



Review of Identifying and preventing cotton diseases

Prof.Ms.B.K.SWETA ,M.Sc, S.Rajasurya ¹ ,B Kaleeswaran².

¹Assistant Professor, Department of Computer Applications, Sri Krishna Arts and Science College, Coimbatore

²Students III BCA-B., Department of Computer Applications,Sri Krishna Arts and Science College, Coimbatore

Abstract—This India, with its vast agricultural sector, is a significant player in global crop production, particularly in cotton cultivation. Despite its extensive cultivation area of approximately 350,000 hectares and a production volume of around 5.3 million tons, the sector faces substantial challenges due to crop diseases and insect pests. Blights, affecting 10-30% of the total crop, are a major concern, necessitating accurate and efficient identification methods to prevent significant yield losses. Cotton, often referred to as “white gold,” is economically vital, with its global trade value projected to rise from 40 billion to 60 billion US dollars by 2030. However, diseases affecting cotton leaves have reduced production by about 25% in India. Insect pests such as *Aphis gossypii*, Thrips, and Cotton Bollworm further exacerbate the problem, causing significant damage and necessitating effective pest management strategies. This review highlights the importance of developing advanced monitoring and prediction systems to mitigate these challenges and enhance cotton production sustainability.

Keywords—IPPS(Insect pest Predict System), SVM(Support Vectot Machine), MLP(Multilayer Perceptron) KNN (K-Nearest neighbor)Cotton

INTRODUCTION

[1]India is a country with a majority of the population relying heavily on the agricultural sector. The crop cultivation area .According to the United States Department of Agriculture (USDA), Bangladesh is the world's second-largest cotton importer and fourth-largest cotton consumer. In 2021, Bangladesh spent 3 billion US dollars importing 8.5 million bales of cotton fiber in order to meet the country's rising demand in the textile industry.

India spans around 3,50,000 hectares approximately and the production quantities roughly sum up to 53,00,000 tons, making India the third largest agricultural producer in the world. The sensitivity of crops coupled with climatic conditions have made blights common in the crop during all the stages of its growth. Blight affected leaves constitute 10-30% of the total crop loss. Identification of such blights in the leaf is very important in preventing any heavy losses in yield as well as the quantity of the

Figure 5

agricultural product. Monitoring the leaf blights manually is a difficult task due to its complex nature and is a time consuming process. Therefore, there is a need to reduce the manual effort put into this task,

while making accurate predictions and ensuring that the farmer's lives are hassle free. Visually observable patterns are difficult To decipher at a single glance, leading to many farmers making inaccurate assumptions regarding the blight.Cotton is one of the most economical crops, often known as “white gold” or the “king of fibers” (Khairnar & Goje, 2020).The global trade value of cotton is 40 billion US dollars, but by the end of 2030, that value is expected to increase to 60 billion US dollars (Meyer et al., 2023).Around 80% to 90% of the disease that affects cotton plants are found on the leaf which has reduced cotton production by around 25% in India (Khairnar & Goje, 2020)About 15–20% of plants can be saved from death if disease symptoms.[1]

Types of insect damages in cotton

[1]Cotton is primarily damaged by insects such as *Aphis gossypii*, Thrips and Cotton Bollworm. These pests significantly reduce production and cultivation, making them major contributors to crop damage[1].

Cotton Bollworm

[2]In those areas where tobacco budworm was the primary lepidopteran pest, the bollworm (ie, primary economically important species (In those areas where tobacco budworm was the primary lepidopteran pest, the bollworm (ie, corn earworm), *Helicoverpa zea*, has become the primary economically important species (Gore et al., 2008). It is much less susceptible to the proteins expressed in Bt cotton than tobacco budworm (Luttrell et al., 1999), and supplemental foliar insecticide sprays are often needed in Bt cotton to prevent yield losses (Gore et al., 2008).[2]



Figure 1

Thrips

Thrips are tiny, light brown or black insects that feed at night. Three species of thrips are of economic importance as pests of potato. The western flower thrips (*Frankliniella occidentalis*) originated in western North America and has since become a major pest on many crops across the United States and around the world. Also, the onion thrip (*Thrips tabaci*) is thought to have originated in the Mediterranean region but is now worldwide spread, including in Africa. In addition, the cotton thrips (*Frankliniella schultzei*) is commonly found on potato in Africa (Nyasani et al., 2012; Otieno, 2019)[3]

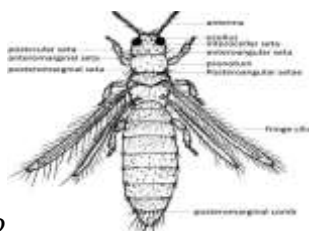


Figure 2

Literature Reviews

INSECT PEST PREDICTION SYSTEM (IPPS)

Figure 5

Insect pest monitoring is typically performed in agriculture and forestry to assess the pest status in given locations (i.e., greenhouse, field, orchard/vineyard, forest) by collecting information about the target pest presence, abundance, and distribution. Within the integrated pest management programs in agriculture, the final goal of insect pest monitoring is to provide growers with a practical decision-making tool. For instance, the intervention thresholds are essential to counteract a given insect pest infestation in a specific field in the best moment, optimizing the control strategy and the grower inputs on that crop. Monitoring data can also be used to implement prediction phenological models to forecast an insect population outbreak, providing additional information to improve the control techniques and optimize insecticide usage (Dent 2000).

Insect pests and beneficials may be monitored using a wide range of techniques, including visual inspection, suction traps and passive methods (McCravy 20)

Data analysis and data preprocessing

[1]A data set containing images of fresh and diseased cotton leaf and plants are collected from Kaggle.com as part of a competition hosted by Dev3v D3V (2022)This data set is comprised of four image types: fresh cotton leaf, fresh cotton plant, diseased cotton leaf and diseased cotton plant with a total of 2310 images. For exasmall. 2and 3 represent the fresh and affected samples of the dataset.[1]



Figure3

[1]The real image of the dataset is not always fresh and hence may contain noise in the data. Before attempting to fit them into the learning module, it is necessary to preprocess them for better accuracy and performance of the module. This refers to the process of preparing the data for further use, which may involve integrating, cleaning, normalization, or transforming the data. In our case, we apply the following types of operation -Resizing - the process of changing the size of an image while keeping the number of pixels constant for every image in the dataset.

Sharpening - the process of yanking noise and making the image sharp to get a better result. We use sharpen filter to sharpen the image.Rescaling - also known as resampling, creates a new picture with a different size, upsampling increases the size of a picture, while downsampling decreases it.

Shearing - refers to changing an image along an axis, usually to create or fix perception angles. It is usually used to improve images so computers can see how people see things from different angles.Zooming - the process of expanding an image so that its individual components become more pronounced and distinct.Horizontal-flip - also called a reversed image, is the horizontal reflection of a static or moving image [1]

SVM(SUPPORT VECTOR MACHINE)

Figure 5

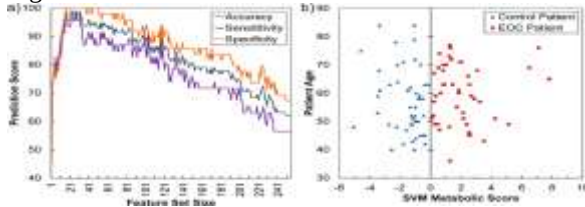
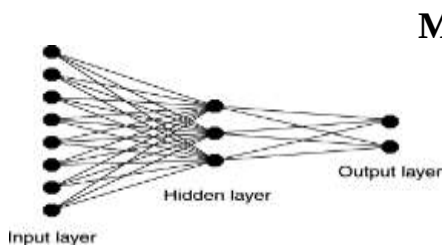


Figure 4

[5]Support Vector Machine (SVM) (Cortes & Vapnik, 1995) is a method for the classification of linear and nonlinear data, and uses nonlinear mapping to transform the original training data into a higher dimension. Support Vector Machines then search for the linear optimal separating hyperplane, which essentially is a boundary which separates the records into classes. The Support Vector Machine finds the optimal hyperplane using support vectors (which can be characterised as the most significant training records) and margins (which are defined by the support vectors) (Han et al., 2011).The Support Vector Machine can be trained using several functions: the Linear Kernel Function, Quadratic, Gaussian Radial Basis (GRB), Multilayer Perceptron Kernel (MP) functions. Fig. 4 shows an example Support Vector Machine applied to ovarian cancer prediction using a linear separation of classes (Gaul et al., 2015). The most suitable training function is often experimentally selected. Support Vector Machines, when trained using the appropriate training function, are highly accurate and capable of modeling complex nonlinear decision boundaries (Han et al., 2011). They are also much less prone to overfitting than other computational intelligence methods, and can be used for classification as well as prediction tasks. Support Vector Machines have been adopted to solve decision-making tasks on data which also include data obtained from images, microarray and other clinical data.[5] [1]SVM classifier is employed for the disease classification step which yields 94% accuracy rate.[1]



A

multilayer perceptron consists of a number of layers containing one or more neurons (see Figure 1 for an example). The role of the input neurons (input layer) is to feed input patterns into the rest of the network. After this layer, there are one or more intermediate layers of units, which are called hidden layers. Subsequently, the hidden layers are followed by a final output layer where the results of the computation are read off. Each unit is connected to all units in the subsequent layer and each unit receives input from all units in the previous layer. Each connection has a certain weight, and this weight illustrates the influence of the unit to the response of the unit in the subsequent layer. The output of a multilayer perceptron depends on the input and on the strength of the connections of the units. When information is offered to a multilayer perceptron by activating the neurons in the input layer, this information is processed layer by layer until finally the output layer is activated. Given enough hidden units and enough data, it has been shown that multilayer perceptrons can approximate virtually any function to any desired accuracy. In other words, multilayer perceptrons are universal approximators. However, these results are valid if and only if there is a sufficiently large number of training data in the series. If there are not enough data to ‘train’ the neural network, the neighbor will not be able to learn the required input–output relationship accurately. Therefore, multilayer perceptrons are valuable tools to solve complex problems when sufficient data are available to train them..[6]

Figure 5

k-nearest neighbor

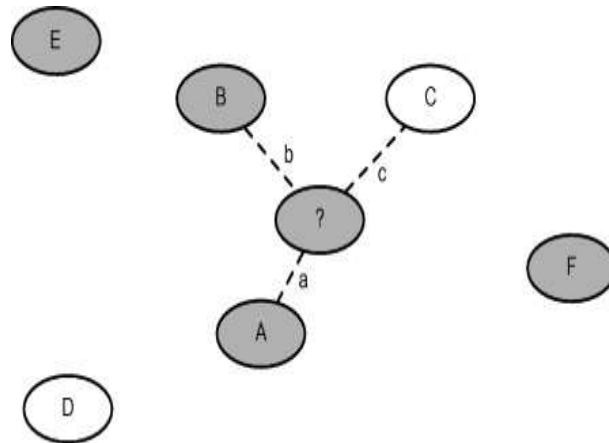


Figure 6

The first, known as k-nearest neighbor classification, involves retrieving the nearest neighboring entities to a new entity and assigning a category, or set of categories, to this new entity based on those already assigned to other entities in the space. For example, Figure 5.10 illustrates the application of the k-nearest neighbor approach (with $k = 3$) as a means to assign a category to an unseen entity (“?”) based on the categories assigned to its nearest neighbors. This is accomplished by assigning scores to the possible categories. In this case, there are only two possible categories, “gray” and “white.” The score for the “gray” category is the sum of the similarity scores assigned to all entities of this type within the neighbors retrieved, which is calculated as $a + b$, where a and b are estimates of the similarity between the new entity and entities A and B respectively. As this score is greater than that assigned to the “white” category, the category “gray” is assigned to the new entity. While this method is relatively simple, it only considers categories assigned to the closest entities in the space, and all entities in the space must be considered in order to determine which of them is closest to the new entity. Therefore, the use of reduced-dimensional geometric representations is desirable as a means to prevent this search process from being prohibitively time consuming with larger datasets. [7]

[1](KNN) classification method to detect cotton disease with an accuracy rate Of 92%.[1]

Artificial Neural Networks

[8]Artificial neural networks (sometimes just called neural networks or connectionist models) provide a means for dealing with complex pattern-oriented problems of both categorization and time-series (trend analysis) types. The nonparametric nature of neural networks enables models to be developed without having any prior knowledge of the distribution of the data population or possible interaction effects between variables as required by commonly used parametric statistical methods.

As an example, multiple regression requires that the error term of the regression equation be distributed normally (with a $\mu = 0$) and also be nonheteroscedastic. Another statistical technique that is frequently used for performing categorization is discriminant analysis, but discriminant analysis requires that the predictor variables be multivariate normally distributed. Because such assumptions are removed from ANN models, the ease of developing a domain problem solution is increased with artificial neural networks. Another factor contributing to the success of ANN applications is their ability to create nonlinear models as well as traditional linear models and, hence, artificial neural network solutions are applicable across a wider range of problem types (both linear and nonlinear).In the following sections, a brief history of artificial neural networks is presented. Next, a detailed examination of the components of

Figure 5
an artificial neural network model is given with respect to the design of artificial neural network models of business and scientific domain problems.

A Biological Basis of Artificial Neural Networks

Artificial neural networks are a technology based on studies of the brain and nervous system as depicted in Fig. 1. These networks emulate a biological neural network but they use a reduced set of concepts from biological neural systems. Specifically, ANN models simulate the electrical activity of the brain and nervous system. Processing elements (also known as either a neurode or perceptron) are connected to other processing elements. Typically the neurodes are arranged in a layer or vector, with the output of one layer serving as the input to the next layer and possibly other layers. A neurode may be connected to all or a subset of the neurodes in the subsequent layer, with these connections simulating the synaptic connections of the brain. Weighted data signals entering a neurode simulate the electrical excitation of a nerve cell and consequently the transference of information within the network or brain. The input values to a processing element, in, are multiplied by a connection weight, $w_{n,m}$, that simulates the strengthening of neural pathways in the brain. It is through the adjustment of the connection strengths or weights that learning is emulated in ANNs[8]

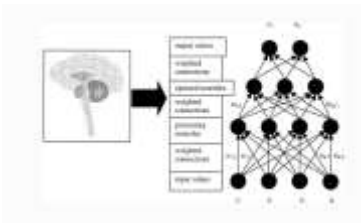


Figure 7

SVM	94%
KNN	92%
MLP	96.9%
ANN	almost 99%

Table 1

Conclusion

The agricultural sector in India, particularly cotton cultivation, is crucial for the country’s economy and global trade. However, the prevalence of blights and insect pests poses significant threats to crop yield and quality. Effective identification and management of these issues are essential to minimize losses and ensure sustainable production. The development and implementation of advanced monitoring systems, such as the Insect Pest Prediction System (IPPS), can significantly reduce manual efforts and improve accuracy in pest detection. By adopting these technologies, farmers can make informed decisions, leading to better crop management and increased productivity. Continued research and innovation in pest management strategies are vital to address the evolving challenges in cotton cultivation and secure the future of this critical agricultural sector.

Figure 5

References

- [1]"A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture":
- [2]J. Leland, J. Gore, in
Microbial Control of Insect and Mite Pe bollwarms
- [3] Joseph E. Munyaneza, Benoit Bizimungu, in
Insect Pests of Potato (Second Edition)
, 2022 thrips
- [4]Insect pests and beneficials may be monitored using a wide range of techniques, including visual inspection, suction traps and passive methods (McCravy 2018).
- [5]An SVM-based regression method is presented
Sarangdhar and Pawar (2017)
for both the detection and categorization of five different diseases
- [6]N.L.W. Keijsers, in
Encyclopedia of Movement Disorders
, 2010
- [7]Khajamoinuddin Syed, ... Preetam Ghosh, in
Artificial Intelligence in Precision Health
, 2020
- [8]Artificial Neural Networks
Steven Walczak, Narciso Cerpa, in
Encyclopedia of Physical Science and Technology (Third Edition)
, 2003
- [9]Reinaldo Padilha França, ... Yuzo Iano, in
Trends in Deep Learning Methodologies
, 2021
Machine learning
- [10]Juan Li, ... Hua-Feng Kong, in
Computers in Biology and Medicine
, 2024
Deep learning

Figure 5