VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM, KARNATAKA

MINOR-PROJECT-I

REPORT ON,

"FOLIAR DISEASE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES"

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by

USN	NAME
2SD20CS032	BHAVANA BORAH
2SD20CS056	MOHAN LAMANI
2SD20CS073	PRAGATI
2SD20CS078	PRASANNAKUMAR

Under the Guidance of

Prof. Yashodha Sambrani
Dept. of CSE, SDMCET, Dharwad

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
S.D.M. COLLEGE OF ENGINEERING & TECHNOLOGY,
DHARWAD-580002

2023

S.D.M COLLEGE OF ENGINEERING & TECHNOLOGY,

DHARWAD –580002

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the Minor-Project-1 work and presentation entitled **"FOLIAR DISEASE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES"** *is a bonafide work carried out by* **BHAVANA BORAH (2SD20CS032), MOHAN LAMANI (2SD20CS056), PRAGATI (2SD20CS073),** and **PRASANNAKUMAR (2SD20CS078),** *students of S. D. M. College of Engineering & Technology, Dharwad, in partial fulfillment for the award of* **Bachelor of Engineering in Computer Science and Engineering of** ***Visvesvaraya Technological University*, Belgaum**, *during the year 2022-2023. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The Minor-Project-1 has been approved, as it satisfies the academic requirements in respect of project report prescribed for the said degree.*

----------------------	-----------------------
Prof.Yashodha Sambrani	**Dr. S M Joshi**
Project Guide	HOD-CSE
Name of Examiners	**Signature with date**

1.

2.

ABSTRACT

Agriculture is one of the main factor that decides the growth of any country.

IT is an important sector of the Indian economy as it contributes 19% of total GDP in the year 2020-21. In recent years, the rate of growth in agriculture production has been declining due to various diseases in plants with varying climatic conditions. In India itself around 65% of the population is based on agriculture. Due to various seasonal conditions the crops get infected by various kind of diseases. These diseases firstly affect the leaves of the plant and later infected the whole plant which in turn affect the quality and quantity of crop cultivated. As there are large number of plants in the farm, it becomes very difficult for the human eye to detect and classify the disease of each plant in the field. And it is very important to diagnose each plant because these diseases may spread. Hence an artificial intelligence based automatic plant leaf disease detection and classification for early detection of disease is a much needed system.

Table of Contents

PROBLEM STATEMENT

Agriculture is an important sector of the Indian economy as it contributes 19% of total GDP in the year 2020-21. In recent years, the rate of growth in agriculture production has been declining due to various diseases in plants with varying climatic conditions. Diseases are the major reason for crop loss every year and it is really a challenge to control diseases. Adoption of precision agriculture is a key solution to overcome the above problems. Hence an Artificial Intelligence based disease recognition system would help the farmers in early detection of diseases, reduction on cost of pesticides and can get good returns for the efforts.

CHAPTER 1: INTRODUCTION

In agriculture, plant diseases happen because of climate changes and changes in weather cycles from place to place that deteriorates the health of a crop. Crops may get infected by the virus, fungal and bacterial infections. Early detection of these diseases can allow taking preventive measures and mitigate economic and production losses. In fact, about 60% to 70% of disease appears on leafs only.

Pest and disease are the major obstacles in increasing chili production. Chili plant diseases like Anthracnose and Thrips make 35%-90% chili productions are lost. Incorrect diagnosis in detecting pests and diseases of chili plants make countermeasures become ineffective and inefficient. Consultation with experts in horticulture is one of the ways to diagnose pests and diseases of chili plant, which is considered to be the correct action. But the limited number of experts in horticulture and the limited consultation time become the obstacles in detecting pests and diseases of chili plants.

One solution for these problems is by transferring knowledge from experts in horticulture. The knowledge base of an expert can be transferred into computer system. This computer system contains a distribution of knowledge and new communication channel that contains the expert knowledge base and a system called as expert system. Expert system is computer programs that emulsify expert behavior in problem solving related to knowledge domain in particular field.

CHAPTER 2: LITERATURE SURVEY

Kantha Raju Kanaparthi et al. in their study has considered chilli crop diseases, the dataset consisting mainly of two types of diseases Geminivirus and Mosaic. The dataset consisted of 160 leaf images taken kaggle open source database. The author has used Squeeze-Net convolutional neural networks . He has worked on various training optimizer algorithms such as Stochastic Gradient Descent with Momentum (SGDM) optimizer, Adam optimizer, Root Mean Squared Propagation (RMSProp). SGDM has reached 50% accuracy whereas RMSProp and ADAM both reach 100% accuracy with 35 to epoch runs. The author has concluded that RMSProp with SqueezNet Deep Convolution Neural Network suits best for chilli disease dataset reaching 100% accuracy [1].

Sufola Das Chagas Silva Araujo et al. observed that Goa was facing many problems during the growth of chilli crops, to study the same they had captured images of chilli leaf using a digital camera. The dataset consisted of 192 leaf images which was later resized to 30x45x35cm. The dataset consisted of leaf images ranging from one to five which were classified into different stages. For the segmentation process FCM and PSO methods were used and the features were extracted using the GLCM algorithm. CNN is used for training the dataset. After all these processes they obtained an accuracy of 98.77%, precision of 72% and sensitivity of 83.5%. [2]

Dyah K. Agustika et al. collected 24 samples of chilli leaves from 3 different fields of Indonesia situated at different locations, to examine if FTIR spectroscopy can still detect PYLCV attacks on chilli plants of different regions. Eight samples were taken from two different trees, 4 from a healthy plant and 4 from an unhealthy plant. During Fourier transform infrared (FTIR) spectroscopy, the environment was kept at a constant 25 C and 75% humidity. The measurements were taken at a spectral resolution of 8 cm, resulting in a data spacing of 0.964 cm. The transmittance spectrum of airborne KBr was employed as the background during measurements in the transmittance mode. The raw data spectra in the 4000-400 cm wavenumber range without preprocessing were passed through a classification model. The same process is repeated with the preprocessed image. Then the data were cut into 1800-900 cm bio fingerprint regions and the same technique used earlier is repeated for these bio fingerprint regions in order to remove the C02, water absorption spectra or to remove baseline distortion impacts(if exists). [3]

Yuslena Sari et al. considered chilli leaf image disease detection. The images were captured from the Tambak Anyar area, Banjar District through their smartphone with an 8 Megapixel camera. The dataset consist of 25 images, each 5 belonged to different classes of disease i.e Fusarium Wilt, Leaf Curl, Leaf Spots, Ralstonia Bacterial Wilt, and Yellow Virus. The author has used Gray Level Cooccurrence Matrix (GLCM) for extracting features from the image and Support Vector Machine (SVM) for classification. This model chilli disease detection has achieved an accuracy of 88% [4].

B. Nageswararao Naik et al. examined the classification capabilities of machine learning and deep learning for the categorization of chilli leaf disease, twelve distinct pre-trained deep learning networks both with and without augmentation were used to classify images of the five most common leaf diseases. Without augmentation, VGG19 seemed to have the best accuracy, and DarkNet53 got the best possible outcome with augmentation. The best accuracy was produced by a squeeze-and-excitation-based convolutional neural network (SECNN) model, which reached an accuracy of 99.28% in categorizing 43 different classes of plant leaf datasets. Its accuracy was 98.63% without augmentation and 99.12% with it. [5]

CHAPTER 3: DETAILED DESIGN

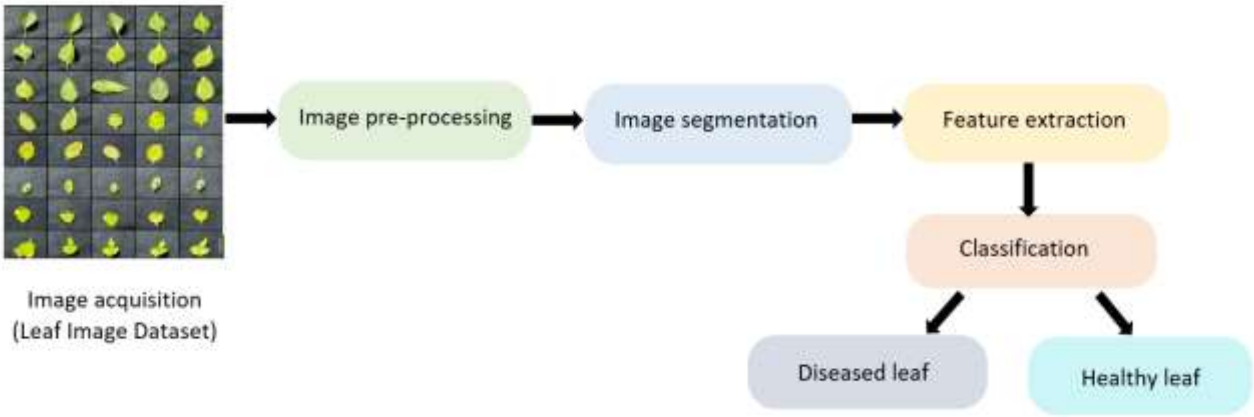


Fig.: Framework for chilli disease classification

According to Fig.,

Image Acquisition:

Image acquisition is the first and foremost step for the other process. This process involves capturing of digital images from a physical or digital source and it's an essential step for various applications such as plant disease detection and leaf image processing. The impact of image acquisition depends on the quality of images captured, which is affected by factors such as camera specifications, lighting conditions, and distance from the object. The steps involved in image acquisition typically include image capture, processing, and storage, and may involve pre-programmed movement sequences of the camera and the object being captured, as well as image resizing and feature extraction for improving accuracy in image processing. The resulting digital images can then be stored, manipulated, and analyzed using digital tools and software.

Image pre-processing:

Preprocessing is an essential step that involves transforming raw data into a format that can be analyzed effectively, accurately and easily understandable by algorithms. Preprocessing layer improves the quality of the data, reduces errors and noise, and prepares the data for effective analysis and modeling.

Preprocessing involves data cleaning by removing or filling in missing values, dealing with outliers or anomalies, and correcting any inconsistencies or errors in the data; data transformation by converting categorical variables into numerical variables, scaling or normalizing the data, transforming the data using mathematical functions; data reduction includes reducing the dimensionality of the data by eliminating irrelevant features, reducing the number of observations, sampling the data; data integration includes integrating data from multiple sources by merging or joining them together; data formatting

includes formatting the data in a way that is suitable for analysis; feature engineering includes creating new features from existing data

Image Segmentation:

Image segmentation refers to the process of dividing a digital image into multiple segments or regions based on similar characteristics such as color, texture, intensity, or other visual properties. The goal of image segmentation is to simplify or separate the input image into distinct regions or segments based on their visual features, by dividing the image into smaller and more manageable pieces that can be analyzed and processed independently. Image segmentation involves extracting visual features using convolutional filters to identify relevant patterns, spatial localization by mapping the visual features to the corresponding pixels in the input image to identify the boundaries between different segments, feature fusion uses skip connections to fuse features from earlier layers with features from later layers to improve the accuracy of segmentation, semantic segmentation by assigning a class label to each pixel in the image to separate image from the background, instance segmentation by identifying and separating individual instances of objects in the image for object detection and tracking.

Feature Extraction:

The process of extracting or selecting a subset of relevant information or features from a larger and more complex dataset. It involves identifying and extracting features of interest. It help to reduce the dimensionality of a dataset, handles missing data, improves performance so that the data can be more easily analyzed, interpreted that leads to more efficient and accurate models . The purpose of this step is to convert the input image into a numerical array of features that can be used for representing the image uniquely.

Image Classification:

Image classification is a process of categorizing or labeling an image using machine learning algorithms to recognize and categorize the objects, scenes, or patterns present in an image. The importance of the classification layer lies in its ability to predict the target variable or class label of input data based on its features or attributes, it produces an output in the form of a probability distribution or a categorical label that assigns the input data to one of several predefined classes. The classification layer uses a model that has been trained on labeled data to make these predictions. The performance of the classification layer can be evaluated using various metrics to assess how well the model is able to classify new data.

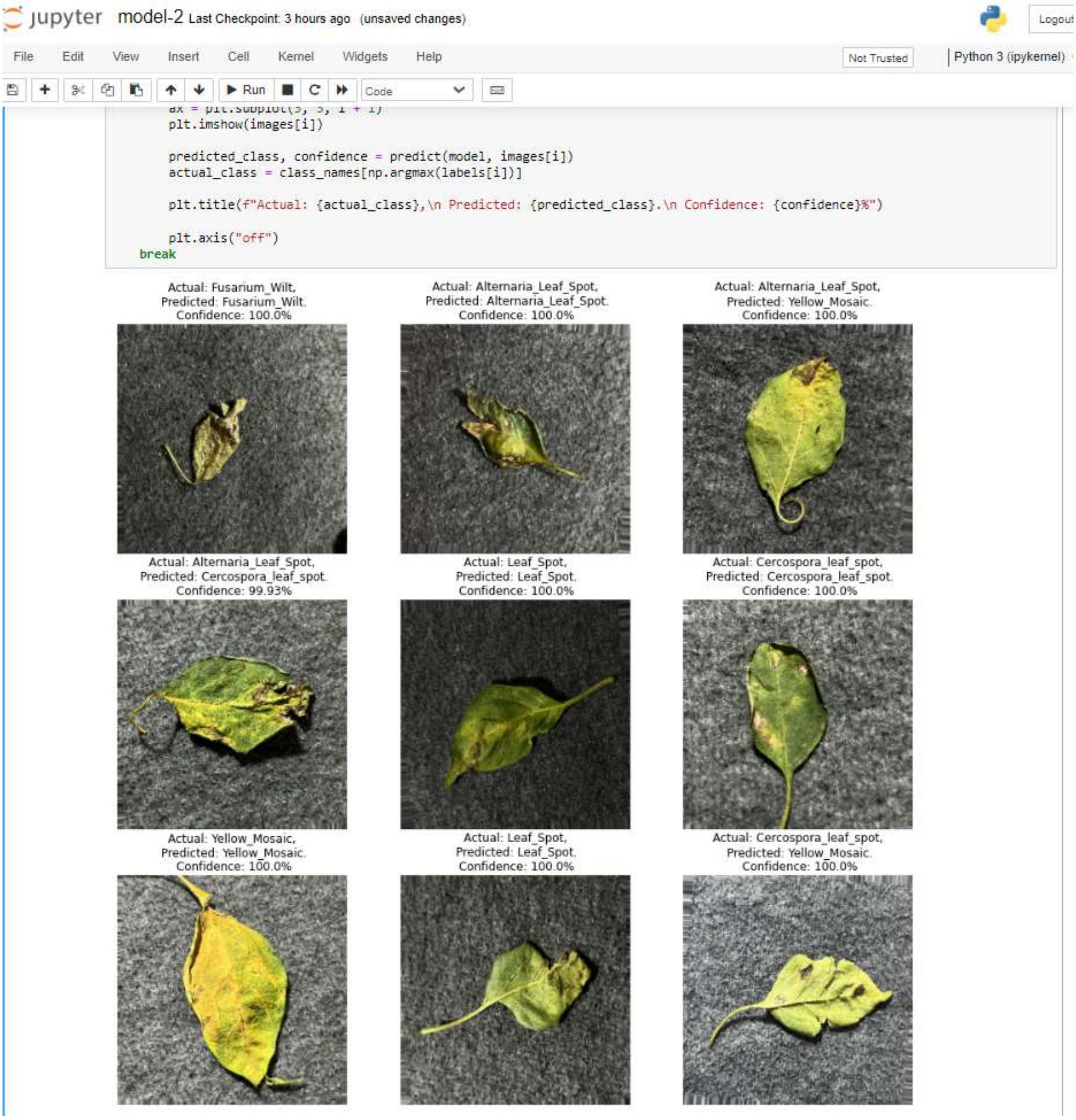
CHAPTER 4: PROJECT SPECIFIC REQUIREMENTS

- Construction of a real time dataset following these device configurations
  Distance: 20 cm from leaf to camera and 1.8x zoom on camera
  Lighting Conditions:
    Time: with led bulb around 10 pm
    Dimensions: 3024x4032
    Camera:Apple iphone 13 mini and realme narzo 50A phone
    F-stop: f/1.6 and f/1.8
    focal length:5mm and 4mm

- System Requirements

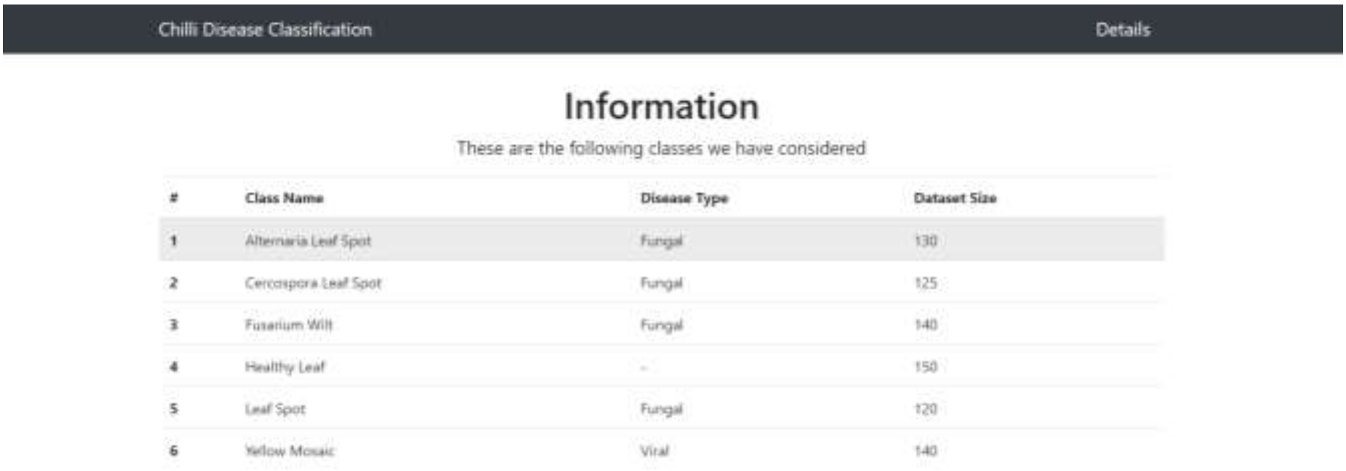
  1) Operating System - win10(64-bit) and above
  2) IDLE (Python GUI) - 3.10.4 version and above
  3) Deep/machine learning library – Tensorflow
  4) Web Application Backend-Flask
  5) Jupyter Notebook and Google Colab with GPU

CHAPTER 5: IMPLEMENTATION



FRONT END:

DETAIL PAGE:



#	Class Name	Disease Type	Dataset Size
1	Alternaria Leaf Spot	Fungal	130
2	Cercospora Leaf Spot	Fungal	125
3	Fusarium Wilt	Fungal	140
4	Healthy Leaf	-	150
5	Leaf Spot	Fungal	120
6	Yellow Mosaic	Viral	140

DATASET CONSTRUCTION

Sl no	Disease Name	1st Month	2nd Month	3rd Month	4th Month	5th Month	6th Month
1	Alternaria Leaf Spot	-	15	34	27	25	35
2	Cercospora Leaf Spot	-	7	25	26	28	30
3	Downy Mildew	-	-	2	6	8	12
4	Leaf Spot	12	23	21	32	21	26
5	Leaf Curl	8	12	21	25	26	21
6	Necrosis	-	-	21	22	25	27
7	Chlorosis	-	-	12	23	21	32
8	Bacterial Leaf Spot	4	23	26	28	24	32
9	Healthy Leaf	45	34	37	42	29	38
10	Anthracnoze	-	-	2	6	7	9
11	Fusarium Wilt	23	30	32	43	32	34
12	Powdery Mildew	-	-	2	3	5	8
13	Vein Banding	8	10	12	24	32	34
14	Yellow Mosaic Disease	-	23	34	36	43	47
15	Irregular Leaf	-	-	2	4	7	9

CHAPTER 6: RESULTS

Chilli Disease Detection

Using Our Own Dataset

CNN:

1) Import preprocessing layer in front of VGG
2) Using imagenet weights
3) Image Size = 224x224
4) Train set : rescale = 1./255, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = true
   Test set: rescale = 1./255
5) Parameters:
   Total params: 138,56,4
   Trainable params: 138,56,4
   Non-trainable params: 0
6) Uses a batch size = 32, Classification mode = categorical
7) Loss method = categorical_crossentropy
8) Optimizer method = adam
9) Epochs = 20
10) Accuracy = 80%

InceptionV3:

1) Import preprocessing layer in front of VGG
2) Using imagenet weights
3) Image Size = 224x224
4) Train set : rescale = 1./2, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = true
   Test set: rescale = 1./255
5) Parameters:
   Total params: 22,109,990
   Trainable params: 307,206
   Non-trainable params: 21,802,784
6) Uses a batch_size = 32, Classification_mode = categorical
7) Loss method = categorical_crossentropy
8) Optimizer method = adam
9) Epochs = 20
10) Accuracy = 93%

ResNet152V2:

1) Import preprocessing layer in front of VGG
2) Using imagenet weights
3) Image Size = 224x224
4) Train set : rescale = 1./255, shear_range = 0.2, zoom_range = 0.2, horizontal_flip = true
   Test set: rescale = 1./255
5) Parameters:
   Total params: 58,933,766
   Trainable params: 602,118
   Non-trainable params: 58,331,648
6) Uses a batch size = 32, Classification mode = categorical
7) Loss method = categorical_crossentropy
8) Optimizer method = adam
9) Epochs = 20
10) Accuracy = 94%

Using Online Dataset

CNN:

1) Epochs = 20
2) Parameters:
   Total params: 138,56,4
   Trainable params: 138,56,4
   Non-trainable params: 0
3) Accuracy = 96%

InceptionV3:

1) Epoch = 20
2) Parameters:
   Total params: 22,109,990
   Trainable params: 307,206
   Non-trainable params: 21,802,784
3) Accuracy = 97%

ResNet152V2:

1) Epochs = 20
2) Parameters:
   Total params: 58,933,766
   Trainable params: 602,118
   Non-trainable params: 58,331,648
3) Accuracy = 96%

CHAPTER 7: CONCLUSION and FUTURE SCOPE

CONCLUSION

In this study, five main chilli leaf diseases i.e. Alternaria Leaf Spot, Generic Leaf Spot, Cercospora leaf spot, Yellow Mosaic, and Fusarium Wilt leaf disease were identified along with a healthy leaf class altogether labelled as six classes. The self-built chilli leaf dataset was a balanced dataset, which means all classes had an equal number of samples, i.e. around 150 images per class and around 900 total images.

The key conclusions drawn from our project are as follows:

Dataset Availability and Quality: The availability of a comprehensive and high-quality dataset plays a critical role in training accurate deep learning models for chili leaf disease identification. However, we found that there is a scarcity of publicly available datasets specifically designed for this task. Furthermore, the quality and diversity of the existing datasets vary significantly, which affects the performance and generalizability of the trained models.

Annotation and Labeling: Accurate annotation and labeling of the chili leaf images are essential for effective training. However, manual annotation of large-scale datasets can be time-consuming, tedious, and subjective. The development of efficient annotation tools or exploring semi-supervised and weakly supervised learning approaches could address this challenge.

Class Imbalance: Imbalanced class distribution, where some chili leaf diseases are more prevalent than others, can lead to biased models. There is a need for addressing class imbalance issues through techniques such as data augmentation, oversampling, or cost-sensitive learning to ensure that the trained models are capable of accurately identifying all types of chili leaf diseases.

Model Selection and Hyperparameter Tuning: Choosing an appropriate deep learning architecture and optimizing hyperparameters is crucial for achieving high accuracy. We should emphasize on the importance of considering the specific characteristics of chili leaf diseases, such as shape, texture, and color variations, when selecting the model architecture.

FUTURE SCOPE

1. Large-scale and Diverse Datasets: Efforts should be made to create large-scale and diverse datasets that cover a wide range of chili leaf diseases, including different stages of disease progression and variations due to factors like lighting conditions, background, and growth stage. Such datasets would enable more robust and generalized models.

2. Transfer Learning and Pretrained Models: Leveraging transfer learning techniques by utilizing pretrained models trained on large-scale image datasets (e.g., ImageNet) can enhance the performance of chili leaf disease identification models.

3. Ensemble and Hybrid Models: We can investigate ensemble learning techniques and hybrid models that combine deep learning with traditional machine learning algorithms. Ensemble methods can improve model accuracy by aggregating predictions from multiple models, while hybrid models can leverage the strengths of both deep learning and conventional feature engineering approaches.

4. Explainability and Interpretability: As deep learning models are often considered black boxes, providing explanations and interpretability becomes crucial, especially in critical applications such as disease identification. One can explore techniques like attention mechanisms, saliency maps, and feature visualization to improve the interpretability of the trained models.

5. Real-Time and Edge Computing: The deployment of chili leaf disease identification models on edge devices can enable real-time monitoring and early detection of diseases in agricultural fields. Future research can be carried out to focus on developing lightweight deep learning models that can run efficiently on resource-constrained devices.

Overall, this survey provides valuable insights into the training issues in chili leaf disease identification using deep learning techniques and offers directions for future research to address these.

REFERENCES

[1] Kantha Raju Kanaparthi , S.Sudhakar Ilango (2023). "*A Survey on Training Issues in Chili Leaf Diseases Identification Using Deep Learning Techniques*".-doi: 10.1016/j.procs.2023.01.188.

[2] Sufola Araujo, Virendra S Malemath., Meenakshi Sundaram Karuppaswamy, February (2021). "*Automated Disease Identification in Chilli Leaves Using FCM and PSO Techniques*".- https://doi.org/10.1007/978-981-16-0493-5_19

[3] Dyah K. Agustika , Ixora Mercuriani , Chandra W. Purnomo , Sedyo Hartono , Kuwat Triyana ,Doina D. Iliescu , Mark S. Leeson (2022). "*Fourier transform infrared spectrum pre-processing technique selection for detecting PYLCV-infected chilli plants*". - https://doi.org/10.1016/j.saa.2022.121339

[4] Yuslena Sari , Andreyan Rizky Baskara , Rika Wahyuni(2022)."*Classification of Chili Leaf Disease Using the GrayLevel Co-occurrence Matrix (GLCM) and the Support Vector Machine (SVM) Methods*". - doi : 10.1109/ICIC54025 2021.9632920

[5] B. Nageswararao Naik , R. Malmathanraj , P. Palanisamy(2022). "*Detection and classification of chilli leaf disease using a squeeze-and-excitation-based CNN model*".- https://doi.org/10.1016/j.ecoinf.2022.101663

[6] Jagadeesh D.Pujari , Rajesh Yakkundimath ,Abdulmunaf S. Byadagi(2015). *"Image Processing Based Detection of Fungal Diseases in Plants"*. - (http://creative commons.org/licenses/by-nc-nd/4.0/).

[7] Yeong Hyeon Gu , Helin Yin , Dong Jin , Jong-Han Park and Seong Joon Yoo(2021). "*Image-Based Hot Pepper Disease and Pest Diagnosis Using Transfer Learning and Fine Tuning*". - doi: 10.3389/fpls.2021.724487

[8] Sufola Das Chagas Silva E Araujo , V.S.Malemathh , K. Meenakshi Sundaram(2022*). "Affine Invariant approach for disease detection on chili plant"*. - DOI: 10.1109/ICECET52533:2021.9698472

[9] Dipak P. Patil , Swapnil R. Kurkute , Pallavi S. Sonar , Svetlin I. Antonov(2017). "*An Advanced Method for Chilli Plant Disease Detection Using Image Processing*".

[10] Rosalina , Genta Sahuri(2020). "*Implementation of Deep Learning Methods in Detecting Disease on Chili Leaf*". WWW.NEW.IJASCSE.ORG

[11] Ian Faizal Idenugraha , Diana Rahmawati , Kunto Aji Wibisono , Miftachul Ulum (2020). "*Automatic Pesticide Spray Based on Digital Image Processing in Chili Plants by Classification Method*". - 10.21070/jeeeu.v4i1.317

[12] Mr. I. Rama Koteswara Rao(2017). "*Chilli Leaf Diseases Detection And Classification Using Various Algorithms*". ISSN : 0731-6755

[13] Sufola Das Chagas Silva Araujo , V. S. Malemath and K. Meenakshi Sundaram(2021). "*Symptom-Based Identification of G-4 Chili Leaf Diseases Based on Rotation Invariant*". - doi: 10.3389/frobt.2021.650134

[14] Zulkifli Bin Husin , Abdul Hallis Bin Abdul Aziz , Ali Yeon Bin Md Shakaff , Rohani Binti S Mohamed Farook(2012). "*Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques"*. DOI 10.1109/ISMS.2012.33

[15] Abdul Hafiz Bin Abdul Wahab , Rahimi Zahari , Tiong Hoo Lim(2019). "*Detecting diseases in Chilli Plants Using K-Means Segmented Support Vector Machine"*. 978-1-7281-3663-9/19

[16] Sufola Das Chagas Silva Araujo, V S Malemath, K. Meenakshi Sundaram(2019). "*Disease Identification in Chilli Leaves using Machine Learning Techniques"*. - DOI:10.35940/ijeat.A1061.1291S319

[17] M. Karuna , B. S Varsh , Sneha. R. M , G. K. Meghana(2019). "*Early Detection of Chili Plant Leaf Diseases using Machine Learning"*. - http://ijesc.org/