International Journal of Computer Applications (0975 – 8887)
Volume 187 – No.16, June 2025

Cotton Disease Detection using Machine Learning Techniques for Crop Health and Yield: A Study

Arwadeep Gupta
M. Tech. Scholar
Dept. of CSE
LNCT, Bhopal

Vivek Rechhariya, PhD
Professor
Dept. of CSE
LNCT, Bhopal

ABSTRACT
Cotton is one of the most important cash crops globally, contributing significantly to the agricultural economy and textile industries. However, cotton production is often affected by various diseases such as bacterial blight, leaf curl virus, and fungal infections, which lead to substantial yield losses and reduced fiber quality. Traditionally, disease detection in cotton relies on manual observation and expert knowledge, which is time-consuming, labor-intensive, and prone to human error. Cotton is a vital cash crop whose productivity is significantly affected by various diseases. Early and accurate detection of these diseases is essential to prevent crop loss and improve yield. This study explores the application of machine learning techniques for detecting cotton leaf diseases using image processing and classification models. Various algorithms, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forests, are evaluated for their effectiveness in identifying common cotton diseases. The study aims to assist farmers and agronomists in disease management, promoting healthier crops and improved yield through technology-driven solutions.

Keywords
Cotton Disease, Machine Learning, Image Processing

1. INTRODUCTION
Most agricultural countries, such as India, rely heavily on agriculture for their finances. Cotton is primarily grown in northern India [1]. Several diseases have been negatively affecting cotton plant productivity in recent years. Infection and preparation hospitalization is the most important requirement. To reduce the loss of on cotton plant leaves, the plant disease must be accurately diagnosed, and should be treated as soon as possible [2]. It is sufficient knowledge that farmers identify illnesses too late, making treatment extremely difficult and expensive. The shortage of experts near is the main cause due to the difficulty of that produces a considerable number of experts across the country. As a result, farmers can benefit from the availability of computer-aided professional programs. The purpose of this study is to develop a system that allows farmers to acquire knowledge on cotton plants by uploading images to the central authorities with specialized expertise.

The proposed system provides a practical solution for the challenge of farmers being exposed to leaf disease identification and treatment, thus taking appropriate measurements of and undermining the adverse effects of the disease on the yield and quality of crops [3]. In this way, the farmer receives diagnostic techniques from the human expert. To ensure that the photo is able to recognize diseases in a sufficient amount of, computer scientists use this information to generate a training set that is applied to the image. Pattern adaptation algorithms quickly identify diseases and identify diseases at their early stages. Capture of diseases by is a test of its physical properties called a diagnosis of cotton disease. Characteristic extraction removes the identifiable structure from the digital image [4]. This process can use several functional descriptors.

2. IMAGE PROCESSING
In Image processing section, initially the image is captured from the camera and further the image is processed using k means clustering for segmenting the image. The processed image is then edge detected using three different edge detection techniques. The edge detection techniques used are sobel, prewitt and canny algorithm. The diseased sample banana leaf has been taken for the edge detection analysis. Amongst the three edges detection methods used, canny edge detection algorithm gives the better and reliable detection. Owing to its optimality to meet with the three criteria for edge detection and the simplicity of process for implementation, it became one of the most popular algorithms for edge detection method.

As discussed earlier, IoT and Image processing are combined together in agricultural field in order to increase product yield and to reduce the crop failure. We focused on plant failure due to environmental factors through IoT technology. IoT system includes sensors, Arduino and a camera that regularly captures the plant.

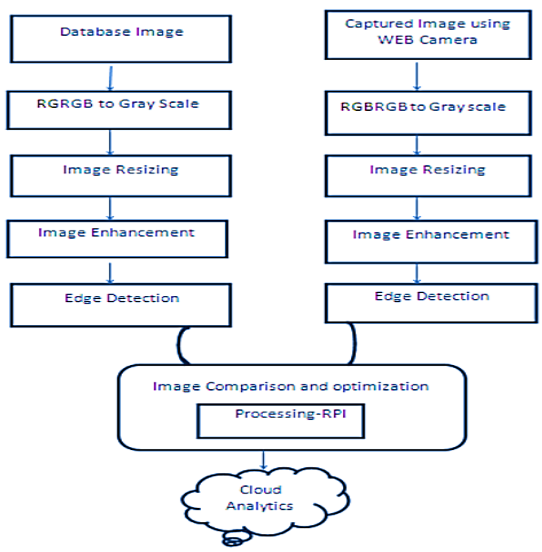

Fig 1: Block diagram of Image Processing

The color, texture, shape and area of the leaf are the parameters also considered in this work. After examine the conditions of the plants we go for image processing. The initial test is done by using MATLAB software. In addition to the environmental factors, the plant with a diseased leaf can also be identified using Image processing. Based on the output and constraints the pesticides will be sprayed for the crop/plant where the disease is identified. If there is any change that corresponds to the deterioration in the plants growth, the farmer is immediately informed. Early diagnosis will thus help in taking the necessary actions to increase the produce and reduce failure of crops.

3. LITERATURE REVIEW

R. Kumar et al. [1], the presence of cotton plant wells is extremely important for maintaining agricultural productivity, and early detection of diseases plays a critical role in achieving this goal. This study presents a comprehensive approach to creating machine learning systems that can identify diseases in cotton systems by analyzing leaf images. The survey includes stages such as obtaining data records, preprocessing data, training models, developing ensemble models, assessing models, and analyzing results. Several machine learning models are trained and evaluated to determine whether cotton leaves can be classified as "healthy or "disease." These models include Random Forest, Support Vector Machines (SVMs), Multi-Class SVMs, and Ensemble models. This test provides a practical and visually beneficial system for recognizing diseases that may contribute to disease prevention, thus improving both yield and quality. This work highlights the importance of continuous improvement by periodically updating the model and examining the possibilities of advanced technologies such as deep learning.

I. Ahmed et al. [2], agricultural harvest production is possible through photographs of plant leaves and automated identification of agricultural harvest diseases and unhealthy regions, and advances in technology that increases production. Machine learning is used to identify agricultural harvest diseases and image processing techniques such as image intake, preprocessing, segmentation, characterization extraction, support vector machines (SVMs), folding networks (CNNs), K-nearest neightre, and Gray-level co-ocular matrix (GLCMs). This method uses photographs to recognize plant diseases. GLCM extracts structural properties from photographs. k-mean clustering segment input image. The SVM classifier improves the performance of existing algorithms by splitting the input image. The approach is to eliminate human participation and improve farmer procurement. Farmers rarely produce high harvests or reduce their income. Nutrient deficiency, soil moisture, temperature fluctuations, etc. Because of this, this analytical approach can use different techniques to improve the accuracy of the victim. During this evaluation, we will develop predictive tools for plant diseases and propose new ways to become uncomfortable. The system provides information about certified users.

S. Bondre et al. [3], awareness of agricultural diseases plays an important role in the world of intellectual agriculture. For agricultural production to increase sustainably, identifying plant diseases must be efficient. Plant abnormalities affected by viruses, insects, malnutrition and lack of bad weather have traditionally been diagnosed by human experts. Two highly developed technologies for machine learning synonymous with deep learning (DL) and transfer learning, further developments of artificial intelligence technology in recent years, have been used to identify agricultural diseases. However, there are various obstacles to the widespread use of these approaches. This paper specifically examines DL and transfer learning and discusses recent developments using these advanced technologies to recognize agricultural disease photos. Many DL architectures have recently been adopted along with visualization tools. This is extremely important for identifying labels and classifying plant diseases. The analysis and evaluation of these two approaches indicate that transfer learning is a better choice when data is available for agricultural diseases. Next, this paper examines the most important issues that must be addressed for research in this field in order to advance. B. Optimizing image data records creation, collection of big data auxiliary domains, and transfer learning. Building a technically feasible image identification system for agricultural diseases involves the creation of image data sets collected under actual cultivation conditions. The purpose of this study is to learn more about DL skills in plant disease detection to increase the efficiency and accuracy of devices in future research.

M. M. Islam et al. [4], cotton is known in the agricultural industry as "white gold." Agriculture is the main source of Bangladesh's economic revenue, and the country's economy is heavily dependent on agriculture. Our country's soil and water resources are fertile and the climate is mild. However, many diseases affect plant production, causing major harvest losses, which put the lives of helpless farmers at risk. Previous reports have shown that approximately 70-80% of cotton diseases are leaf disease and 30-20% pest disease. Experts usually use the naked eye to identify and identify plant diseases and pests that can result from poor identification accuracy. As a result, early detection of cotton disease using an AI-based system can help increase cotton production by significantly demonstrating leaf disease. In this study, we proposed a DL-based cotton leaf disease approach with detection of fine-tuning transfer learning (TL) by setting layers and parameters of existing TL algorithms. We also examined the performance of several fine-tuning models such as VGG-16, VGG-19, Inception-V3, and Xception of cotton data records that have been published to predict cotton disease. The study showed that the Xception model yielded the highest accuracy of 98.70% and was chosen to increase cotton production to develop web-based intelligent applications for the prediction of cotton disease in agriculture. Therefore, our model can diagnose the accurate diagnosis of cotton disease and provide a new window for automatic leaf disease diagnosis of other plants.

R. Mahum et al. [5], potato disease management plays a valuable role in the agricultural sector. This is because it can lead to significant losses in plant production. Therefore, timely detection and classification of potato disease is required to minimize losses. However, it is time to take on the task and demand human effort. Therefore, accurate automated technology is required for timely detection and classification to address the above challenges. There are methods based on machine learning and deep learning methods that use existing data records. Run the H. Village data record and run only two classes with potato leaves. A highly efficient Dennenet model was used along with the additional transition layer of Dersenet-2010 to efficiently classify potato leaf diseases. Furthermore, because the training data is very unbalanced, the proposed algorithm becomes lobster when using the newly weighted cross-entry loss function. The tight connection with the power of regularization contributes to minimizing overwork during nighttime during training with small training rates of potato leaves. The proposed algorithm is a new first technique for addressing and reporting successful implementation of recognition and classification of four diseases in potato leaves. The performance of the algorithm was evaluated on a test set, resulting in an accuracy of 97.2%. Various experiments were conducted to confirm that our proposed algorithm is more consistent and capable of recognizing potato blade disease and classifying it as an existing model.

L. Goel et al. [6], agriculture is one of the most important sectors and meets the basic food needs of humanity. Plant diseases raise concerns about economic and nutritional security of federal states and hinder agricultural planning. Traditional methods to recognize plant diseases require a lot of work and time. As a result, many researchers and institutions strive to tackle these issues with the help of sophisticated technical methods. Deep learning-based detection of plant diseases offers considerable advancement and hope compared to classical methods. When these techniques are trained on large, high-quality datasets, these techniques recognize robust diseases of plant leaves at the early stages. This study was examined in three different areas: disease classification, detection and segmentation of system leaf disease, and thoroughly checked data records that were published simultaneously. This systematic overview provides a comprehensive assessment of the current literature and describes the most popular deep learning architectures, the most frequently investigated systems, data records, challenges, and various perspectives. Provides researchers working in the agricultural sector. Furthermore, the biggest challenges in the field of disease detection in agriculture are involved. Therefore, this study provides valuable information and appropriate solutions based on deep learning applications for agricultural sustainability.

4. METHODOLOGY

In this section we will explain about the methodologies which we have used.

A **decision tree** is a graphical representation of possible solutions to a decision based on certain conditions. It's called a decision tree because it starts with a single box (or root), which then branches off into a number of solutions, just like a tree.Decision trees are helpful, not only because they are graphics that help you 'see' what you are thinking, but also because making a decision tree requires a systematic, documented thought process. Often, the biggest limitation of our decision making is that we can only select from the known alternatives. Decision trees help formalize the brainstorming process so we can identify more potential solutions.

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set.The first algorithm for random decision forests was created by using the random subspace method which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An **Artificial Neural Network**, often just called a neural network, is a mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase. Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data.The inspiration for neural networks came from examination of central nervous systems. In an artificial neural network, simple artificial nodes, called “neurons”, “neuroses”, “processing elements” or “units”, are connected together to form a network which mimics a biological neural network.

In machine learning, **support vector machines** (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970’s as a non-parametric technique [14-15].

5. CONCLUSION

The detection of cotton leaf diseases using machine learning techniques offers a promising solution for improving crop health and increasing yield. This study demonstrates that models like Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forests can effectively classify common cotton diseases from leaf images with good accuracy. By enabling early disease detection, these methods help farmers take timely action, reducing crop losses and minimizing the excessive use of pesticides.

Machine learning-based detection systems not only support precision agriculture but also promote sustainable farming practices. With further improvements and integration into mobile or IoT-based applications, these techniques can be made accessible to farmers, especially in rural areas, making disease management easier and more efficient. Future work can focus on real-time monitoring, larger datasets, and advanced deep learning models to enhance the system’s accuracy and practical usability in the agricultural sector.

6. REFERENCES

[1] R. Kumar *et al.*, "Hybrid Approach of Cotton Disease Detection for Enhanced Crop Health and Yield," in *IEEE Access*, vol. 12, pp. 132495-132507, 2024

[2] I. Ahmed, G. Habib and P. K. Yadav, "An approach to identify and classify agricultural crop diseases using machine learning and deep learning techniques", *Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI)*, pp. 1-6, Mar. 2023.

[3] S. Bondre and D. Patil, "Recent advances in agricultural disease image recognition technologies: A review", *Concurrency Computation Pract. Exper.*, vol. 35, no. 9, Apr. 2023.

[4] M. M. Islam, M. A. Talukder, M. R. A. Sarker, M. A. Uddin, A. Akhter, S. Sharmin, et al., "A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture", *Intell. Syst. Appl.*, vol. 20, Nov. 2023.

[5] R. Mahum, H. Munir, Z.-U.-N. Mughal, M. Awais, F. S. Khan, M. Saqlain, et al., "A novel framework for potato leaf disease detection using an efficient deep learning model", *Hum. Ecol. Risk Assessment Int. J.*, vol. 29, no. 2, pp. 303-326, Feb. 2023.

[6] L. Goel and J. Nagpal, "A systematic review of recent machine learning techniques for plant disease identification and classification", *IETE Tech. Rev.*, vol. 40, no. 3, pp. 423-439, May 2023.

[7] C. Sarkar, D. Gupta, U. Gupta and B. B. Hazarika, "Leaf disease detection using machine learning and deep

learning: Review and challenges", *Appl. Soft Comput.*, vol. 145, Sep. 2023.

[8] J. Karthika, M. Santhose and T. Sharan, "Disease detection in cotton leaf spot using image processing", *J. Phys. Conf. Ser.*, vol. 1916, no. 1, 2021.

[9] Y. Yuan, L. Chen, H.Wu, and L. Li, ‘‘Advanced agricultural disease image recognition technologies: A review,’’ Inf. Process. Agricult., vol. 9, no. 1, pp. 48–59, Mar. 2022.

[10] V. Kathole and M. Munot, ‘‘Deep learning models for tomato plant disease detection,’’ in Advanced Machine Intelligence and Signal Processing. Singapore: Springer, 2022, pp. 679–686.

[11] Q. Pan, J.-F. Qiao, R. Wang, H.-L. Yu, C. Wang, K. Taylor, and H.-Y. Pan, ‘‘Intelligent diagnosis of northern corn leaf blight with deep learning model,’’ J. Integrative Agricult., vol. 21, no. 4, pp. 1094–1105,Apr. 2022. V. Petsiuk. (2019).

[12] U. Chauhan, ‘‘ResTS: Residual deep interpretable architecture for plant disease detection,’’ Inf. Process. Agricult., vol. 9, no. 2, pp. 212–223, Jun. 2022. M. Turkoglu, B. Yanikoğlu, and D. Hanbay, ‘‘Plantdiseasenet: Convolutional neural network ensemble for plant disease and pest detection,’’ Signal, Image Video Process., vol. 16, no. 2, pp. 301–309, 2022.

[13] Vallabhajosyula, V. Sistla, and V. K. K. Kolli, ‘‘Transfer learning-based deep ensemble neural network for plant leaf disease detection,’’ J. Plant Diseases Protection, vol. 129, no. 3, pp. 545–558, Jun. 2022.

[14] Chen, Z. Yuan, S. Chen, and X. Zou, ‘‘Plant disease recognition model based on improved YOLOv5,’’ Agronomy, vol. 12, no. 2, p. 365, Jan. 2022. A. M. Roy, R. Bose, and J. Bhaduri, ‘‘A fast accurate fine-grain object detection model based on YOLOv4 deep neural network,’’ Neural Comput. Appl., vol. 34, no. 5, pp. 3895–3921, Mar. 2022.

[15] Wang, X. He, K. Feng, and H. Zhu, ‘‘Plant disease detection and classification method based on the optimized lightweight YOLOv5 model,’’ Agriculture, vol. 12, no. 7, p. 931, Jun. 2022.