International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Detecting and Classifying Plant Diseases Automatically Using Machine Learning Algorithms

Dr. Sreedhar Bhukya [1] & Y. Monica [2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Food production must rise in tandem with the rapid growth of the human population. Diseases that spread quickly can seriously reduce plant yields and possibly wipe off entire crops. It is essential to increase food production given the rapidly expanding world population. But the danger of quickly spreading illnesses is real, with the potential to completely destroy agricultural products as well as severely damage crop yields. Acknowledging the critical significance of early disease identification and prevention, this study explores innovative approaches that leverage the widespread availability of cell phones, even in the most remote rural locations. This research focuses on the application of automated image analysis as a workable substitute for traditional methods that depend on expensive laboratory processes and human expertise—materials that are frequently noticeably lacking in less developed areas. The study carefully compares the efficacy of latest developments in deep learningtechniques with more conventional machine learning algorithms to present the most recent developments in this rapidly developing subject. The main objective is to improve agricultural practices to lessen the detrimental effects of disease on crop yields and promote sustainable food production in light of the world's rapidly growing population.

***Keywords** : K - Nearest Neighbor, Support Vector Machine and Convolutional Neural Network.*

1. Introduction

Food manufacturing needs to increase to keep up with the world's population growth. Diseases that spread quickly have the potential to completely kill entire harvests and severely reduce plant production. It is essential to increase food production given the rapidly expanding world population. But the danger of quickly spreading illnesses is real, with the potential to completely destroy agricultural outputs as well as severely damage crop yields. Acknowledging the critical significance of early disease identification and prevention, this study explores innovative approaches that leverage the widespread availability of cell phones, even in the most remote rural locations.

This research focuses on the application of automated image analysis as a workable substitute for traditional methods that depend on expensive laboratory processes and human knowledge—foundations that are frequently noticeably lacking in less developed areas. The study carefully compares the efficacy of the latest in deep learning techniques with more conventional machine learning algorithms to present the most recent developments in this rapidly developing subject. The main objective is to maximize agricultural methods to lessen the detrimental consequences of diseases on crop yields and promote sustainable food production in light of the world's rapidly growing population.

The introduction describes the current global trends in smart automobile technology, specifically in the areas of complicated embedded computerization and digitization, which are giving rise to new system designs and intelligent platforms. It emphasizes the value of soft computing approaches, especially deep learning and convolutional neural networks, in tackling problems with tracking, identifying, and simulating important events and drivers' emotional strain in autonomous vehicles. To ensure safety and comfort while driving, the introduction also emphasizes the importance of quick detection and monitoring of specific car parameters as well as the driver's emotional and attentive state. In several domains, including information security, identity authentication, law enforcement, and access control systems, it presents facial recognition as a crucial biometric identification tool.

Fundamentally, the introduction establishes the framework for the proposed paper's emphasis on creating efficient soft computing techniques to improve smart automobile technology, especially in resolving issues linked to drivers. In the automobile sector, digitization, automation, and the integration of information and communication technology (ICT) at all levels of car control and diagnostics are just a few of the developments that are being highlighted, along with the necessity of adjusting to them.

[1]Professor, Department of Computer Science Engineering, Sreenidhi Institute of Science and Technology Hyderabad, India Email: sreedharb@sreenidhi.edu.in

[2]Student of Mtech, Department of Computer Science Engineering, Sreenidhi Institute of Science and Technology Hyderabad, India Email: monicavijaya99@gmail.com

2. Literature Survey

Initial Applying appropriate management practices, such as these, can yield details regarding crop health and illness identification. The administration of fungicides, chemical treatments tailored to specific diseases, and pesticides for vector control. Both productivity and illness control may benefit from this.

Authors [1] define, evaluate, and accept the necessity of creating a health-monitoring sensor that is dependable, quick, and affordable in order to support agricultural improvements.In order to create a ground-based sensor system that will support field monitoring of plant health and diseases, they outlined the current technologies in use, which Spectroscopy, imaging, and volatile profiling-based methods for identifying plant diseases.There are now numerous applications of computer vision for the detection of illnesses in plants. As the authors have demonstrated, one of these is the identification of illness using the extraction of colour characteristicsLuma (brightness), blue minus luma (B-Y), red minus luma (R-Y), and yellow minus luma (Cb) are all represented by the symbols YcbCr.Hue, Saturation, Intensity Model (HSI)
Since it depicts colours in the same way that the human eye perceives them, it is a highly significant and appealing colour model.
Colour models from CIELB In CIELAB, the French term for the International Commission on Illumination, Commission Internationale de l'Eclairage, is abbreviated as CIE. To measure objective colour and compute colour differences in the CIELAB colour space, the three values are represented by the letters L*, a*, and b*. were used in the investigation and successfully locate illness sites despite external noise such as flash from cameras. To detect diseases on maize leaves, they experimented using colour extraction in conjunction with [3].

When all these elements are combined, a strong feature set is produced that can be used to improve images and improve categorization. An overview of common classical feature extraction methods has been given by the authors in.Owing to the swift advancement Regarding the discipline of Artificial Intelligence, the primary emphasis of this study is the implementation of these approaches and strategies. It is also possible to recognise and differentiate between several plant diseases using support vector machine algorithms. In [4] Moreover, the classification accuracy ranged from 65% to 90% based on the kind and stage of the disease. This approach was used to treat problems of sugar beets and was presented in.The authors of the paper [5] described methods for using deep learning to tackle the most difficult issues in a range of biological, Biological Informatics, biomedical, robotics, and 3D technology research fields. Motivated by the practical application of deep learning techniques and their evolution, We employ the deep learning method in our work to identify plant diseases.

A thorough review of the most recent studies on the subject turned up no proof that scientists have investigated the use of deep learning methods to recognise plant illnesses from images of leaves.The parts that follow will outline our deep CNN recognition technique.

3. Methodology

In the conventional approach, before feeding the input into machine learning (ML) algorithms, well-liked substitutes like It is necessary to perform feature extraction and image pre-processing before using k-Nearest Neighbours (k-NN) algorithm with support vector machines (SVM). neural networks using convolutions(CNN) and additional methods for deep learning (DL), on the other hand, have shown a noteworthy trend in recent years towards their application to picture categorization challenges. This change is explained by the fact that DL methods regularly perform better than classical algorithms, especially when given reasonably sized datasets. Most notably, hand-engineered features are no longer necessary with DL techniques, which improves efficiency and streamlines implementation. The purpose of the suggested system is to compare, particularly with regard to plant disease categorization, the effectiveness of DL techniques and traditional ML algorithms.

4. System Architecture

This architecture combines deep learning and traditional machine learning methods for comparison, and it covers the entire process of plant disease detection, from data collecting to model deployment. Depending on the particular objectivesand the features of the dataset, adjustments can be necessary.

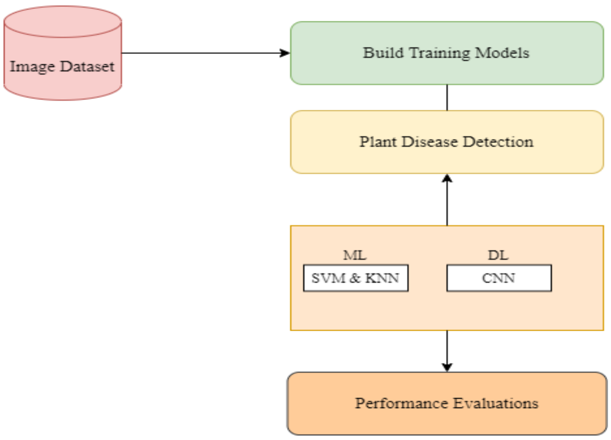

Fig 1

1. Image Dataset:

Collect a well-labeled dataset with pictures of both healthy and diseased plants. Make sure the dataset is representative of many plant diseases and varieties, as well as diversified.

2. **Data Preprocessing:**

Resize photos, adjust pixel values, and use data augmentation methods to clean up and preprocess the dataset. This guarantees that the model performs properly when applied to new data.

3. Training Traditional ML Models:

Using the features that were extracted, train SVM and KNN models. Utilising the validation set, adjust hyperparameters to maximise performance.

4. Deep learning,

Convolutional neural networks (CNNs) are used for feature learning. can be used for deep learning to automatically extract hierarchical features from unprocessed image data.

5. CNN Model Training:

Train the CNN model with the previously processed images.To get better performance, especially with little data, use transfer learning using pre-trained models.

6. Performance Evaluation:

To evaluate how well the deep learning model (CNN) and the more traditional ML models (SVM, KNN) fared on the test set, use metrics like accuracy, precision, and recall.

7. Comparison and Selection:

Examine how well the CNN, KNN, and SVM models perform in comparison. Choose the model that performs best overall based on pertinent measures such as accuracy.

8. Deployment:

Install the chosen model in a working context. Numerous techniques, including web apps, APIs, and interaction with current agricultural systems, can be used to accomplish this.

9. Monitoring and Updates:

To monitor the model's performance in real-world circumstances, use monitoring. Update the model on a regular basis with fresh data to increase its robustness and accuracy over time.

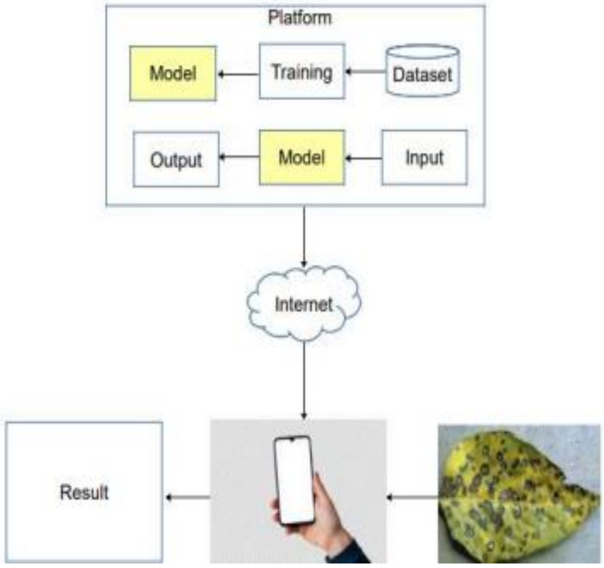

Fig 2

FUNCTIONAL REQUIREMENTS

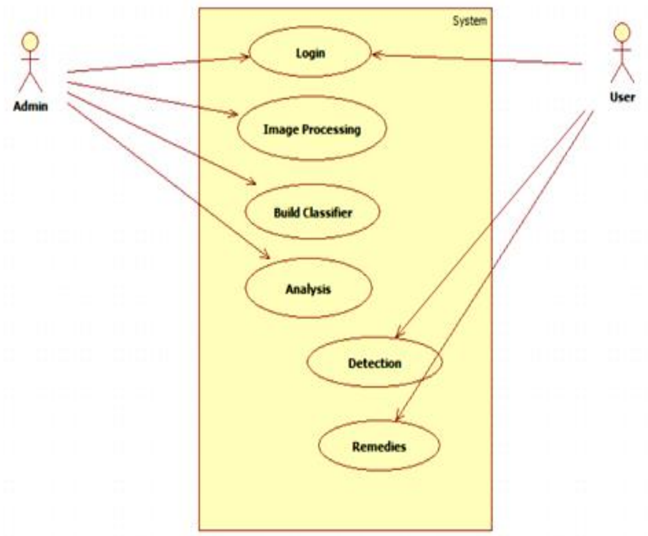

Fig 3

IMPLEMENTATION

An offline augmentation of the original dataset is used to recreate this one. This GitHub repository has the original dataset available for you to access. The about 87K rgb photos in this collection, which are divided into 38 groups, show both healthy and damaged crop leaves. The training and validation sets are split up into 80/20 ratios while maintaining the directory structure of the entire dataset. To aid with prediction, a new directory with 33 test photos is later generated.[4,5]

Convolutionary A neural network that shares its parameters is called a covnet.

Types of layers:

1.The input layer
2. Convolutional Layer

3. Layer of Activation Functions
4. The Pool Level
5. Completely Networked Layer

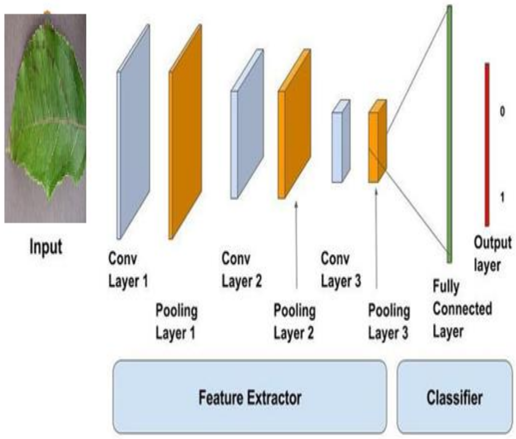

Fig 4

1. **The Input Layer:**
   In this layer stores raw input images, which has dimensions 32 by 32 by 3.
2. **Convolution Layer:**
   In this layer determines of output volume by taking the image patch and the dot product of all the refiners.Should we apply twelve filters in total to this layer, the output volume will measure thirty-two by thirty-two by twelve.
3. **Layer ofActivation Function :**
   This layer takes the output of the convolution layer and appn element-wise activation function to it.The activation function layer adds an element-wise activation function to the output of the convolution layer.
4. **Pooling Layer**
   The pooling layer is crucial to an image's pre-processing. When the image size becomes too huge, the pooling layer reduces the number of parameters. Pooling involves decreasing in size. the image that was acquired from the earlier layers. It's like reducing the size of an image to lower the pixel density. Another name for spatial pooling is downsampling or subsampling, which lowers each map's dimensionality while keeping the crucial details.

5. **Fully Connected Layer**

   The input from the other layers will be transferred to the fully linked layer after being flattened into a vector. By using the network, it will convert the output into the required number of classes.

SYSTEM TESTING

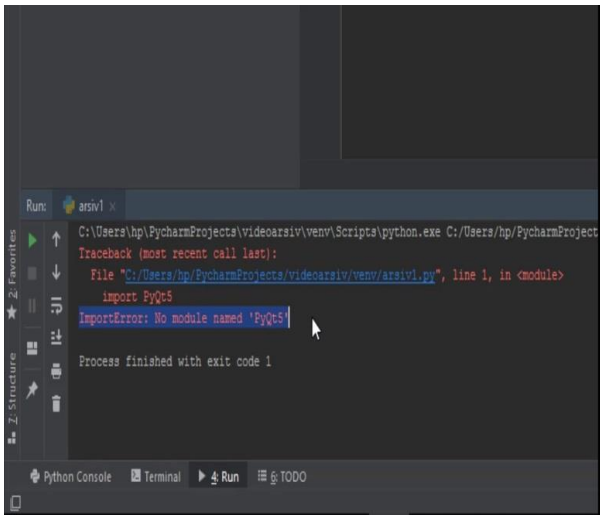

Fig 5

Figure illustrates the error that appears when the pyqt5 module is not found.

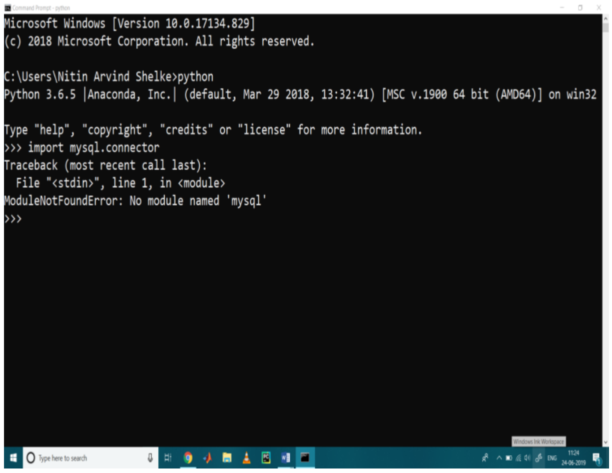

Fig 6

The error that appears in fig. when the MySQL module cannot be found

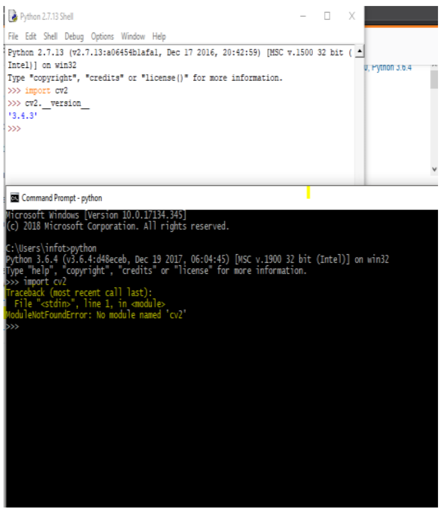

Fig 7

If the CV2 module cannot be located, the error message displayed in Figure

TABLE.1.TESTING CASES

SNO	Test Case	Test	Expected
1	CNN Training	Folder Structure	YES
2	Image Upload	NULL	YES
3	Prediction	Input Image	YES

DATASET

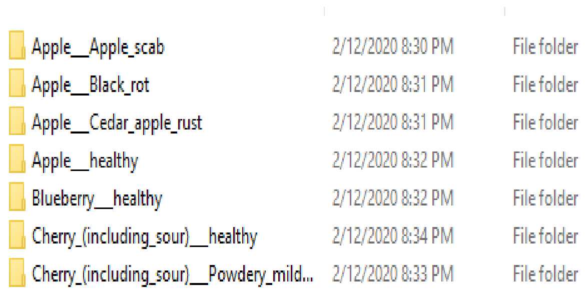

Fig 8

Apple___Apple_scab

Fig 9

Apple___Black_rot

Fig 10

Blueberry healthy

Fig 11

SCREEN SHORTS

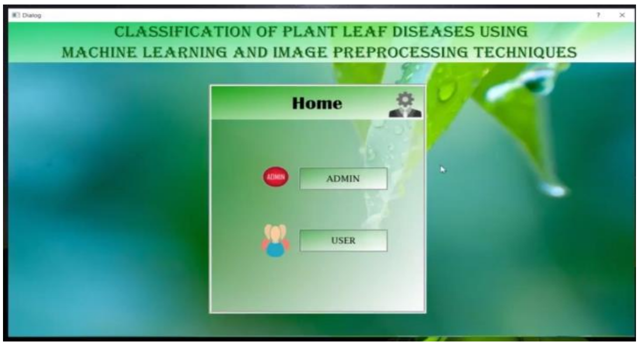

Fig 12: Home Dialog box

Fig 13: Admin Login dialog box

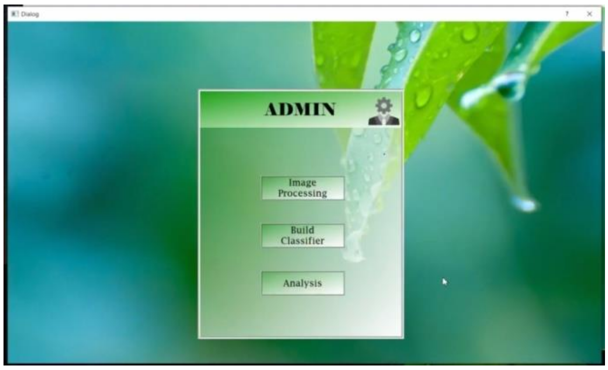

Fig 14: The dialogue window for the admin portal

```
[INFO] Loading images ...
[INFO] Processing Tomato_Septoria_leaf_spot ...
[INFO] Processing Tomato__Tomato_mosaic_virus ...
[INFO] Processing Tomato_Late_blight ...
[INFO] Processing Tomato_Spider_mites_Two_spotted_spider_mite ...
[INFO] Processing Tomato__Tomato_YellowLeaf__Curl_Virus ...
[INFO] Processing Tomato_healthy ...
[INFO] Processing Pepper__bell___Bacterial_spot ...
[INFO] Processing Potato___healthy ...
[INFO] Processing Tomato__Target_Spot ...
[INFO] Processing Tomato_Leaf_Mold ...
[INFO] Processing Tomato_Bacterial_spot ...
[INFO] Processing Tomato_Early_blight ...
[INFO] Processing Potato___Late_blight ...
[INFO] Processing Potato___Early_blight ...
[INFO] Processing Pepper__bell___healthy ...
[INFO] Image loading completed
```

Fig 15: Outcomes of Preprocessing Images

```
[INFO] Training KNN model...
[INFO] Training KNN model created successfully..!
[INFO] Training SVM model...
[INFO]  Training SVM model created successfully..!
```

Fig 16: ML training Model process

```
Epoch 1/25
73/73 [==============================] - 47s 644ms/step - loss: 0.1989 - acc: 0.9367 - val_loss: 0.4419 - val_acc: 0.9122
Epoch 2/25
73/73 [==============================] - 39s 532ms/step - loss: 0.1316 - acc: 0.9538 - val_loss: 0.2397 - val_acc: 0.9382
Epoch 3/25
73/73 [==============================] - 41s 557ms/step - loss: 0.1191 - acc: 0.9570 - val_loss: 1.2767 - val_acc: 0.8832
Epoch 4/25
73/73 [==============================] - 39s 529ms/step - loss: 0.1093 - acc: 0.9605 - val_loss: 0.5535 - val_acc: 0.9276
Epoch 5/25
73/73 [==============================] - 39s 531ms/step - loss: 0.1226 - acc: 0.9575 - val_loss: 0.6102 - val_acc: 0.9175
Epoch 6/25
73/73 [==============================] - 39s 537ms/step - loss: 0.1196 - acc: 0.9577 - val_loss: 0.5770 - val_acc: 0.9063
Epoch 7/25
73/73 [==============================] - 39s 530ms/step - loss: 0.1023 - acc: 0.9632 - val_loss: 0.5210 - val_acc: 0.9248
Epoch 8/25
73/73 [==============================] - 40s 550ms/step - loss: 0.1026 - acc: 0.9633 - val_loss: 1.7190 - val_acc: 0.8799
Epoch 9/25
73/73 [==============================] - 40s 550ms/step - loss: 0.1238 - acc: 0.9570 - val_loss: 1.9751 - val_acc: 0.8761
Epoch 10/25
73/73 [==============================] - 41s 559ms/step - loss: 0.1035 - acc: 0.9632 - val_loss: 1.0540 - val_acc: 0.8950
```

Fig 17: DL training Model process

DL Training and Validation Accuracy

The estimation of validation accuracy serves the purpose of determining the model's generalization ability to fresh, untested data. To detect overfitting, a model that performs poorly on the training set yet uses fresh data should be identified.

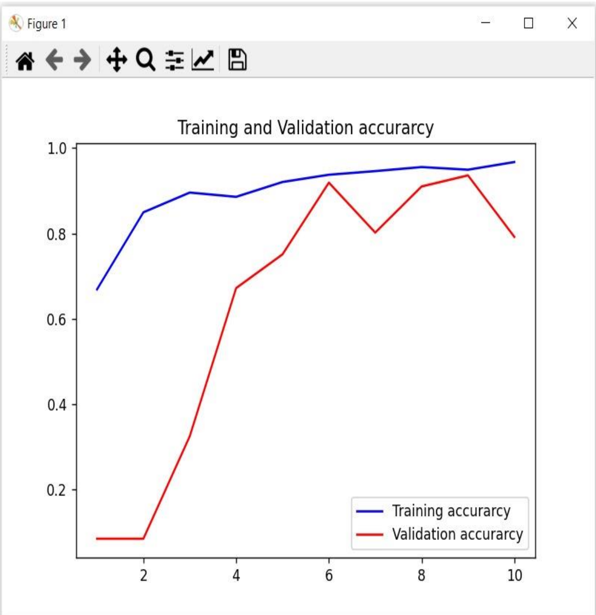

Fig 18: Accuracy of DL Training and Validation

DL Loss of Training and Validation

Validation loss aids in evaluating the generalizability of the model. Over fitting may be indicated by an increase in validation loss. when new data causes the model to perform poorly, although it performs well on training data.

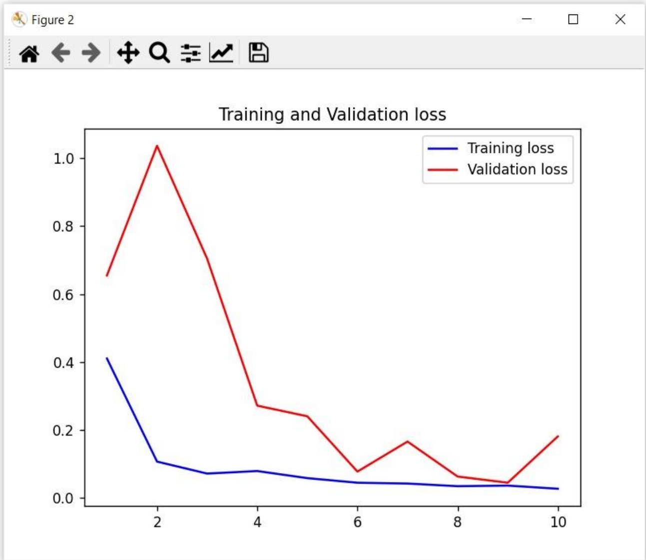

Fig 19: DL Training and Validation Loss

Table.2.CNN model training process:

Epochs	Training Loss	Training Accuracy	Val loss	Val accuracy
1	0.4107	0.6690	0.6551	0.0850
2	0.1072	0.8495	1.0367	0.0850
3	0.0719	0.8957	0.7043	0.3250
4	0.0794	0.8857	0.2719	0.6720
5	0.0585	0.9205	0.2405	0.7510
6	0.0450	0.9375	0.0778	0.9190
7	0.0426	0.9460	0.1661	0.8020
8	0.0348	0.9555	0.0632	0.9100
9	0.0366	0.9492	0.0450	0.9360
10	0.0273	0.9672	0.1813	0.7920

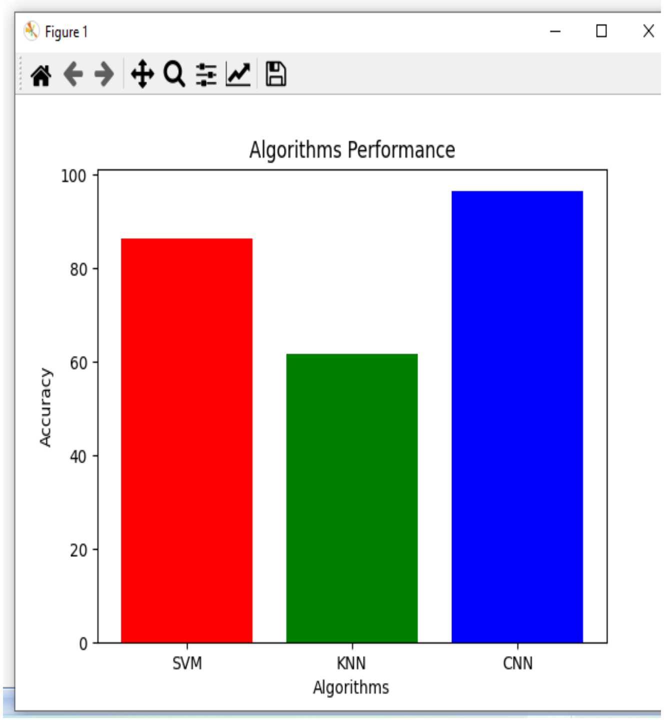

Fig 20: Comparisons between ML and DL models

Table.3. ML and DL models performance comparisons:

Algorithm name	**Accuracy**
SVM	86.44
KNN	61.77
CNN	96.49

In evaluating different machine learning and deep learning algorithms for the goal of plant disease detection and classification, convolutional neural networks (CNN) clearly perform better than other conventional techniques. CNN exhibits outstanding skill in accurately identifying and categorizing plant diseases from photos, with an accuracy rate of 96.49%. Support Vector Machines (SVM) performs admirably as well, with an accuracy of 86.44%. With an accuracy of 61.77%, the K-Nearest Neighbors (KNN) method, however, lags behind considerably. These findings demonstrate how well deep learning methods—in particular, CNN—handle complicated visual data, including pictures of sick plants, opening up potential new possibilities for the creation of reliable and effective agricultural diagnostic instruments.

5. Result Screen Shorts

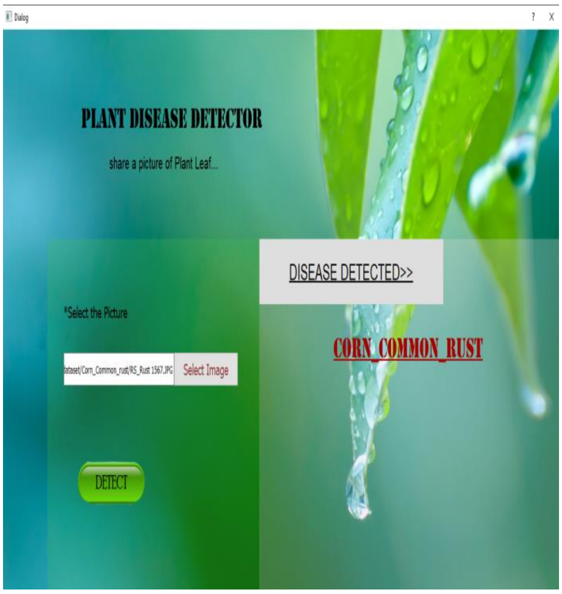

Fig.21 Plant disease detection

6. Conclusion

Despite the wide range of methods available for Plant disease categorization and identification using automated or computer vision techniques, this field of studystill has shortcomings.Furthermore, aside from commercial systems that identify plant species from photographs of their leaves, there are currently none on the market. In order to automatically identify and classify plant illnesses using photographs of leaves, this article explored a novel deep learning technique. The created model could identify when leaves were present and differentiate 13 distinct diseases that could be identified visually from healthy leaves.[6] Every step of process was covered, including gathering the training and validation photographs, preprocessing and enhancing the images, and lastly training and fine-tuning the deep neural network. To evaluate the newly developed model's performance, several experiments were run. A freshly created database of plant disease images was produced, with over 3,000 original photos that were taken from the internet and enhanced to over 30,000 with the right editing.For tests conducted on different classes, The precision of the experiment ranged from 91% to 98%. 96.3% was the final accuracy of the trained model.[9] Although the augmentation method had more of an impact to produce acceptable outcomes, Adjustment hasnot demonstrated appreciable alterations in the total accuracy. There hasn't been a comparison with similar outcomes utilising the exact procedure as the suggested procedure hasn't, been used, to the best of our knowledge, to identify plant diseases. Outcomes were equivalent, or even better, than with other approaches employed and given in Section 2, especially when the larger number of classes in provided study is taken into consideration.

References

[1] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers, “Climate change effects on plant disease: genomes to ecosystems,” Annual Review of Phytopathology, vol. 44, pp. 489– 509, 2006.

[2] S. M. Coakley, H. Scherm, and S. Chakraborty, “Climate change and plant disease management,” Annual Review of Phytopathology, vol. 37, no. 1, pp.399–426,1999.

[3] S. Chakraborty, A. V. Tiedemann, and P. S. Teng, “Climate change: potential impact on plant diseases, Environmental Pollution, vol. 108, no. 3, pp. 317–326, 2000.

[4] A. J. Tatem, D. J. Rogers, and S. I. Hay, “Global transport networks and infectious disease spread,” Advances in Parasitology, vol. 62, pp. 293–343, 2006.

[5] J. R. Rohr, T. R. Raffel, J. M. Romansic, H. McCallum, and P. J. Hudson, “Evaluating the links between climate, disease spread, and amphibian declines,” National Academy of Sciences of the United States of America Proceedings vol. 105, no. 45, pp. 17436–17441, 2008.

[6] T. Van der Zwet, “Present worldwide distribution of fire blight,”in Volume 590 of the Proceedings of the 9th International Workshop on Fire Blight, October 2001, Napier, New Zealand.

[7] S. A. Miller, F. D. Beed, and C. L. Harmon, “Plant disease diagnostic capabilities and networks,” Annual Review of Phytopathology, vol. 47, pp. 15–38, 2009.

[8] M. B. Riley, M. R. Williamson, and O. Maloy, “Plant disease diagnosis. The Plant Health Instructor,”2002.

[9] J. G. Arnal Barbedo, “Digital image processing techniques for detecting, quantifying and classifying plant diseases,” SpringerPlus, vol. 2, article 660, pp.1–12,2013.

[10] H. Cartwright, Ed., Artificial Neural Networks, 2015; Humana Press.