the data matrix, where rows represent the different images. A larger explained variance is expressed by the axis with the highest eigenvalues.The principal component is data covariance matrix. Its result the new matrix denoted by $\boldsymbol{D_{PCA}}$ has shown in (2).

$$\boldsymbol{D_{PCA}} = \boldsymbol{X}(f(\{\mathbf{x_j^r}, \mathbf{y_j}\}_{\mathbf{j=1}}^{\mathbf{k}}, \mathbf{p} \times \mathbf{q}), \boldsymbol{n}) \tag{2}$$

n→ no. of features after extraction using PCA
3.4Splitting Dataset into Train, Test Set, &Validation
Initially, the datasethas been splitting each class image into three parts: Training set, Test set,and Validation set.Let X=(X_1,………..,X_k) be the dataset and k be the no. of images.[16] It is $\boldsymbol{X^r}$split into test, train, validation andTest dataset namely $\boldsymbol{T_{rn}}, \boldsymbol{V_{ld}}, \boldsymbol{and\, T_{st}}$ respectively.i.e.,$\boldsymbol{T_{rn}} \subset \boldsymbol{X^r}$, $\boldsymbol{V_{ld}} \subset \boldsymbol{X^r}$ and$\boldsymbol{T_{st}} \subset \boldsymbol{X^r}$.
Where, $\boldsymbol{X^r} = \boldsymbol{T_{rn}} \cup \boldsymbol{V_{ld}} \cup \boldsymbol{T_{st}}$ and $\boldsymbol{T_{rn}}$ and $\{\boldsymbol{T_{rn}} \cap \boldsymbol{V_{ld}}\} \cup \{\boldsymbol{T_{rn}} \cap \boldsymbol{T_{st}}\} \cup \{\boldsymbol{V_{ld}} \cap \boldsymbol{T_{st}}\} = \emptyset$.
3.5-Applying ML Classifier
After splitting the classifier Ø Trained ($\boldsymbol{T_{rn}}$) and Validation ($\boldsymbol{V_{ld}}$) is denoted as$\emptyset(\mathbf{T_{rn}}, \mathbf{V_{ld}})$.For Prediction then, Let predictor function $\boldsymbol{f(i)}$ which is obtained for test samples or data point *i* consisting of **P** illustrative variables and**m →no.** of test samples.
For *i* = 1,..., *n*, has shown in eq^{n}(3).

$$\boldsymbol{f(i)} = \{\boldsymbol{P_i^{\emptyset}}\}_{\boldsymbol{i}}^{\boldsymbol{m}}, \tag{3}$$

3.6: ObtainAccuracy& Prediction:

One parameter for evaluating classification models is accuracy. Informally, accuracy refers to the percentage
Of correct predictions made by our modelusing accuracy definition:

$$\text{Acc} = \frac{\text{No. of correct Predictions}}{\text{Total No. of Predictions}}$$

Now, using 0-1 loss functionto accuracy can be calculated, so let's my set of class labels is ***y_test$_i$*** and Function as *L* .The function look asshown in eqn (4).

$L(i, j) =$	{	0	$i=j$	$i,j \in \boldsymbol{y\_test_i}$
		1	$i \neq j$	

(4)

So, a model on this training data, using ML classifier, and this model classified several objects correctly and assigned them the correct class labels shown in (5)

$$Acc = L(\{y\_test_i, f(i)\}) \text{ or } L\left(\{y\_test_i\}_{i=1}^{m}, \{P_i^{\emptyset}\}_{i=1}^{m}\right) \quad (5)$$

Where $\boldsymbol{f(i)}$ is a prediction obtained by classifier $\emptyset$ for the test samples, and m denotes the number of test sample.

IV. EXPERIMENT

4.1 DATASET DESCRIPTIONS

As datasets for experimentation, we have taken images of leaves of 21 categories of crops. A total of 5931 images have been taken from all categories combined. Brassica juncea (Mustard), Cajanus Cajan (Pigeon pea),Hordeum vulgare (Barley), Lens culinaris (Lentil), Ricinus(Castor oil) etc. are the few examples from these 21 categories. Each class has at least more than 250+ images in it. The detailed data set description has been **shown in Table 1**, where the sample image of each class along with their botanical/scientific name as well as their common name and number on sample images, just below each class sample image, has been mentioned. The method divides each class image into three parts: Training set, Validation set, and Test set. The validation and test sets each contain 50 images; whereas the training set has more than 250 images from each category.According to this passage, data and its experimental process and result are given below respectively.

Different classifiers have now been used to undertake a comparative analysis, and the top-performing classifier has been identified. After the data has been pre-processed, the ML algorithm can run, and the features of these leaf images have been extracted in the feature extraction section. The approach divides each class into train, validation, and test sets after feature extraction, and several classifiers are applied one after the other to obtain performance in the form of accuracy, precision, recall, and so on. Finally, at the step of crop plant leaves classification, we were able to classify plants depending on the results of the previous stage.

Table 1 List of All 21 Categories of Crop Plants leaf with Sample Image





Brassica junceaBrassica oleracea varCajanus cajanDaucus carrotHordeum vulgare
(Mustard) : 260 (Cabbage) :250 (Pigeon pea) :250 (wild carrot) :269 (Barley) :343





IpomoeaBatatasLuffa acutangulaRaphanus sativusPachyrhizusErosusVicia faba
(Sweet potato) :298 (Thurai) :338 (Radis) :308 (Kesaur) :186 (broad bean) :250





RicinusSaccharum officinarumLinum usitatissimumPisum sativum Phaseolus vulgaris
(Castor oil) :250 (Sugarcane) :251 (Flex) :300 (Pea) :357(Kidney bean):301





Zea maysSolanumTriticumZingiberIndian pea
(Maize) :275 (Potato) :266 (Wheat) :347 (Ginger) :267 (Grass pea) :272

Vigna radiata
(Mung bean):293

4.2 Experiments Design

The first phase is the collection of data set. In this we collected more than 250 images of leaves in each 21 categories of cropplants. The data collection (Crop Leaves) we went for the data pre-processing phase in which the collected images had been cropped, resized, and organised based on their class.Then firstly we had collected more than 250 images of each 21 categories of crop plants' leaves both side on a white board. Then the pictures were cropped to removes the defects in the data images as much as possible. After this the images were resized into 250×250 size and each category of images was stored in separate folders with their names.After that, we have used the classifier of machine learning to find accuracy by analyzing these datasets. The Flowchart of data processing can be seen in Figure-4.

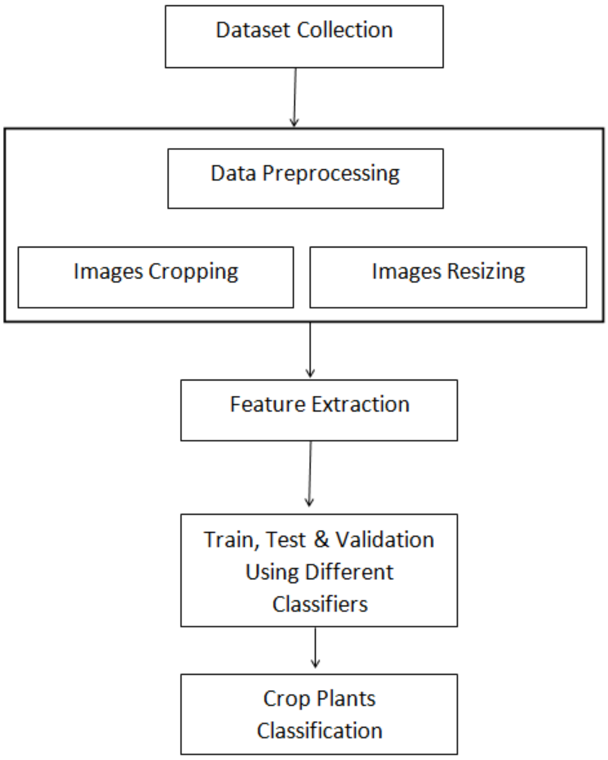

Fig-4: The Workflow of Crop Plants Leaf Classifications

The hardware and software used for this research*:* The hardware and software used in this research has been shown in Table 2.

Table 2. Shows the Hardware & Software Requires for the Classification

Hardware Requirements	• An ordinary smartphone capable to capture decent normal images of leaves. • System/machine requirements to process the images: - Processor Intel(R) Xeon(R) Silver 4210 CPU @ 2.20GHz 2.19 GHz Installed RAM 64.0 GB (63.7 GB usable) System type 64-bit operating system, x64-basedprocessor Edition Windows 10 Pro for Workstations Version 21H2
Software Requirements	• Machine Learning Library-scikit-learn[7] • IDE - Anaconda Navigator 2.1.1, JupyterLab 3.2. • Programming Language- Python 3.9.7 64-bit

Performance Evaluation Parameter:

Accuracy, Precision, Recall, F-Score, and ROC results are used to draw conclusions about the experimental results. To conduct a comparative analysis, a bar plot and a box plot of the accuracy of various ML classifiers were created. Finally, the heat map depicts the relationship between several categories.

V. Results and Analysis

5.1 Results

The suggested approach was examined using a crop dataset that includes 5931 images of 21 different crops as table 1 gives the names of all the crops. Typically, the collection includes images of a single leaf of a plant. The majority of the leaves are mature and in good condition. Only a very small percentage of the leaves are slightly bent or twisted.The dataset exhibits substantial intra-class similarity and, in a very small number of instances, inter-class similarity.Figure 4& 5 illustrates the classification performance of the six classifiers, LR, KNN, SVM, DT, MLP, and NB, as a bar plot and box plot in text. The x-axis in thisgraph reflects the classifiers' names, while the y-axis represents their accuracy. The accuracy of each classifier has been displayed directly above the bar plot and image for easy viewing.

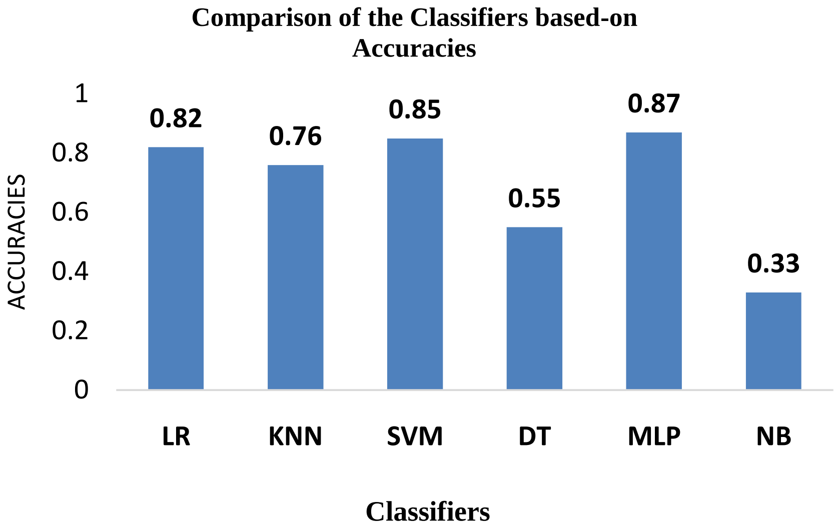

Figure 4. Bar Plot of ML algorithm Classifiers

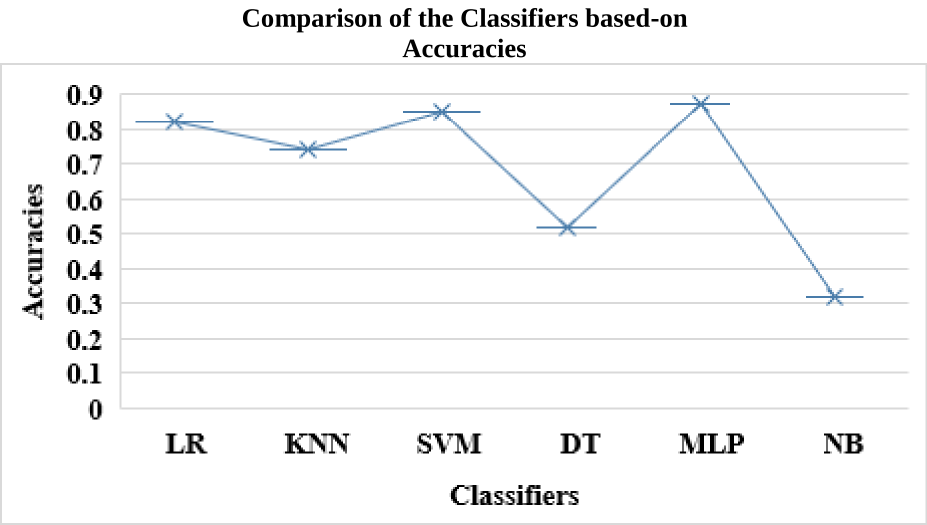

Figure 5. Bar Plot of ML algorithm Classifiers

Table-03Best classifier's (MLP) classification performance

Iteration= 3 Algo.: MLPClassifier				
S.No	**Precision**	**Recall**	**f1-score**	**Support**
0	0.90	0.85	0.87	91
1	0.75	0.84	0.79	56
2	0.92	0.88	0.90	69
3	0.98	0.92	0.95	60
4	0.81	0.88	0.85	59
5	0.96	0.95	0.96	57
6	0.78	0.81	0.80	63
7	1.00	0.96	0.98	48
8	0.96	0.93	0.94	69

9	0.88	0.87	0.87	84
10	0.64	0.75	0.69	59
11	0.89	0.89	0.89	88
12	0.92	0.97	0.94	61
13	0.71	0.74	0.73	70
14	0.71	0.71	0.71	73
15	0.83	0.91	0.87	79
16	0.94	0.94	0.94	53
17	1.00	0.97	0.99	77
18	0.82	0.58	0.68	92
19	0.85	0.92	0.88	96
20	0.91	0.94	0.92	79
accuracy			0.86	1483
macro avg	0.87	0.87	0.86	1483
weighted avg	0.86	0.86	0.86	1483
Algo:	**MLPClassifier Accuracy: 0.8742414025623736**			

Table 3shows the best classifier's (MLP) classification performance (precision, recall, f1-score, and support).

Figure 6 uses a heat map to display the confusion matrix for our crop dataset. It shows how various crop categories have been classified as well as incorrectly classified. The x-axis and y-axis both display the names of the crop plants categorized in this study. The color scheme used in this heat map spans from 0 to 7. The brighter color of a cell in the diagonal elements, which run from left-top to right-bottom, denotes a considerably greater correlation between the category and itself, demonstrating that the classifier's classification was accurate. A number of misclassifications have occurred, though, if the colour of these diagonal cells is darker (closer to 0).However, the narrative will be completely the opposite save for this diagonal cell. There is no connection between that crop class and the corresponding other crop class in that situation, as the color is darker (closer to 0).

Ploting confusion matrix....

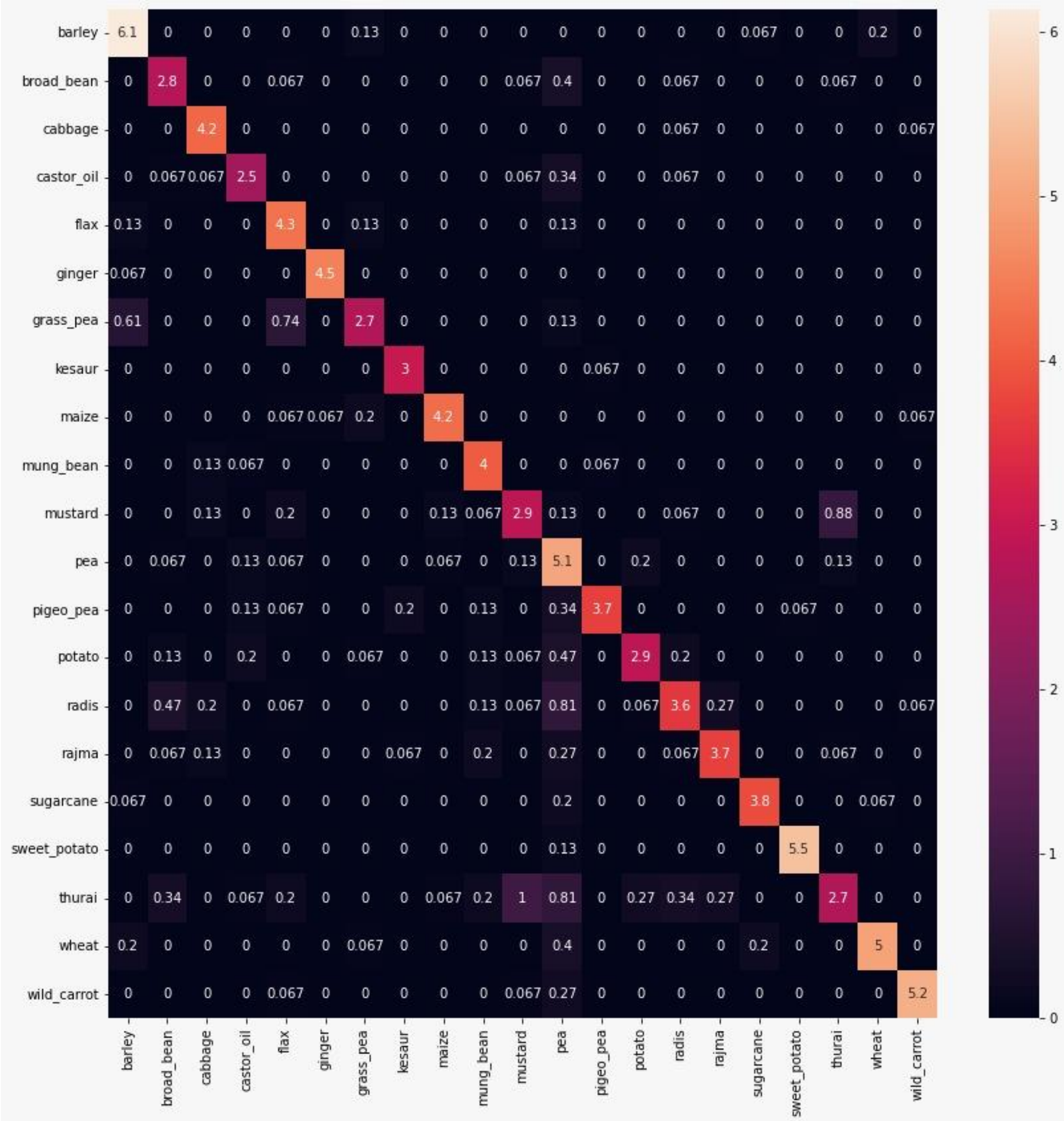

Figure 6. Heat-Map Showing the Correlation amongst Different Classes of crops

5.2. Analysis of Results

Classification comparison using bar plot: As we can see from the bar plot in Figure 4, Linear Regression (LR) has 82.58%, k-Nearest Neighbours (KNN) has 76.36%, Support vector Machine (SVM) has 85.82%, Decision Tree (DT) has 55.09%, Neural Network (MLP-Multi layer perceptron) has 87.89%, and the Naïve Bayes (NB) has 33.90% classification accuracy. So, on basis of classification accuracy if we will choose MLP, which has the maximum accuracy, then this classifier for the classification of crop plants will give best result as compared to other classifiers. And the NB will give worst classification report because it has least accuracy.

Classification comparison using box plot: The evaluation criteria for the analysis of a box plot are Max Maximization, Min Maximization, Size of box plot and the median of the box plots. From figure 5 we can see that MLP has the maximum max, the maximum min, the median has maximum value, and it has the smallest box size as compared to other classifiers. So again, MLP classifier is performing best based on all four evaluation criteria of a box plot.

Classification comparison based on Performance: Theperformance of the MLP classifier has been shown in Table 3.

Performance analysis of best performer based on confusion matrix: Figure 6 demonstrates that Pisum sativum is receiving the best classification (Pea). The level of misclassification is quite high for both Ipomoea Batatas (sweet potatoes) and Zea mays (maize) at the same time. This is because of how similar its leaf image looks to the leaf images of other crops in its class, such Zea Mays (Maize).

VI. Conclusion

In this paper, we classified 21 different agricultural plant leaf photos into 21 different groups. For each of the 21 crop plant classifications, more than 250 pictures were painstakingly collected. As a result, the dataset that is being used in this work is unique. To begin the classification process, the shape feature is extracted. Second, to obtain classification accuracies, several approaches such as k-Nearest Neighbours, Linear Regression, Support Vector Machine, Decision Tree, Neural Network, and Nave Bayes were used. The analysis was drawn using box plots, bar plots, and heat maps based on these findings. Using the Multilayer Perceptron(MLP) Classifier of the Machine Learning model, we acquire accuracy of 87.90 percent.

In the future, the Deep Learning(DL) model will be applied to the same 21 types of crops dataset to achieve higher accuracy than the best Machine learning model, and the correlation of some crops, such as Pisum sativum (Pea), Ipomoea Batatas (Sweet Potato), and Zea Mays (Maize), will be improved through proper data analysis. Other algorithms can also be evaluated to choose the best classifier. DL can be used as a more advanced solution for severely misclassified categories. More crop picture categories can be gathered for classification.By using advanced technology and classifier's, better results and conclusions can be drawn from these collected crop leaves.

References:

[1]. V. S. Arti N. Rathod, Bhavesh Tanawal, "Image Processing Techniques for Detection of Leaf Disease," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 19, no. Suppl 1, p. P557, 2013.

[2]. P. Sharma, P. Hans, and S. C. Gupta, "Classification of plant leaf diseases using machine learning and image preprocessing techniques," in *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, Jan. 2020, pp. 480–484. doi: 10.1109/Confluence47617.2020.9057889.

[3]. I. Ranggadara *et al.*, "Machine Learning from Theory to Algorithms: An Overview," *J. Phys. Conf. Ser.*, vol. 1142, no. 1, p. 012012, Nov. 2018, doi: 10.1088/1742-6596/1142/1/012012.

[4]. K. Chithra and T. Santhanam, "A novel denoising technique for mixed noise removal from grayscale and color images," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 3, pp. 626–642, 2018.

[5]. G. Wang, Y. Sun, and J. Wang, "Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning," *Comput. Intell. Neurosci.*, vol. 2017, 2017, doi: 10.1155/2017/2917536.

[6]. J. Liu, S. Yang, Y. Cheng, and Z. Song, "Plant Leaf Classification Based on Deep Learning," *Proc. 2018 Chinese Autom. Congr. CAC 2018*, pp. 3165–3169, Jan. 2019, doi: 10.1109/CAC.2018.8623427.

[7]. S. Kaur and P. Kaur, "Plant Species Identification based on Plant Leaf Using Computer Vision and Machine Learning Techniques," *J. Multimed. Inf. Syst.*, vol. 6, no. 2, pp. 49–60, 2019, doi: 10.33851/jmis.2019.6.2.49.

[8]. A. Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, "Automatic Recognition of Medicinal Plants using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 166–175, 2017, doi: 10.14569/ijacsa.2017.080424.

[9]. Nirbhay Bhuyar, "Crop Classification with Multi-Temporal Satellite Image Data," *Int. J. Eng. Res.*, vol. V9, no. 06, 2020, doi: 10.17577/ijertv9is060208.

[10]. M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. Abdullah, and S. R. W. Alwi, "Review on techniques for plant leaf classification and recognition," *Computers*, vol. 8, no. 4, 2019, doi: 10.3390/computers8040077.

[11]. P. Sharma, P. Hans, and S. C. Gupta, "Classification of plant leaf diseases using machine learning and image preprocessing techniques," *Proc. Conflu. 2020 - 10th Int. Conf. Cloud Comput. Data Sci. Eng.*, pp. 480–484, Jan. 2020, doi: 10.1109/CONFLUENCE47617.2020.9057889.

[12]. G. Yang, Y. He, Y. Yang, and B. Xu, "Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism," *Front. Plant Sci.*, vol. 11, Dec. 2020, doi: 10.3389/FPLS.2020.600854/FULL.

[13]. J. Wäldchen and P. Mäder, *Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review*, vol. 25, no. 2. Springer Netherlands, 2018. doi: 10.1007/s11831-016-9206-z.

[14]. J. Wäldchen, M. Rzanny, M. Seeland, and P. Mäder, "Automated plant species identification—Trends and future directions," *PLoS Comput. Biol.*, vol. 14, no. 4, Apr. 2018, doi: 10.1371/JOURNAL.PCBI.1005993.

[15]. M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Applied Artificial Intelligence An International Journal Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," 2017, doi: 10.1080/08839514.2017.1315516.

[16]. S. X. Wu, H. T. Wai, L. Li, and A. Scaglione, "A Review of Distributed Algorithms for Principal Component Analysis," *Proc. IEEE*, vol. 106, no. 8, pp. 1321–1340, 2018, doi: 10.1109/JPROC.2018.2846568.