\

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

COTTON YIELD PREDICTION USING MACHINE LEARNING TECHNIQUES

Uzma Rafeeq[1], Hafiz. M. Sana Ullah Badar[*2], Israr Hussain[3], Muhammad Ali Imran[4], Farrukh Baig[5]

[1,3] Institute of Computing, MNS-University of Agriculture, Multan 66000, Pakistan

[*2] School of Computer and information Engineering, Henan University, Kaifeng, Xinjiang 475001 China.

[4]Department of Agribusiness and Entrepreneurship Development, MNS University of Agriculture, Multan 66000, Pakistan

[5]Institute of Plant Protection, MNS-University of Agriculture, Multan 66000, Pakistan

[*2]badar@henu.edu.cn

DOI: https://doi.org/ 10.5281/zenodo.16727757

Keywords
SVM, Crop genotype, Resnet-50, Efficient Net, weather data, whitefly, resilience.

Article History
Received: 30 April, 2025
Accepted: 16 July, 2025
Published: 31 July, 2025

Corresponding Author: *
M. Sana Ullah Badar

Abstract

Predicting cotton yield is a critical task influenced by various factors, including environmental conditions, crop genotype and management practices. In yield forecast many crop models have been used in technical and strategic agricultural decision-making. The short- and long-term risk of the climate and weather was the primary cause of the expected to variability of crop production related parameters, as well as the use of natural resources. In this Research I willpredict White fly attacks on cotton plants leaves that can significantly impact yield, making early detection and prediction play crucial role for effective crop management. This research proposed a novel approach combining machine learning and weather analysis to detect white fly attacks andpredict their impact on cotton yield. We utilize ResNet-50 and Efficient Net models on image datato identify white fly infestations, achieving an accuracy of 98%. Additionally, I integrate weather analysis using Naive Bayes, SVM, and Logistic Regression to predict the likelihood of white fly infestations, achieving accuracies of 100%, 98.63%, and 90.21%, respectively. Our results show that the combined approach integrates cotton varieties andcultivation area with image data and weather to predict cotton yield as well and this research willhelp farmers to make decisions enabling farmers to take preemptive actions to mitigate losses. This research contributes to the development of precision agriculture techniques, enhancing crop resilience and productivity.

INTRODUCTION

Cotton crop play an important role in textile sector and provide good support to national economy. Pakistan ranks as 4th in cotton production while 3rd largest consumer of the world. Cotton is a natural crop extending world's largest textile industry with yearly business of $600 billion around the world [1].

Cotton yield is influenced by a variety of factors, including the crop's genotype, environment, and management practices. Variations in the environment over time and space havea significant impact on the yield of cotton crops from year to year and from location to location [2]. Growers can develop their management practices and financial decision through yieldforecasting [3]. It is evident that the cotton crop yield could be affected by several insect pests of which the whitefly (*Bemisia tabaci*) is

recognized to have a severe effect on the yield of the cottoncrop [4].
Heavy infestation of whitefly seen on the cotton plants, they were presented whitefly as global threat to cotton crops hence a serious threat to sustainable development in the agriculture sector across the globe. We will predict white fly infestation through mL models that weather is suitable for cotton production or it will less attack of whitefly on cotton leaves [5].
There are more instances of whiteflies on the plants, people are demanding the development of more effective, efficient, and, at the same time, organic forms of control. Analyzing the various challenges highlighted, researchers have been able to design various solutions ranging from machine learning, image processing, and computer vision

Methodologies [6].
Researchers have investigated to predict crop output due to critical and variant factors such as crop disease, pests, fertilizer, nature of soil, and change in temperature and rainfall. They explore the elements affecting in inspect agricultural yield, explore the features utilized, and examine carefully the use of deep learning techniques and performance indicators in crop yield prediction [7]. The number of cotton bolls has a big impact on fiber yield, which is perhaps the most important phenotypic feature for both farmers and plant breeders. Moreover, boll number helps to understand the genetic and physiological principles underlying crop growth and development, allowing for timely decisions on crop management to optimize profit. The accuracy recorded for boll counting was approximately 84.6% [8].
Whiteflies are tiny insects that can cause significant damage to cotton plants by feeding on the sap of the leaves. They thrive in warm temperatures generally between 68°F and 86°F, or 20°C and 30°C. In these temperature ranges, whiteflies reproduce rapidly, and their populations can explode if conditions are favorable. High temperatures, particularly above 30°C (86°F), can accelerate their reproduction and development. When whiteflies infest cotton plants, they can cause direct damage by sucking sap from the leaves, which weakens the plant and reduces its ability to photosynthesize effectively. Researchers have employed machine learning, image processing, and computer vision methodologies to devise numerous solutions to address these challenges [9]. This study demonstrated food security issues and serve as a foundation for farmers and policymakers to make informed decisions regarding field operations, crop and resources import export and make decision. The objectives of this research are: (1) to develop machine learning models for cotton yield prediction. (2) To improve decision making for farmers and stakeholders.

A. RESEARCH CONTRIBUTIONS
The contributions of this study encompass several key challenges associated with the utilization of machine learning methods, such as Naïve Bayes, logistic regression and SVM for predicting cotton yield based on weather parameters. This research contributes by tackling the formidable challenge of obtaining reliable and comprehensive datasets that encompass historical weather data and corresponding cotton yield information for Multan, Pakistan. The dedicated effort and resources invested in data collection and quality assurance enhance the foundation of this predictive model.
Another significant contribution Applying Resnet-50 and Efficient Net for Cotton Yield Prediction through Plant Health substantial influence on cotton yield. ResNet-(50) is a deep convolutional neural network including of 50 layers. It is a specialized convolutional neural network (CNN) model specifically designed for predicting Whitefly attack on cotton leaves. Efficient-Net is a CNN built upon a conception called "compound scaling." This concept talks the ancient trade-off amongst model size, accuracy, and computational effectiveness. The clue behind multiple scaling is to scale three essential dimensions of a neural network: width, depth, and resolution.
This paper will discuss the following section. The literature review is covered in section II, the methodology is covered in section III, the experiment and findings are covered in section IV, and the conclusion and future work are included in section V.

I. LITERATURE REVIEW
Machine learning has been experimented with in order to forecast yields in recent years. They aimed

to predict India's major crop yields. They obtained historical data from the Indian government between 1997 and 2012 and utilized the Random Forest classification algorithm to it. The accuracy of the Random Forest model was compared by the authors with three other classifiers: KNN, XGBoost, and Logistic Regression. The Random Forest classifier, according generates decision trees for various training data sets and then averages to produce a single extremely precise decision tree. The above results in results, 98%, with lower error while compared to other models [10]. A model that suggests the best crop to grow with the highest yield based on climate and other factors was developed by another study conducted in India. Rf gibes accuracy of 67%, then all other models, including KNN, XBoost, Logistic Regression, and Random Forest [11]. He focused on predicting barley yield throughout its growth period using a model with field data, remote sensing data, and climate data. The research, conducted in a major barley production area in Iran, employed with machine learning models, with the Gaussian process regression algorithm proving most effective. Results showed an estimated yield with an r2 of 0.84, RMSE of 737 kg ha−1, and MAE of 650 kg ha−1, one month before harvest. The study highlighted variations in estimation results across temporal contexts and agro - based zones, emphasizing the model's potentiality tool to predict barely yield and for food security [12]. Machine learning has shown a lot of promise recently for predicting crop productivity both globally and regionally [11]. It is uncertain how well crop models predict the differing, site- specific effects of expanded soil health measurements on crop productivity and other ecological systems [13]. To improve yield estimation accuracy, integrated plant height, cotton fiber index (CFI), canopy temperature, and the a* component of the CIELAB color space. This comprehensive approach achieved a higher correlation coefficient of 0.94. The study highlighted the trade-off between improved accuracy and the practical challenges associated with implementing a sophisticated, multi-camera system for cotton yield estimation [14]. They developed a cotton yield prediction model using various factors like VCI, GDD, SPI, LST, and past yield data. They used a method called Random Forest (RF) algorithm. Their study showed that this model provided accurate and quick predictions. The model was measured using SSR (a statistical measure), and it achieved values of 0.69, 0.60, and 0.39 in September, December, and February, respectively. This indicates the reliability of the model in forecasting cotton yields during these months. [15]. To train the SVM classifier used to recognize the cotton pixel in each plot image. The SVM classifier recognized cotton pixels with an accuracy of 89%, precision of 86%, recall of 75%, and F1-score of 80%. This work shows that pre-harvest cotton production prediction using aerial images and Ml approaches can be a dependable, and effective tool [16]. They leveraged a historical dataset of US sweet corn from 1992 to 2018 to (a) evaluate the performance of machine learning models in predicting sweet corn yields and (b) Identify the most significant variables influencing crop yield predictions. This dataset encompasses over 25 years of field-level data from the primary sweet corn production regions in the Upper Midwest and Pacific Northwest. Machine learning models were trained using 67 variables related to crop genetics, management, weather, and soil factors. Among the models tested, the random forest model outperformed the others, demonstrating the lowest Root Mean Square Error (3.29 Mt/ha) and the highest Pearson's correlation coefficient (0.77) between predicted and observed yields [17].

I. PROPOSED CONCEPTUAL MODEL

Experiment I: Weather Analysis for Detection of Whitefly Attack on Cotton Yield

Weather data is gathered from the Meteorological department Muhammad Nawaz Sharif University of Agriculture, Multan. Weather data is collected since last four years and we will predict through ML model that overall weather is suitable for cotton growth or it will be affected by whiteflies In this experiment we will use three models Naïve Bayes, logistic regression and SVM.

Naïve Bayes: Naive Bayes is a basic but effective machine learning method used for sorting things. It uses Bayes' idea, which explains how likely something is to happen based on what we know before. The "naive" part comes from the assumption that each part of the information has no connection to the others. This makes the calculations easier and faster, especially when there's a lot of data.

SVM: SVM is a Machine Learning algorithm used for classification and regression purposes. How does SVM work - At a high level, SVM finds the hyperplane which separates different classes/groups of data points in higher dimension space. The idea of finding a hyperplane that maximizes margin, distance between the separating planes (boundaries) and nearest data points from each category. This allows SVM to create a decision boundary which can actually separate different classes while maximizing the generalization power of model. Logistic Regression: Logistic Regression is a basic conventional method broadly used to solve binary classification issues. Surprisingly, it is actually a classification algorithm despite its name of being a regression algorithm. Based on the description of the idea behind Logistic Regression, it involves estimating the probability of the input variable belonging to a certain class. It is called so because it uses a logistic function or also referred to as sigmoid function to transform input features into a value between 0 and 1, depicting the probability of the input data belonging to the positive class. This logistic function helps to confine the model's output within the interval [0, 1] making it
ideal for problems with two classes. Usually, if the computed probability is greater than a given value (it is commonly set at 0. 5), the treated sample belongs to the positive class; otherwise, it falls into the Negative class.

The research process comprises the following stages: 1) Data Acquisition 2) Feature selection 3) Model training and
validation 4) Model testing and results. Data acquisition: Weather data gathered through MNSUAM Weather station which is installed in MNS University of Agriculture, Block (C) includes the relevant variables for weather analysis, such as temperature, humidity, rainfall, wet, dry and pay attention in term of space and time. Feature selection: Feature extraction for weather analysis using the parameters Temperature Max, Dry, Wet, Relative Humidity, Rainfall, and possibly a target variable. Here's how you might approach feature extraction for this analysis:

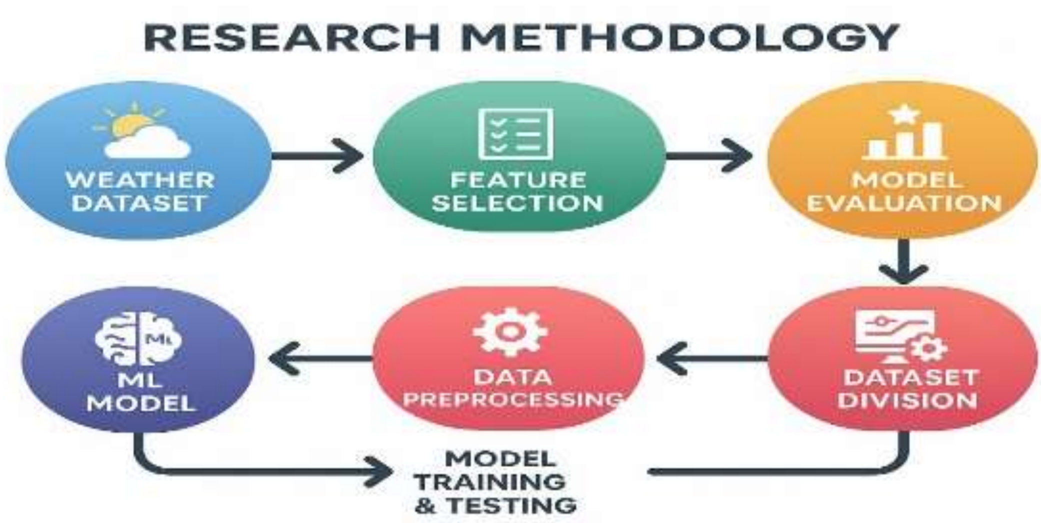


FIGURE 1. Proposed methodology of Weather Dataset

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

Model training and validation: In this step we train our ml models on these parameters Temperature Max, Dry, Wet, Relative Humidity, Rainfall to predict cotton yields here we will train three machine learning models to predict weather is suitable for cotton yield or white fly attacks SVM, Naïve Bayes and logistic regression model will be trained to get better results.

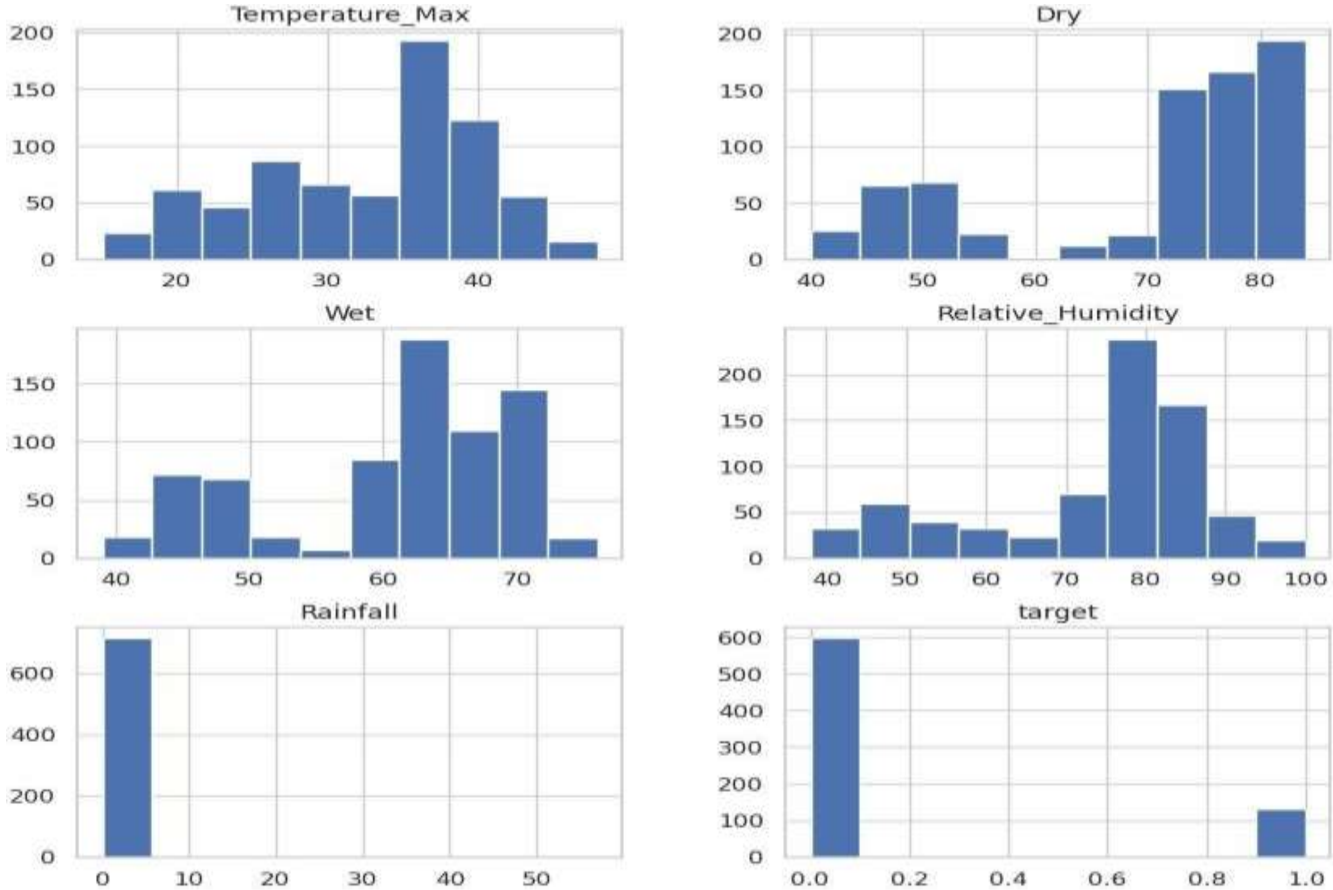
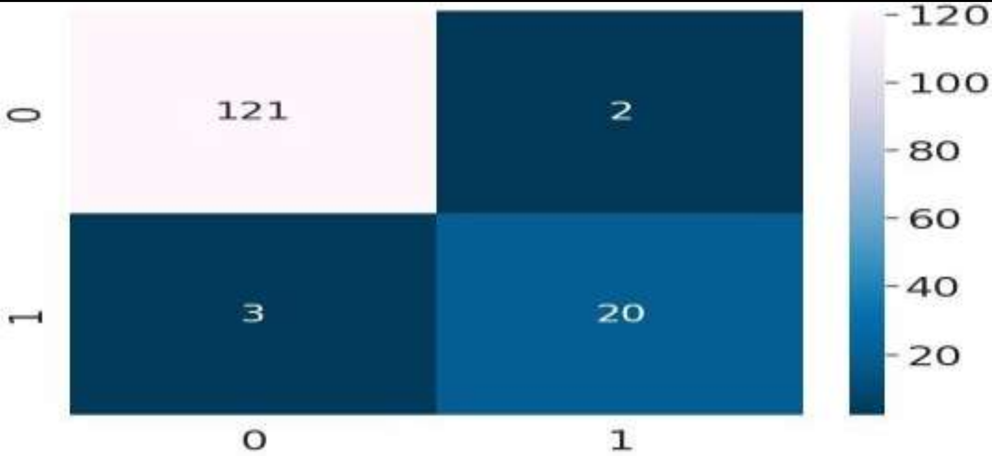


FIGURE 2. Weather analysis features on which we predict whitefly infestation

	Temperature_Max	Dry	Wet	Relative_Humidity	Rainfall	target	target_encoded
count	728.000000	728.000000	728.000000	728.000000	728.000000	728.000000	728.000000
mean	32.899725	70.034341	61.046703	73.656593	0.550275	0.179945	0.179945
std	7.604900	13.090634	9.152717	14.834073	4.006518	0.384406	0.384406
min	15.000000	40.000000	39.000000	38.000000	0.000000	0.000000	0.000000
25%	27.000000	59.500000	55.750000	67.750000	0.000000	0.000000	0.000000
50%	35.000000	75.000000	63.000000	78.000000	0.000000	0.000000	0.000000
75%	39.000000	80.000000	68.000000	85.000000	0.000000	0.000000	0.000000
max	48.000000	84.000000	76.000000	100.000000	57.000000	1.000000	1.000000

FIGURE 3. Training and Validation of weather analysis features

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025



Model testing and results: We tested and validate out weather data on three model SVM, Logistic Regression and Naïve Bayes. After testing the ML models achieved the following results. SVM confusion matrix shows that out of 146 predictions 141 prediction were correct. These results shows that SVM gives high accuracy of 98.63%. IF prediction = 1, There is cotton yield loss predicted due to whitefly attack, IF prediction = 0, it shows High cotton yield prediction. Logistics Regression confusion matrix shows that out of 146 predictions 132 prediction were

correct. These results shows that Logistics Regression gives accuracy of 90.21%. IF prediction = 1, There is cotton yield loss predicted due to whitefly attack, IF prediction = 0, it shows High cotton yield prediction. Native Bayes matrix shows that out of 146 predictions 146 prediction were correct. These results shows that Naive Bayes has accuracy of 100%. IF prediction = 1, There is cotton yield loss predicted due to whitefly attack, IF prediction = 0, it shows High cotton yield prediction.

Figure 4: SVM Confusion matrix

Figure 5: Logistics Regression Confusion matrix

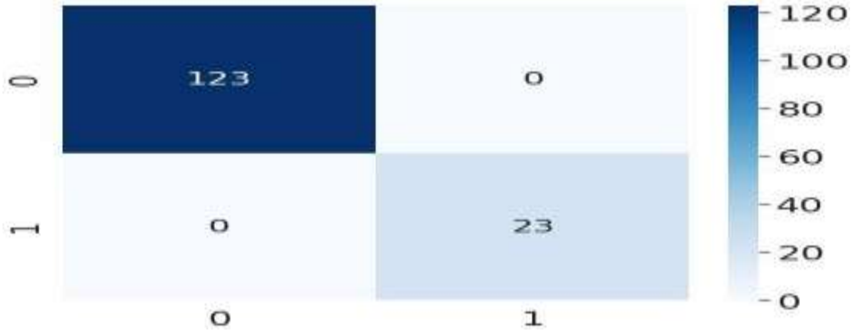


Figure 6: Native Bayes Confusion matrix **Model Evaluation for weather based data:**

Table 1: Model Evaluation Models Accuracy

NATIVE BAYES	100 %
SVM	98.63%
Logistic Regression	90.21%

We checked the trained models using accuracy as the main measure. Naïve Bayes got a perfect score of 200%, meaning it was always correct. This suggests that the model's assumptions, like the independence between features, matched the dataset very well. It did a good job predicting cotton yield and white fly attacks based on weather data, showing clear patterns in the dataset. Support Vector Machine (SVM) got an accuracy of 98.63%, which is almost perfect. It performed very well, understanding the complex connections between weather variables, cotton yield, and white fly attacks. The model likely found a clear distinction between different categories, making accurate predictions. Logistic Regression achieved an accuracy of 90.21%. While it still did well, it was slightly less accurate than Naïve Bayes and SVM. It may have had more difficulty capturing all the details in the dataset.

Experiment II Applying Resnet-50 and Efficient Net for Cotton Yield Prediction through Plant Health

ResNet-(50) is a deep convolutional neural network including of 50 layers. It is a specialized convolutional neural network (CNN) model specifically designed for predicting Whitefly attack on cotton leaves. It is a most common insect of cotton that effect yield by suckling on the base of the leaf and increasing diseases like Cotton Leaf Curl Virus. Whiteflies suck the sap of foliage and discharge a liquid on to the leaves on which a black fungus develops. This affect the photosynthesis, the food making procedure of the plant, and so hand down the power of the plant. To effectively learn and recognize patterns in images of white fly-affected cotton leaves, Resnet 50 leverages deep learning techniques. Subsequently, the neural network has developed complete and useful feature demonstrations for a wide spectrum of images. The image response size is set to 224*224 pixels. It was accomplished on 20 epochs using Google Colab. ResNet-50, highlighting the use of residual connections to train deep neural networks efficiently. Renet-50, a convolutional neural network architecture, has gained widespread recognition in the realm of image classification due to its remarkable performance across diverse tasks. [18].

Efficient Net model: Efficient-Net is a CNN built upon a conception called "compound scaling." This concept talks the ancient trade-off amongst model size, accuracy, and computational effectiveness. The clue behind multiple scaling is to scale three essential dimensions of a neural network: width, depth, and resolution. Efficient net model in this studies is designed to predict white fly attacks on cotton leaves. To effectively learn and recognize patterns in images of whitefly-affected cotton leaves, efficient net leverages deep learning techniques. As a result, the neural network now possesses thorough and insightful feature representations for a wide range of images. It has 224 × 224 pixels selected as the picture input size. It had 20 epochs of training.

Dataset was collected through keggle.com on cotton leaf images consist of white fly attacks on cotton leaves and healthy cotton leaves. The dataset is divided into training (70%), validation (20%) and testing sets (10) as shown in Figure 8.

To confirm the consistency and effectiveness of the RES-NET-50 and Efficient-Net model for identifying whitefly cotton leaf disease, it is crucial to highlight the importance of maintaining appropriate training, validation, and testing datasets. Independent and representative datasets that accurately reflect the problem domain are necessary. By carefully selecting samples that resemble the real-world data the model will encounter, we enable effective learning and generalization. The careful separation and allocation of data into training, validation, and testing sets enable training on labeled examples, fine-tuning performance through validation, and ultimately assessing the model's ability to identify white fly infestation on cotton leaf in unseen testing data. This meticulous approach facilitates the construction of a robust model capable of accurately detecting the disease in practical applications.

Results for Resnet-50 model:

Model training is the process when model learns about given cotton leaf image data, the next step is evaluation of this learning process. This can be represented in the form of model results of 49 training and validation. In the following Figure 10 (a) accuracy of model is demonstrated as a line graph, and loss in Figure 10 (b). Accuracy is a commonly used evaluation metric that provides an indication of model performance. It counts the percentage of correct forecasts made by the model, specifically measuring the entire number of images accurately classified during the testing phase.

Precision is calculated by dividing the number of properly forecast true labels by the total number of tags forecast as correct by the model.

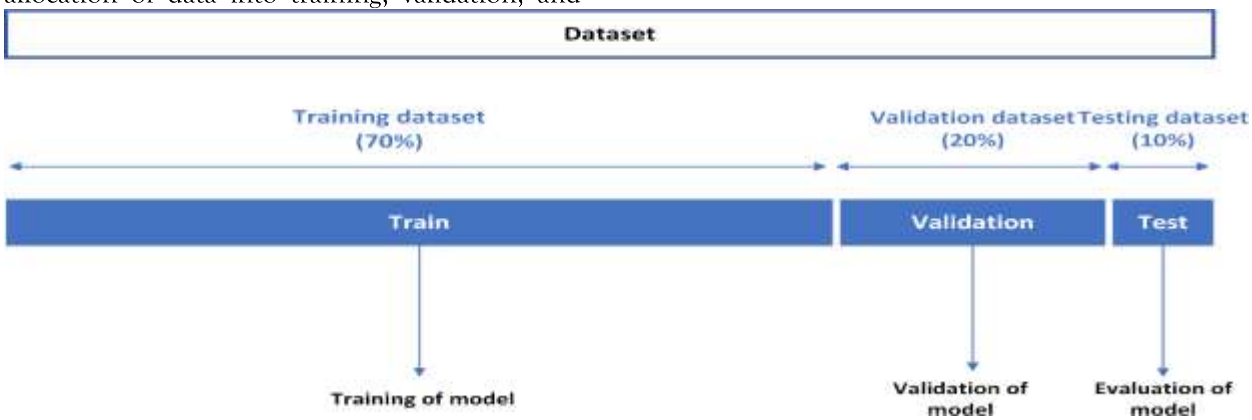


Figure 7. Dataset division for Resnet-50 and Efficient Net Models

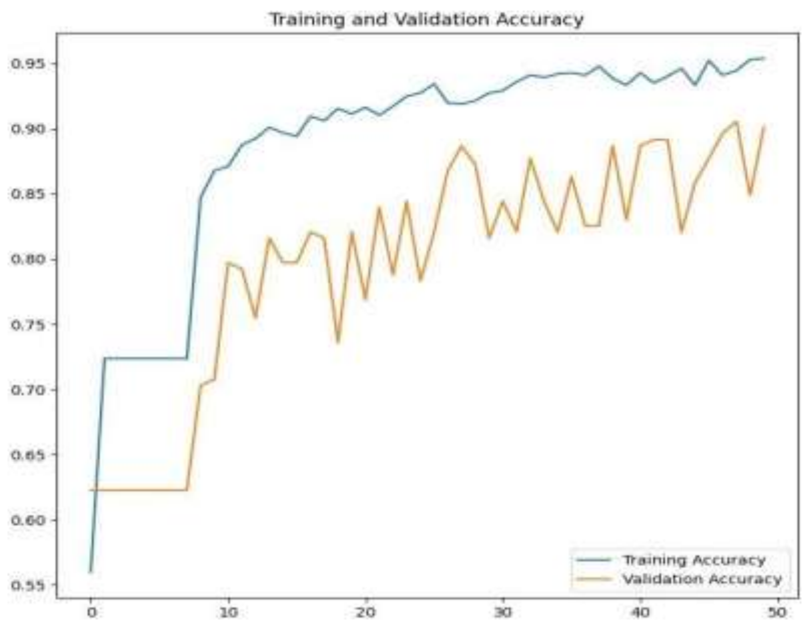


Figure 8. Random sample of healthy vs infected leaves

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

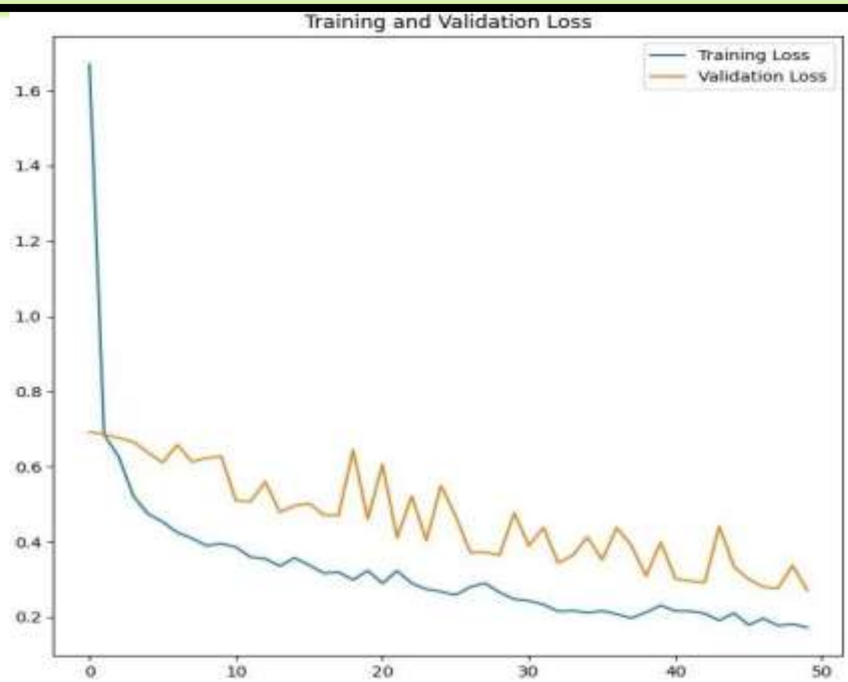


Figure 9 (a): Training and validation accuracyof Resnet-50 model **Figure 9 (b):** Training and validation loss of Resnet-50 model

Accuracy graph of ResNet-50 model. Blue line indicates training accuracy and orange line indicated validation accuracy (95%) at Y-axis. In X-axis, the number of epochs are represented these are 50 epochs to trained the model for achieving accuracy rate in resnet-50. Confusion Matrix for ResNet-50. This matrix shows that 386 predictions were tested precise out of 392 for high yield prediction on test sampled data. 6 predictions were gone wrong out of 392. Which shows ResNet-50 gives accuracy of 98%.

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

Figure 10. ResNet-50 Yield Prediction Based On white fly attack

Figure 11. Confusion Matrix for ResNet-50.

Results for Efficient model:

The process of acquiring knowledge from a set of cotton leaf image data is referred to as model training. The evaluation of this learning process is subsequently conducted, which is evident from the model's training and validation results. Precision is a commonly used evaluation quantity that specifies how well a model performs. It precisely quantifies the number of photos correctly identified during the testing phase, providing an accurate measurement of the model's prediction accuracy rate.

Confusion matrix for Efficient Net: Confusion matrix is computed by dividing the total number of accurate predictions—that is, the true positive and true negative outcomes—by the total number of samples. As the aforementioned matrix demonstrate [19]. Confusion Matrix for Efficient Net. This shows that 386 predictions were tested correct out of 392 for high yield prediction on test sampled data. 6 predictions were gone wrong out. Which shows Efficient Net gives accuracy of 98%.

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 7, 2025

Figure 12 (a). Training and validation accuracyof efficient net model

Figure 12 (b). Training and validation loss of efficient net model

Predicted Yield results with efficient net:

Figure 13. Efficient Net images based predicted results

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

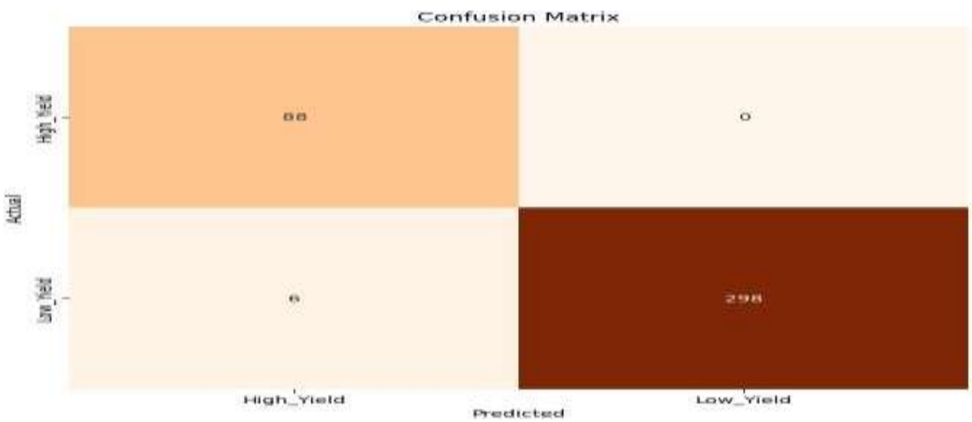


Figure 15. Confusion Matrix for Efficient Net.

Table 2: Training accuracy evaluation for image based models				
	Training-loss	Training-accuracy	Val-loss	Val-accuracy
	0.1719%	95%	2%	95%
	5.9151e-05	100%	0.0497%	98%

Yield prediction: Machine learning models integrated with cotton variety and land area when the code runs, it will first ask for the type of cotton. Then, will check the image for whitefly infestation. If detected, it will provide details, otherwise, it will indicate that the image is good in health. Next, it will determine the area of cotton and calculate the expected yield, showing the loss due to the disease of white fly infestation. It will then display the actual yield and analyze the image to show the unhealthy cotton along with the yield and cotton type and areas. It will also specify the type of cotton, whether it is CKC-3 or CKC-6 or unknown and results are given below.

Figure 16 (a) is showing that cotton plant is unhealthy with ckc-6, cultivation area is 10 acres and expected yield 276.25 Maunds and total yield loss due to weather is 48.75 Maunds. Figure 16 (b) is showing that image is healthy and cotton type is ckc-6 and cultivation area is about 10 acres and expected yield is 325.00 Maunds. Figure 16 (c) is showing that image is healthy and cotton type is ckc-6 and cultivation area is about 10 acres and expected yield is 500.00 Maunds. Figure 17 (d) showing that healthy cotton type of ckc-3 and cultivation of farmers is 20 acres gives us the expected yield is 362.12 Maunds. Total yield loss due to weather is 137.88 Maunds.

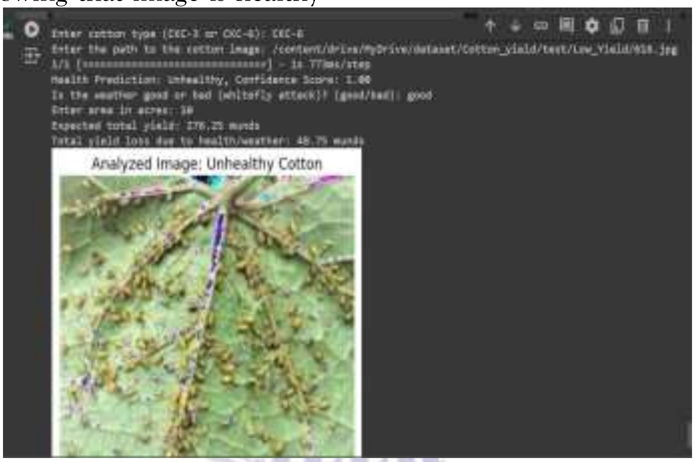


Figure 16 (a). Image Yield Prediction of Unhealthy plants.
Figure 16 (b): Image Yield Prediction of Healthy plant leaves.

Figure 16 (c). Image Yield Prediction of Healthy plant leaves
Figure 17 (d). Image Yield Prediction of Unhealthy plant leaves.

Discussion:
In this study, two CNN based models namely Resnet-50, and Efficient Net are applied for the identification of white fly cotton leaf disease using image dataset of cotton leaves to address the current limitations of previous studies. Both of these models are optimized in terms of accuracy and time required for processing the given input image dataset. Both of applied models have outperformed for the identification of white fly on cotton leaf disease using image data. It has been investigated that both of these models Resnet-50 and Efficient Net can be utilized for concerned disease identification at real time in the field. Evaluation metrics are used to check the performance of proposed methodologies. The suggested technique efficient net model consistently achieves high accuracy in identification, regardless of variations in image shapes, sizes, backgrounds, and resolutions. Processing time for each image is not increased when using these models. Both the Resnet-50 and Efficient Net models excel in real- time picture recognition, making them highly valuable in practical applications. Additionally, I integrate weather analysis using Naive Bayes, SVM, and Logistic Regression to predict the likelihood of white fly infestations, achieving accuracies of 100%, 98.63%, and 90.21%, respectively. Our results show that the combined approach integrates cotton varieties and cultivation area with image data and weather to predict cotton yield as well and this research will help farmers to make decisions enabling farmers to take preemptive actions to mitigate losses. It successfully captured the complex relationships between weather variables, cotton yield, and white fly attacks. The combined approach integrates cotton varieties and cultivation

area with image data and weather to predict cotton yield as well and this research will help farmers to make decisions enabling farmers to take preemptive actions to mitigate losses. This research contributes to the development of precision agriculture techniques, enhancing crop resilience and productivity.

Reference:

[1]. J. Ashraf, D. Zuo, Q. Wang, W. Malik, Y. Zhang, M.A. Abid and G. Song, "Recent insights into cotton functional genomics: progress and future perspectives.'' Plant Biotechnology, vol. 16, no.3, pp. 699-713, 2018.

[2]. T. Horie, Y. Masaharu, and N. Hiroshi, "Yield forecasting." Agricultural systems vol. 40, no. 1-3, pp. 211-236, 1992.

[3]. S. Khaki and L. Wang, "Crop yield prediction using deep neural networks. Frontiers in plant science," vol. 10, pp. 621, 2019.

[4]. M. Abubakar, B. Koul, K. Chandrashekar, A. Raut and D.Yadav, "Whitefly (Bemisia tabaci) management (WFM) strategies for sustainable agriculture: a review. Agriculture," vol. 12, p.1317, 2022.

[5]. L. Zhang, , Z. Zhang, Y. Luo, J. Cao and F. Tao, "Combining optical fluorescence thermal satellite and environmental data to predict county-level maize yield in China using machine learning approaches", Remote Sens., vol. 12, p. 21, 2019.

[6]. B.R. Hussein, O.A. Malik, W.H. Ong and J.W.F. Slik, "Applications of computer vision and machine learning techniques for digitized herbarium specimens," A systematic literature review. Ecological Informatics, vol. 69, p.101641, 2022.

[7]. S. Archana and P.S. Kumar, "A Survey on Deep Learning Based Crop Yield Prediction," Nature Environment & Pollution Technology, vol. 22, no. 2, 2023.

[8]. B. Peng, K. Guan, W. Zhou, C. Jiang, C. Frankenberg and Y. Sun, "Assessing the benefit of satellite-based solar-induced chlorophyll fluorescence in crop yield prediction", Int. J. Appl. Earth Observ. Geoinf, vol. 90, 2020.

[9]. M. Shahhosseini, G. Hu, I. Huber and S.V. Archontoulis, "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt," Scientific reports, vol. 11, pp. 1-15, 2021.

[10]. P. Kamath, P. Patil, S. Shrilatha and S. Sowmya, "Crop yield forecasting using data mining. Global Transitions Proceedings," vol. 2 pp. 402-407, 2021.

[11]. W. Dodge, "Image based yield estimation in cotton using UAS (Ph.D. dissertation)". Texas Tech University, Lubbock, TX, United States.

[12]. A. Sharifi, "Yield prediction with machine learning algorithms and satellite images," Journal of the Science of Food and Agriculture, vol. 101, pp. 891-896, 2021.

[13]. D. Geisseler and R. Wilson. "Nitrogen in potato rotations with cover crops: Field trial and simulations using DSSAT". Agronomy Journal vol. 112, pp. 2275-2287, 2020.

[14]. A. Feng, J. Zhou, E. D. Vories, K. Sudduth and M. Zhang, "Yield estimation in cotton using UAV-based multi-sensor imagery," Biosyst.Eng, vol. 193, pp.101–114, 2020.

[15]. N. Prasad, N. Patel and A. Danodia, "Crop yield prediction in cotton for regional level using random forest approach," Spatial Inf. Res., vol. 29, pp. 1-12, 2020.

[16]. J. Rodriguez-Sanchez, C. Li and A.H. Paterson, "Cotton yield estimation from aerial imagery using machine learning approaches," Frontiers in plant science, vol. 13, p. 870181, 2022.

[17]. D.S. Dhaliwal and M.M. Williams, "Sweet corn yield prediction using machine learning models and field-level data." Precision Agriculture, vol. 25, pp. 51-64, 2024.

[18]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770-778, 2016.

[19]. B. Basso and L. Liu, "Seasonal crop yield forecast: Methods, applications, and accuracies," Advances in agronomy, vol. 154, pp. 201-255, 2019.