JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Cotton Disease Prediction Using Machine Learning

[1] **Dr. Deepali Sale ,[2]Omkar Chaudhari, [3] Avinash Baikare, [4]Sumit Chaskar, [5]Anil Gujar,**
[1] Professor, [2,3,4,5] Students
I. Department of Computer Engineering,
[1]SPPU, Dr. D.Y.Patil College of Engg. & Innovation, Varale, Talegaon, Pune, Maharashtra, India.

Abstract: **Cotton cultivation is an important economic activity all over the world. However, it faces many challenges due to various pests and diseases that adversely affect the plant's leaf area. This project aims to develop a web application by using deep learning algorithms to identify and distinguish common cotton diseases, such as those caused by sucking and chewing pests, curl virus etc. The application will not only diagnose diseases but also recommend suitable solutions to farmers. Early disease detection is crucial for preserving crop yield. The proposed model, employing meta-deep learning and Convolutional Neural Networks (CNNs), promises accurate disease identification, contributing to the protection of cotton crops**

Keywords: Deep Learning, Convolutional Neural Network (CNN), cotton leaf disease, Diagnose, Yield, Disease identification

I. Introduction

Agriculture serves as an important source of economic, revenue and food supply, particularly as the world's population continues to grow. Meeting the rising food demand is crucial for both maintaining local populations and also to the global economy through agricultural exports. However, the agriculture sector faces significant challenges, with crop diseases emerging as a major concern. These diseases can lead to reduced crop yield, impact food quality and the quantity. Cotton, a major cash crop, plays a significant role in the textile industry by providing fiber for clothing production. The success of cotton cultivation relies heavily on effective plant protection methods. Traditional methods of identifying crop diseases often involve domain specialists physically visiting the sites for inspections. This approach is time-consuming, labor-intensive, and also need constant monitoring. Furthermore, it may not be accessible to farmers in remote areas without access to experts, rendering it unreliable. Emerging technologies, such as Deep Learning and Computer Vision, have made the way for autonomous and accurate crop disease detection, eliminating the need for human intervention. This automated approach is not only cost-effective but also significantly faster, ultimately contributing to the agriculture sector's efficiency. Plant diseases are a global concern, causing substantial economic losses. Early and precise disease diagnosis is critical to preventing further spread and implementing effective treatment measures. The adoption of these methods for disease detection provides agronomists with objective tools for identifying plant diseases. One of the primary challenges in Computer Vision is the early and accurate identification of diseased plant leaves, which can be hard due to busy backgrounds, varying angles, and subtle symptoms. Visual symptoms are essential for diagnosis, but they can be challenging to extract from complex backgrounds. The central aim of this study is to develop a generalized model for accurate leaf disease identification. This model, based on meta–Deep Learning, offers generalizations that can be applied to various diseases, including leaf spot, bacterial blight, target spot, powdery mildew, leaf curl, nutrient deficits, and verticillium wilt.

In summary, agriculture's importance in sustaining the population and the economy is undeniable. However, the threat of crop diseases poses a significant challenge. Leveraging advanced technologies, such as Deep Learning and Computer Vision, provides a cost-effective and efficient means of early disease detection, benefitting the agriculture industry and ensuring food quality and quantity.

II. Literature Survey:

According to research [1] cultivating cotton in tropical regions exposes the crop to a diverse range of agricultural pests and diseases, demanding effective solutions. Furthermore, distinguishing symptoms of major pests and diseases during the early stages can be challenging for producers. To address this issue, the present research offers a deep learning-based solution for screening cotton leaves. This solution enables the monitoring of cotton crop health, facilitating improved decision-making for crop management while avoiding plagiarism.

The research [2] states that a TensorFlow model can be built successfully and subsequently converted into a CoreML model to facilitate iOS app development. The iOS mobile app has been successfully developed, and it can accurately detect two cotton leaf diseases: boll rot and fungal leaf spot, in addition to identifying healthy leaves. This app fully adheres to the iOS programming guidelines and human interface design standards. Notably, the app operates without requiring an active internet connection on the device, as the ML model is embedded within the app, enabling offline functionality. To enhance the model's accuracy further, various approaches can be explored. These include modifying the training dataset, incorporating a more diverse range of images, adjusting input parameters, and exploring alternative algorithms. These measures can help improve the model's performance in identifying cotton leaf diseases while maintaining its integrity.

The research [3] is focused to create a system which will be able to diagnose the disease present on the leaf of the cotton plant. Since detecting the disease present on cotton crop with necked eyes might bring some errors, so doing the same with the help of machine will reduce the chances of misprediction. The main module of the system will help the user from analysing the disease, predicting the disease, recommending the remedies for that disease. The other modules of the system will provide the list of agriculture centres which exclusively doing their research on cotton crop, it will also suggest the products to the user that they may use and can purchase it directly. The system will be designed in such a manner that it will help the user all throughout their journey of analysing the present on the cotton crop.

In [4] research, which focuses on the detection and classification of cotton leaf diseases utilizing image processing and machine learning methods, we conducted an extensive survey. Additionally, we delved into the realm of background removal and segmentation techniques. Our findings revealed that for effective background removal, the conversion of the colour space from RGB to HSV proves to be highly beneficial. Furthermore, we identified the thresholding technique as a superior method when compared to other background removal techniques, yielding more accurate results.

To perform colour segmentation, we masked the green pixels in the background-removed image and subsequently applied Otsu thresholding to the resulting masked image, resulting in a binary image. This technique proved to be instrumental in extracting precise disease-related features from the images.

III. PROPOSED METHODOLOGY

1. Data Collection: The first step in building a cotton disease prediction system involves gathering a diverse dataset of cotton plant images. This dataset should encompass various growth stages, including both healthy plants and those affected by different diseases. The dataset's diversity is crucial for training a robust model capable of accurately identifying various diseases.

2. Data Preprocessing: Data preprocessing is essential to ensure the quality and consistency of the dataset. This step involves cleaning the data by removing any noise or irrelevant information. Additionally, images may be resized for uniformity, and environmental variables may be normalized to eliminate potential biases.

3. Data Splitting: To evaluate the model's performance effectively, the dataset is typically divided into three subsets: a training set for model training, a validation set for hyperparameter tuning, and a test set for assessing the model's real-world performance. This division helps ensure that the model generalizes well to unseen data.

4. Feature Extraction: Feature extraction is a critical phase where meaningful characteristics are derived from the images. Techniques such as Convolutional Neural Networks (CNNs) or Transfer Learning can be applied to extract relevant features, such as colour histograms, texture patterns, and shape descriptors. These features are crucial for disease classification.

5. Model Selection: Choosing the right machine learning model is vital for accurate disease prediction. Various models can be considered, including Random Forest, Support Vector Machines (SVMs), or deep learning architectures like Convolutional Neural Networks (CNNs). The selection should be based on the specific classification task and the characteristics of the dataset.

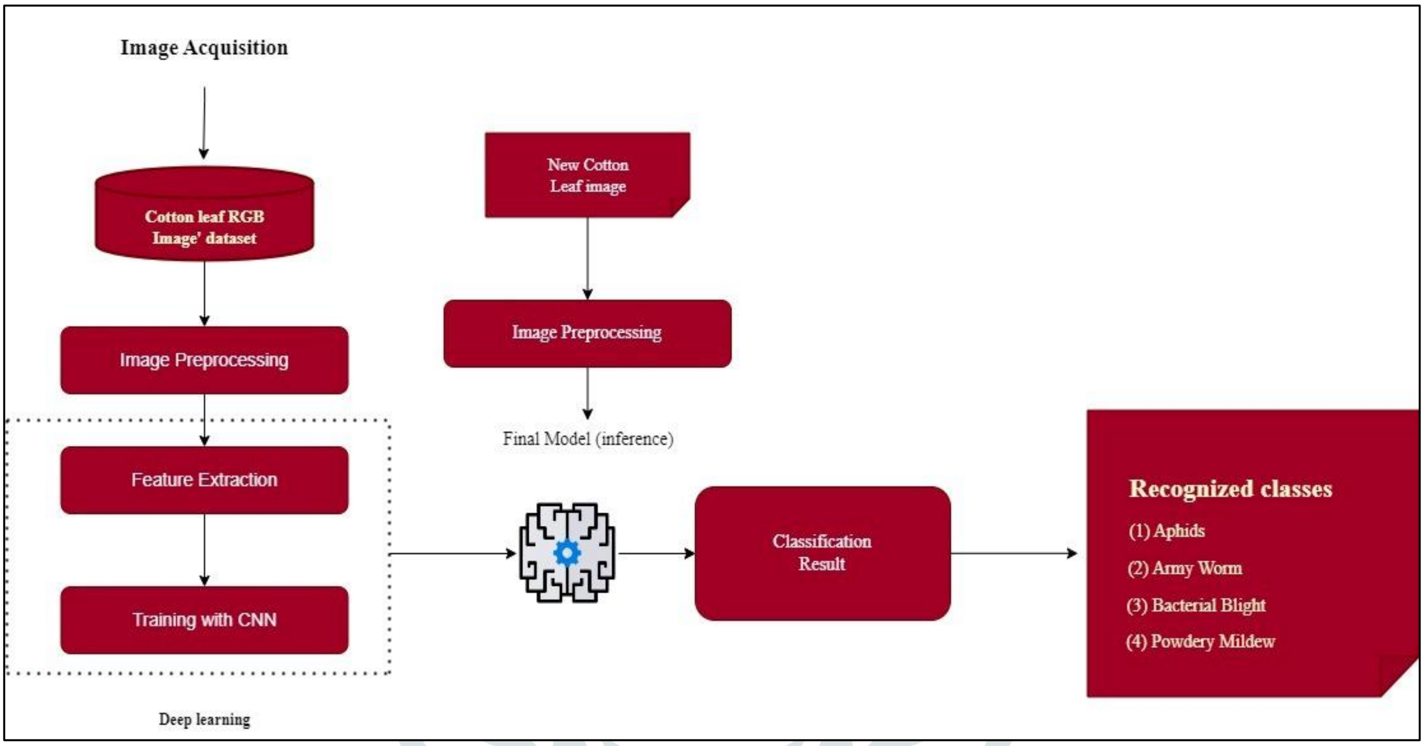
6. Model Training: Once the model is selected, it is trained on the labelled training dataset using both the extracted image features and environmental data. The training process aims to teach the model to recognize disease patterns and make accurate predictions.

7. Testing: After training, the model is rigorously tested using a separate test dataset. This step evaluates the model's real-world performance and its ability to generalize to new, unseen cotton plant images. It provides insights into the model's accuracy and reliability.

8. Deployment: Once the model has been thoroughly tested and proven to perform well, it can be deployed for practical use. Common deployment methods involve utilizing web frameworks like Flask, React, or Django to create a user-friendly application that allows farmers to upload images of their cotton plants for disease prediction

The first stage is data collection phase in this phase the required data is collected for processing,
The content of data solely depends on what you require the model to predict it can be image, text, audio, video etc. The next stage is the data preprocessing stage where the data is cleaned as the data have been gathered from the various sources and can be incomplete in many cases , hence this is a very crucial stage due to data preprocessing the noise in the data , animalities in the data , missing values get detected and are treated .Data Splitting phase , the data is splited in the two components , 1.Train Data and 2.Test Data the train data is used for training the model and the test data is used for testing the model whether the model is correctly predicting on the unseen data .Model selection include which mode is to be picked in our case we consider the CNN model as the most of the work is related to the images .Model training is the phase where the model is trained on the seen data and then tested on the test data , After the rigorous training of the model considering the accuracy of the model the model is now ready for deployment . The model can be deployed hence it will provide a good user experience to it users, the model can be deployed using the Flask, React, Django etc.

BLOCK DIAGRAM

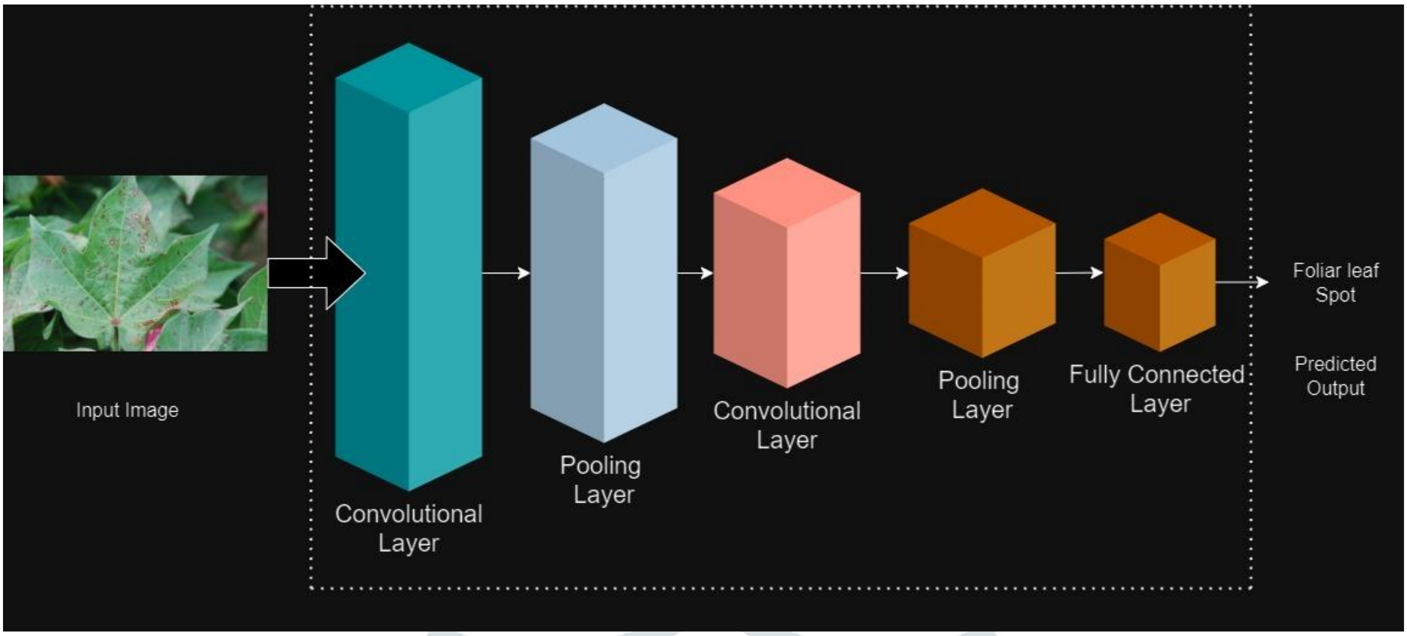


IV. ALGORITHM DESCRIPTION:

CNN and Transfer learning both the approaches can be used to get the accurate results, both of them will be implemented on the dataset and the one which performs better shall be considered for further analysis

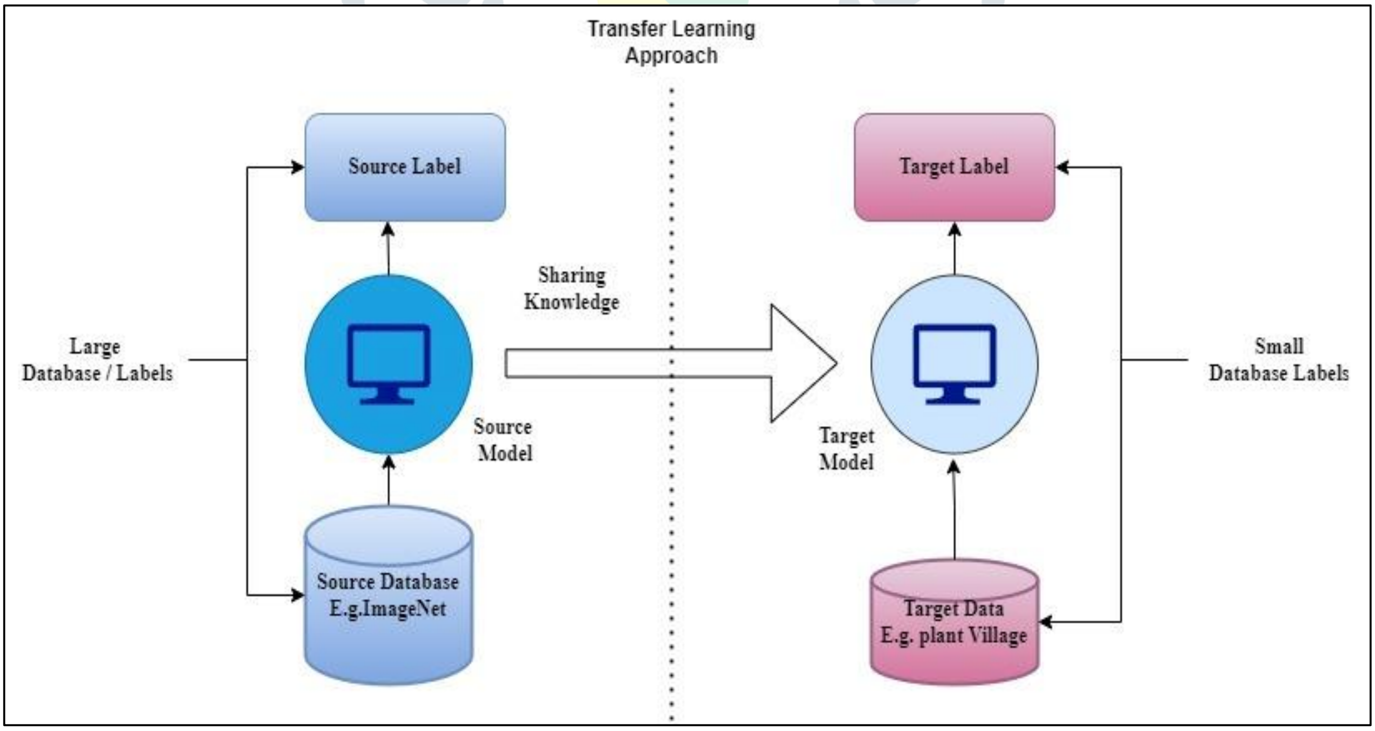
a. CNN:

A Convolutional Neural Network (CNN) is a specialized neural network architecture tailored for processing grid-like data, such as images and spatial data. Unlike traditional neural networks, CNNs leverage convolutional layers to automatically learn and extract relevant features from the input data. These layers employ small filters that slide across the input, enabling the network to detect patterns, edges, and more complex features within the image. This feature extraction process is a crucial element in image analyses also incorporate pooling layers, which reduce the spatial dimensions of the data and decrease computational complexity. This helps make the network more robust to variations in scale and position. Lastly, fully connected layers are used for classification tasks, where the network makes predictions based on the features it has learned. The strength of CNNs lies in their ability to capture hierarchical representations of visual data, making them highly effective for image recognition, object detection, and various computer vision tasks. They have been instrumental in advancements in fields like autonomous vehicles, medical image analysis, and natural language processing.



b. Transfer Learning:

In transfer learning, a pre-trained model serves as a starting point, offering valuable knowledge and feature representations that can be applied to a new problem domain. Fine-tuning involves updating the model's parameters on a smaller dataset specific to the new task, enabling the model to adapt and perform well without requiring extensive training from scratch. This technique is widely used in various machine learning applications, particularly in scenarios where data is limited or training resources are constrained.
Models such as VGG19, Resnet are trained on ImageNet on larger datasets and provides us pre implemented features.
Transfer learning provides a vast not of options where the model can be applied with no of layers for getting the accurate results, Transfer learning will take a model as a base model which has already been trained on the large datasets which will help for improvising the accuracy of the model



V. Datasets:

The dataset typically includes high-resolution images of cotton plants at different growth stages, capturing both healthy plants and those affected by various diseases. Alongside these images, there are detailed annotations specifying the presence and, where relevant, the type and severity of diseases, which serve as crucial data for training and testing disease prediction models. Additionally, environmental data, such as temperature, humidity, and soil quality, are often included to explore the relationship between disease prevalence and environmental factors. The dataset may also feature metadata, providing information on the location, date of image capture, and contextual details, offering valuable insights into disease patterns. In terms of dataset size, the comprehensiveness and potential of the dataset are contingent on the number of images and associated data points. Moreover, a well-constructed cotton disease dataset embraces a variety of diseases commonly affecting cotton plants, ensuring that the model can accurately identify different diseases and their manifestations. These datasets are instrumental in empowering researchers and data scientists to develop robust machine learning models for disease prediction and crop protection within the cotton agriculture sector.

The dataset consists of 7 classes which are again divided in training data and the test data, as we know the training data will be used for the training purpose and the test data shall be used for the testing purpose

VI. Conclusion and Future Scope:

We can observe that using deep learning we can reduce the human effort and also increase the yield. The model accurately predicts the image if the image belongs to the diseased class or not. The methods like transfer learning can be used to improve the accuracy of the model. By providing farmers with the means to detect diseases early and take appropriate measures, the project can lead to increased cotton crop yields, contributing to food security and economic stability for farmers. The implementation of Convolutional Neural Networks (CNNs) and other deep learning architectures empowers the accurate identification of various cotton diseases, ultimately leading to early diagnosis and timely intervention. By leveraging technology, we can mitigate the impact of crop diseases, enhance agricultural productivity, and contribute to global food security. This approach not only saves time and resources but also gives a more objective and optimal solution, benefitting both farmers and the broader agricultural community. The future of cotton disease prediction using deep learning is indeed bright, and it holds the promise of sustaining and elevating the cotton industry to new heights.

In the coming years, as we refine and expand upon these methods, the agricultural community can anticipate a more resilient and productive cotton industry. This amalgamation of data science and agriculture not only enhances the livelihoods of farmers but also plays a pivotal role in ensuring a sustainable and secure food supply for the growing global population. As the data will keep on accumulating the model will give more accurate predictions, the model can be trained on variety of unseen data and made a generic to detect the cotton disease. The predictions made still differs by some number can be improved if enough data is provided. As this is tested on few diseases due to the lack of the data, in future the no. of diseases can be increased

VII. References:

[1] A Survey on Detection and Classification of Cotton Leaf Diseases by Bhumika S.Prajapati Vipul K.Dabhi Harshadkumar B.Prajapati 978-1-4673-9939-5/16/$31.00 ©2016 IEEE

[2] Meta Deep Learn Leaf Disease Identification Model for Cotton Crop Muhammad Suleman Memon 1,* , Pardeep Kumar 2 and Rizwan Iqbal 3 22 June 2022

[3] Cotton Disease Detection Using TensorFlow Machine Learning Technique Sandeep Kumar , Rajeev Ratan , and J. V. Desai 24 August 2022

[4] Leaf Disease Identification: Enhanced Cotton Leaf Disease Identification Using Deep CNN Models P. Sivakumar; N. Sri Ram Mohan; P. Kavya; P. Vinay Sai Teja 02 March 2022

[5] Cotton Plant Fungal Disease Classification using Deep Learning Models Sandhya Dhage; Vijay Kumar Garg 19 June 2023

[6] Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images Aditya Parikh∗ , Mehul S. Raval† , Chandrasinh Parmar‡ and Sanjay Chaudhary 2023

[7] Identification of Cotton Leaf Lesions Using Deep Learning Techniques Rafael Faria Caldeira,[1,2,*] Wesley Esdras Santiago,[2] and Barbara Teruel[1] May 2021

[8] Detection of Cotton Plant Diseases Using Deep Transfer Learning Vani Rajasekar K Venu Soumya Ranjan JenaR. Janani Varthini S. Ishwarya 2021

VIII. AUTHOR'S BIOGRAPHY

Dr. Deepali Sale, Associate Professor, Dr. D.Y.Patil College of Engineering and Innovation, Varale,Talegaon. She has published more than 30 papers in National International conferences and Journal. Her area of expertise is in Image Processing andSignal Processing. She has completed her M.E. in Electronics and Ph.D in E&TC Engineering from College of Engineering, Pune.

Omkar Chaudhari, pursuing his Bachelor of Engineering in Computer Engineering from Dr. D.Y.Patil College of Engineering and Innovation,Varale, Talegaon, Pune. He has done his internship in Web developmentand Django from Elite software, Pune. He has completed his Diploma from Dyanshree Institue of Engineering and Technology, Satara

Avinash Baikare, pursuing his Bachelor of Engineering in Computer Engineering from Dr. D.Y.Patil College of Engineering and Innovation,Varale, Talegaon, Pune. He has done his internship in Web developmentand Django from Elite software, Pune. He has completed his Diploma from Government Polytechnic Nanded

Sumit Chaskar, pursuing his Bachelor of Engineering in Computer Engineering from Dr. D.Y.Patil College of Engineering and Innovation,Varale, Talegaon, Pune. He has done his internship in Web developmentand Django from Elite software, Pune. He has completed his Diploma from Dr Dy Patil Ploytechnic Akurdi

Anil Gujar, pursuing his Bachelor of Engineering in Computer Engineering from Dr. D.Y.Patil College of Engineering and Innovation,Varale, Talegaon, Pune. He has done his internship in Software Testing from Testometer software, Pune. He has completed his Diploma from T.B. girwalkar polytechnic ambejogai