Survey of Methods to Detect Tobacco Leaf Disease

Khushi H P[1], Praveen H M[2], Pragna C P[3], Giridhar K[4], B Uma[5]

[1,2,3,4,5]*Department of Computer Science and Engineering, Malnad College of Engineering, Hassan*

---***---

Abstract

Tobacco is a key commercial crop in karnataka and several parts of india, but its cultivation is frequently hindered by various foliar diseases that reduce yield and quality. Traditional disease detection methods are manual, time-consuming, and require expert knowledge. Recent advances in machine learning (ml) have demonstrated significant potential in automating plant disease classification. This paper reviews ml-based approaches for tobacco leaf disease detection, covering existing datasets, models, performance factors, practical challenges, and future research opportunities.

Keywords— Tobacco Leaf Disease, Plant Disease Detection, Image Classification, Feature Extraction, Classification Algorithms, Infrared Spectroscopy, k-Nearest Neighbors

I. INTRODUCTION

Tobacco is a major high-value commercial crop cultivated extensively in karnataka and several other regions of india. It plays a significant role in the agricultural economy and supports the livelihood of a large number of farmers. However, tobacco cultivation is highly susceptible to a range of foliar diseases that can severely affect both yield and quality. If not identified and managed promptly, these diseases may lead to substantial economic losses.

The common tobacco leaf diseases are:

- **Tobacco Mosaic Virus (TMV):** A viral disease transmitted through contaminated tools, human contact, and insect vectors. It causes mottling of leaves, stunted plant growth, and poor-quality yield.
- **Frog Eye Leaf Spot:** Caused by the fungus *Cercospora nicotianae*, this disease presents as circular brown lesions on leaves and thrives in warm, humid conditions.
- **Powdery Mildew:** It is caused by the fungus Erysiphe cichoracearum and manifests as white, powdery patches on the leaf surface, usually in cool, dry environments.
- **Black Shank:** A soil-borne disease caused by *Phytophthora nicotianae* that affects the roots and lower stem, leading to plant wilting and eventual death.

Both biotic (such as organisms, infections, and microscopic organisms) and abiotic (such as an inadequate water system, unhealthy soil, and climate pressure) factors can cause these diseases. Natural elements like temperature, humidity, and trim administration techniques have a significant influence on their event.The conclusion of an illness has traditionally relied on manual review by qualified specialists. Although effective, this approach is subjective, time-consuming, and frequently irrational in areas with limited resources or distance. Additionally, many farmers lack the opportunity to reach rural masters, which postpones appropriate intervention.Traditionally, disease diagnosis has relied on manual inspection by trained experts. While effective, this method is time-consuming, subjective, and often impractical in remote or resource-limited areas. Additionally, many farmers do not have timely access to agricultural specialists, which delays appropriate intervention.

Recent technological developments, especially in the area of machine learning (ML), have created new avenues for automating the detection of plant diseases. A class of deep learning models called Convolutional Neural Networks (CNNs) has demonstrated great promise in identifying disease types from leaf image analysis by examining patterns and symptoms that are readily apparent. Through mobile applications, these models can provide scalable, real-time, and accurate disease diagnosis that is even accessible to non-experts.

This survey paper investigates the application of ML—especially deep learning—for tobacco leaf disease detection. It reviews existing datasets, model architectures, performance evaluation metrics, and the practical deployment of such systems. The paper also incorporates insights from agricultural experts and discusses the challenges, limitations, and future directions in this critical area of precision agriculture. Through a systematic review of various ML models, including CNNs and traditional classifiers, this study highlights the methodological approaches, data requirements, and research gaps, contributing toward more efficient and sustainable tobacco cultivation.

II. LITERATURE SURVEY

Over the past decade, the integration of machine learning (ML) into agriculture—particularly for plant disease detection—has seen remarkable growth. With growing concerns around food security, crop health, and sustainable farming practices, researchers have increasingly turned to ML models to automate the detection and classification of plant diseases. This section reviews key contributions that are directly or indirectly relevant to tobacco leaf disease detection, focusing on ML techniques, datasets, implementation strategies, and practical limitations.

A. Machine Learning for Leaf Disease Classification

In the research paper "Machine Learning for Leaf Disease Classification: Data, Techniques and Applications" by Yao et al. (2023) [1], the researchers offer an overview of machine learning approaches applied to leaf disease classification problems. The research classifies techniques into two broad categories: Conventional ML algorithms—like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees—and more complex deep learning methods, specifically Convolutional Neural Networks (CNNs).

Conventional models work well on small, neatly organized datasets but struggle when handling high-dimensional image data. Conventional models are outperformed by CNNs, on the other hand, in image-based tasks because CNNs have the capacity to learn spatial patterns and features automatically from input images. The authors infer that CNNs outperform conventional models greatly in terms of accuracy, generalization, and robustness—particularly with large, diverse, and well-annotated datasets. This research emphasizes the reason why deep learning has emerged as the go-to method for plant disease classification.

B. Tobacco Disease Detection Using CNN

The paper *"Tobacco Plant Disease Detection and Classification Using Deep Convolutional Neural Networks"* by Krishna et al. (2022) [2] discusses tobacco leaf diseases in particular. The authors created a dataset of images representing prevalent tobacco diseases, such as Tobacco Mosaic Virus (TMV), Frog Eye Leaf Spot, and Powdery Mildew. The dataset was trained and tested with a custom CNN architecture that resulted in a high classification accuracy rate of 89.71%.

In addition to complications like changing lighting, shadow, and background noise, the model differentiated correctly between healthy and infected leaves. The research also suggested the application of the model in real life through mobile or smart farming platforms, providing immediate diagnosis assistance to farmers. This is an important milestone towards developing useful, crop-specific disease diagnostic tools through deep learning.

C. Mobile-Based Plant Disease Detection Systems

The mobile-based ML solution potential is also investigated in *"Deep Convolutional Neural Networks for Mobile Capture Device-Based Crop Disease Classification in the Wild"* by Picon et al. (2019) [3]. In this research, the researchers created lightweight CNN models trained for mobile hardware so that farmers can take photos of leaves and get real-time predictions on the disease directly on their mobile phones.

These systems were able to operate even in low-connectivity settings, with others engineered to operate offline. The research proved that high performance was attainable without heavy computational needs. For tobacco-producing areas where access to agricultural specialists is limited, such a system offers a scalable and accessible solution to monitoring plant health.

D. Key points on Data Challenges in Agriculture

Even though there's been a lot of progress in developing and applying machine learning models in agriculture, one of the biggest challenges still lies in the availability of good-quality datasets—especially when it comes to detecting tobacco leaf diseases. In the paper *"Deep Learning Models for Plant Disease Detection and Diagnosis"*, Ferentinos (2018) [4] points out that deep learning models really depend on large, well-labeled datasets that are specific to each crop. Datasets like PlantVillage have helped a lot in advancing disease detection for crops like tomato, potato, and maize. But when it comes to tobacco, similar public datasets are either very limited or simply don't exist. This makes it harder for researchers to build and test accurate models that work well in real farming conditions.

Consequently, scientists have to develop personalized datasets by manual annotation and fieldwork, which is time-consuming and labor-intensive. This creates problems like class imbalance, variability in the environment, and data collection inconsistency. Ferentinos also emphasizes the need for data augmentation (e.g., scaling, rotation, flipping), transfer learning from pre-trained models, and the utilization of synthetic data to enhance model generalization and training efficiency. These techniques are imperative to overcome data constraints and allow for sound disease detection models under different field conditions.

IV. PROPOSED METHODOLOGY

The process of creating a machine learning model for detecting tobacco leaf diseases usually involves some major stages, from data collection to model deployment in real-world scenarios.

A. Dataset Collection

The first step is dataset collection, which involves gathering images of both healthy and diseased tobacco leaves. These images may be sourced from publicly available datasets such as plant village or collected manually through field surveys. When collected from the field, images are often labeled by agricultural experts to ensure accuracy. Field-based data collection reflects real agricultural conditions, including variations in lighting, background clutter, leaf orientation, and disease severity—factors essential for training robust models.

B. Data Preprocessing

Once the dataset is compiled, it undergoes preprocessing to enhance the quality and consistency of the data. All images are resized to a uniform resolution to standardize model input. Pixel values are normalized, typically scaled between 0 and 1, which helps accelerate model training and improves convergence. To address dataset limitations and improve generalization, data augmentation techniques—such as flipping, rotation, cropping, and zooming—are applied. These methods efficiently increase the training data and enable the model to become more robust to image appearance variations.

C. Model Design and Architecture Choice

After preprocessing, the subsequent step is choosing a suitable model architecture. In image-based classification problems, Convolutional Neural Networks (CNNs) are most commonly utilized because they are capable of learning spatial and texture-based features from images. CNNs have stacked layers of convolutions, pooling, and fully connected layers and can therefore learn intricate patterns related to various leaf diseases.

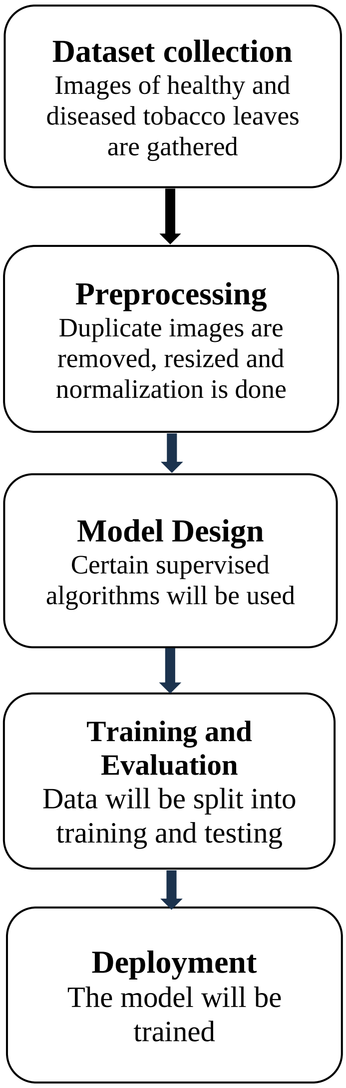
D. Model Training and Evaluation

Once the architecture is selected, the model is trained on the labeled dataset. Typically, the data is split into training and testing sets. The training set is used to teach the model how to classify images, while the testing set evaluates its performance on previously unseen data.

E. Model Deployment

The final phase is model deployment where the trained model is integrated into a functional system. The final output of the model could be deployed as a web application, a mobile app, or even be embedded in hardware devices like drones or handheld tools, depending on the purpose for which it was intended.

For mobile apps, the user is simply able to take a picture of a tobacco leaf with their phone's camera, and it will label it as healthy or infected in real-time. This renders the technology very useful for farmers, allowing field diagnosis without having to go through specialist consulting—dramatically increasing accessibility and disease control in rural regions.



V. EXPECTED RESULTS

As we reviewed various research papers, we came across several studies that explored how machine learning and deep learning models can be used for detecting plant diseases, including those that specifically focus on tobacco leaves. One of the paper used a CNN-based model for tobacco disease classification and achieved an accuracy of 89.71%

In addition to this several papers were based on traditional machine learning models like SVM, KNN, and Random Forest. These are simpler and require manual feature extraction (like measuring color or texture), and they have lower accuracy around 70% to 80%. Even though they are easier to implement and don't need as much computing power, they're not as good at handling complex image data compared to deep learning models.

This survey has helped us understand what's already done and what works well, so the plan is to select the more efficient algorithm and then train the model. Main goal is to create the dataset by our own and provide the high quality and well labelled images. Giving more importance to the evaluation step may improve the accuracy, so the proper evaluation metrices like F1 score, recall etc will be implemented. Like this reviewing of several published paper has helped us to understand topic well, so that the accuracy of our project can be increased.

VII. CONCLUSION

This survey reviewed a range of machine learning methods for identifying diseases in tobacco leaves. Deep learning models, especially Convolutional Neural Networks (CNNs), are shown to be a preferable method for correctly identifying diseases including Tobacco Mosaic Virus, Frog Eye Leaf Spot, and Powdery Mildew. These methods also represent improved accuracy through automation and scalability compared with classical approaches.

A future project will concentrate on creating a machine learning model that can identify tobacco leaf diseases from picture data, based on the results of the literature. The objective is to create a system that can help farmers by offering accurate and timely diagnostics. The survey's findings will be useful in identifying suitable datasets, picking efficient model architectures, and comprehending possible implementation difficulties.

VIII. REFERENCES

[1] J. Yao, S. N. Tran, S. Sawyer, and S. Garg, "Machine Learning for Leaf Disease Classification: Data, Techniques and Applications," arXiv preprint arXiv:2310.12509, 2023.

[2] T. S. R. Krishna, S. M. S. Hussain, M. R. Mohan, and K. S. Venugopal, "Tobacco Plant Disease Detection and Classification Using Deep Convolutional Neural Networks," Computers and Electronics in Agriculture, vol. 200, p. 105123, 2022.

[3] A. Picon, A. Alvarez-Gila, M. Seitz, A. M. Ortiz-Barredo, and A. Echazarra, "Deep Convolutional Neural Networks for Mobile Capture Device-Based Crop Disease Classification in the Wild," Computers and Electronics in Agriculture, vol. 161, pp. 280–290, 2019.

[4] G. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[5] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 2016.

[6] H. A. Khan, N. Yousaf, M. Hussain, and A. Mahmood, "Plant Disease Detection Using Machine Learning: A Review," Computer and Electronics in Agriculture, vol. 186, p. 106125, 2021.

[7] J. Chen, Y. Chen, and X. Zhang, “Crop Leaf Disease Detection Using Deep Learning with Data Augmentation,” IEEE Access, vol. 8, pp. 53784–53792, 2020.