# *Cotton Crop Disease Detection using Decision Tree Classifier*

Jayraj Chopda
Department of Information Technology
Universal College of Engineering
Kaman, Vasai 401212
Email: jayrajchopda1402@gmail.com

Sagar Nakum
Department of Information Technology
Universal College of Engineering
Kaman, Vasai 401212
Email: nakumsagar8248@gmail.com

Vivek Nakrani
Department of Information Technology
Universal College of Engineering
Kaman, Vasai 401212
Email: viveknakrani805@gmail.com

Prof. Hiral Raveshiya
Department of Information Technology
Universal College of Engineering
Kaman, Vasai 401212
Email: hiral.raveshiya.edu.in

***Abstract*—Smart Farming is a technique that offers high-ended application of modern farming by acquiring multiple data from sensors, robots, live stream, social media, etc. The concept of integrating these various data from multiple sources and processing it using multilevel databases is referred to as big data. Till today, there have been various developments in smart farming using image processing, data mining, IOT, etc. but as of now, Machine Learning is the industry which is emerging at a rapid pace. There has been an increasing demand related to real time applications in machine learning using supervised or unsupervised methods .The current scenario of traditional farming includes manually taking data, unpredictable weather conditions, sprinkling pesticides on diseases, etc. to yield productions, which is paving way to life threats for farmers especially in drought areas. With reference to the current scenario in traditional farming, there has been a dire need of predicated data in farming which can help farmers to know about their real time problems and act accordingly. To resolve their problems, we would like to propose a system which can predict cotton crop diseases using 'Decision Tree Classifier' by taking parameters as temperature, soil moisture, etc. This would help farmers by providing better quality productions and we would be also focusing on building an android application which will give real time output to the farmers in efficient ways.**

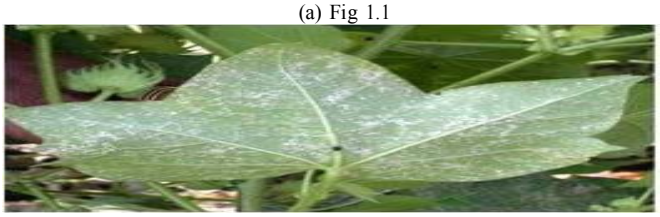
***Keywords—Decision Tree Classifier; Disease Detection; Machine Learning ; Smart Farming;***

## I.INTRODUCTION

Smart Farming provides technological solutions to the agri-cultural businesses as well as to the farmers. It has many applications in precision farming, weather forecasting, quality control, data analyses and collection, etc. With reference to that, crop disease detection is the subset of smart farming. Crop disease detection plays an integral role in the development of the crops. In present era, the approach used for detection by farmers is through naked eye. It requires continuous observation and monitoring which results into loss of time and it is expensive. In some regions, farmers need to go various places to take guidance from the experts.

Henceforth, automatic prediction of various diseases on the crop will provide real time benefits to the farmers that will save time, money, and life of the crop. To tackle the increasing inconvenience caused to farmers, a machine-learning model, which can predict occurrence of cotton crop disease based on the temperature in the real environment and from the soil moisture temperature is proposed in this paper. Here we are considering 3 major cotton diseases of which 2 are evaluated, Anthracnose [Fig 1.1], Areolate or Greymildew [Fig 1.2], Wilt [Fig. 1.3]

(a) Fig 1.1



(b) Fig 1.2

(c) Fig 1.3

Fig.1. Cotton Crop Diseases

## II. LITERATURE SURVEY

In this module, several Crop Disease Detection Systems that have been researched and implemented by other researchers are described . For further understanding on their methods and techniques, refer to the reference page at the end of this paper to search for the papers or text or even websites published.

1.In[1],authors have used Support Vector Machines (SVMs),for detecting and classifying various crop diseases through image processing. The images used for the work were acquired from the cotton field using digital camera. In pre-processing step, background removal technique is applied on the image in order to remove background from the image. Then, the background removed images are further processed for image segmentation using thresholding technique. Different segmented images will be used for extracting the features such as color, shape and texture from the images. At last, these extracted features will be used as inputs of classifier.The technique used here is based on supervised classification.

Drawbacks: The images used for the work were acquired from the cotton field using digital camera which makes the process costly and time consuming.

2.In[2],authors have expressed technological solutions using images of Cotton Leaf Spot and categorize the diseases using neural network.The classifier is trained to achieve smart farming, which includes early detection of disease in the groves, selective application of fungicide, etc. The proposed work is based on Image detection of edges which uses segmentation techniques in which,the acquired images are processed for improvement first. Then image segmentation using RGB color feature is carried out to get spots on diseases. Later, image features such as border, shape, color alongwith texture are extracted for the spots of disease to identify diseases and control the pest recommendation.It is based on supervised classification.

Drawbacks: It requires multiple image features such as boundary, shape, color and texture are extracted for the disease spots detection.

3.In[3],authors have suggested a robot which is a ground based agricultural robot that overcomes challenges existing in massive and complex satellite based solutions available as m-Services. It provides a small, mobile and reliable platform to automatically investigate farmland, detect diseases as well as spray the pesticide and insecticide. In near future, the farmers can obtain a consolidated view of the farm along with decision support statistics for planning purposes.It uses Artificial Neural Networks (Supervised) and K-means algorithm (Unsupervised).

Drawbacks: The designed robot requires higher maintenance and it proves to be a bit expensive.

4.In[4],authors have worked on diseases occurring on leaves of grape plant .The system acquires a single leaf of a plant as an input and process of segmentation is done after background removal. The segmented leaf image is then observed through high pass filter to detect the diseased part of the leaf. The segmented leaf texture is retrieved using unique fractal based texture feature. Fractal based features are locally invariant in nature and therefore provides a good texture model. The texture of every independent disease will be different. The extracted texture pattern is then classified using SVM. The work classified focuses on major diseases commonly observed in Grapes plant which are downy mildew  black rot.

5.In[5],author has presented a SVM-based Multiple Classifier System for pattern recognition of wheat leaf diseases. Further author has used stacked generation structure and Mid level feature generation to improve the performance of recognition of disease of the wheat plant. The proposed approach has obtained better success rate of recognition.

## III. EXISTING SYSTEM

General approach for detecting the leaf disease is composed of two steps, image processing and classification of disease using machine learning techniques. Image acquisition, Image pre-processing, Image segmentation and Feature extraction steps of image processing are discussed in this section. Several classification techniques are:

1] Image Processing: Various phases for detecting diseases using Image processing are Image acquisition, Image pre-processing, Image segmentation, and feature extraction.
Image acquisition: The process of acquiring images from various sources is known as image acquisition.
Image pre-processing: The aim of pre-processing is to improve the quality of image by removing unwanted noise from the image.
Image Segmentation: Image segmentation is used to divide the image into various segments.
Feature extraction: Feature extraction is a type of dimensionality reduction technique, which efficiently represents the most informative parts of the image.

2] Machine Learning: Machine learning explores the study and construction of algorithms that learns from and makes predictions on the given data. In machine learning, mainly classification techniques are available for detecting or classifying diseases. There are many techniques available for classification. Machine learning tasks are broadly classified into 3 broad categories like supervised learning, unsupervised learning, and reinforcement learning.
1) Support Vector Machine (SVM): Support vector machine is based on concept of decision planes that defines the decision boundaries to separate data.
2) K-Nearest Neighbor (K-NN): KNN algorithm is used for classifying objects based on closest training examples in the feature space.

3) Neural Network (NN): The NN is a computing system made up of simple, highly interconnected processing elements, which process their information using states to external inputs.

## IV. PROPOSED SYSTEM

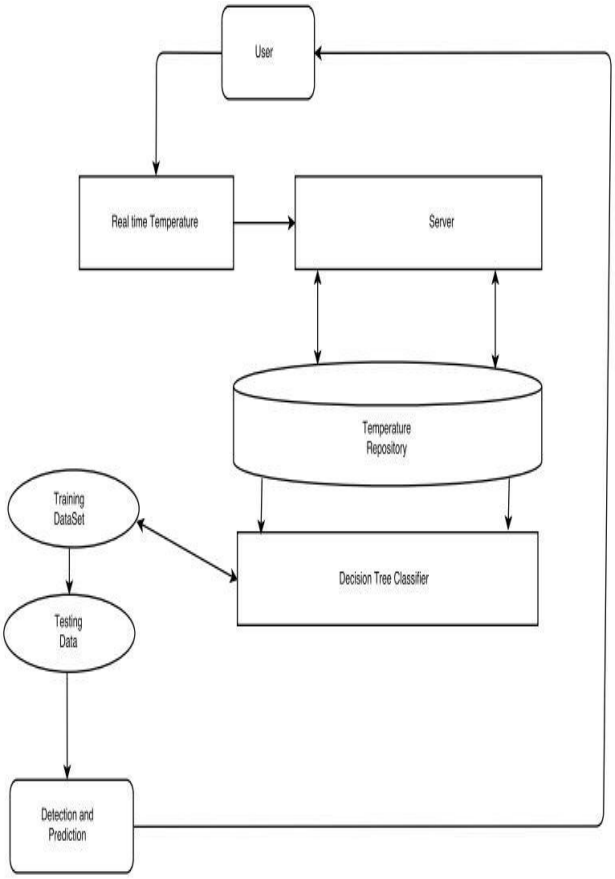



Fig.2. System Architecture

In this proposed system the user will provide real time temperature data to server. The server then authenticates it and transfers it to Temperature Repository i.e Database ,where the user input values are compared with available training dataset and then the decision tree classifier algorithm will predict the relevant output and display it to user.

## V. HARDWARE ARCHITECTURE

After the prerequisites have been achieved, the detailed specifications for the hardware, software, people, and data resources, and the information products that will meet the functional requirements of the proposed system can be determined.The design will act as a guidance for the system and will help detecting problems of the errors or problems which are into the final system. Experts creates the system design, but must review their work with the users to ensure that the design meets users' needs.

Environmental temperature data and soil temperature data will be collected by input sensors. The sensors will provide gathered information to Arduino Uno Board, which in turn will send data to Thingspeak server. The data will be retrieved in csv format to the Decision Tree Classifier, which will predict the relevant result and provide it to the user.

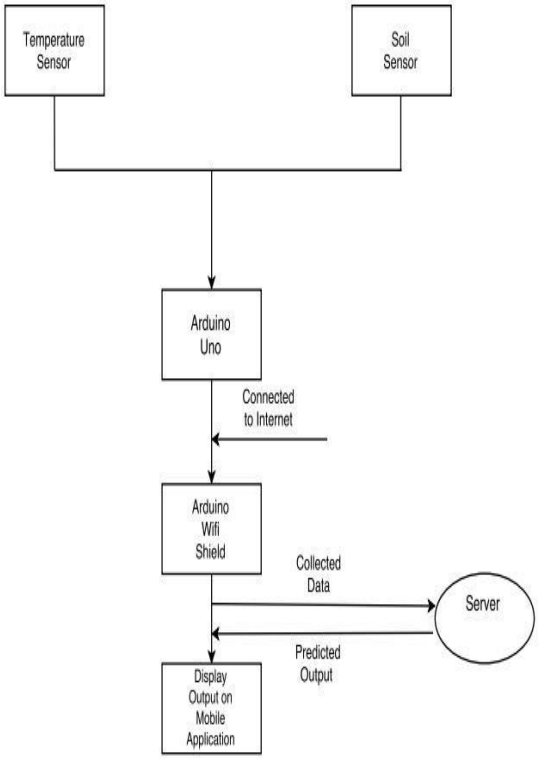



Fig.3. Hardware Architecture

Technical Specification of Components The main components of the system are explained below: Temperature and soil moisture sensors are used. Human input which consists of their past experinces is also considered through mobile phone applications.

Table I: Sensor Details

| Sensor | Name | Manufacturer |
|---|---|---|
| Temperature | DHT 11 | Kitsguru |
| Soil Moisture | KG003 | Kitsguru |

*Arduino Uno R3[10]*

- Microcontroller: ATmega328
- Operating Voltage: 5V
- Input Voltage (recommended): 7-12V
- Input Voltage (limits): 6-20V

- Digital I/O Pins: 14 (of which 6 provide PWM output)
- Analog Input Pins: 6
- DC Current per I/O Pin: 40 mA
- DC Current for 3.3V Pin: 50 mA
- Flash Memory: 32 KB of which 0.5 KB used by boot- loader
- SRAM: 2 KB (ATmega328)
- EEPROM: 1 KB (ATmega328)
- Clock Speed: 16 MHz

*ESP8266 Serial Wifi Wireless Transceiver Module[11]*

- Operating Voltage: 3.0 3.6V
- Operating Current Average value: 80mA
- Operating Temperature Range: -40 125
- Ambient Temperature Range: Normal temperature
- Package Size: 5x5mm
- External Interface: N/A
- WiFi mode: station/softAP/SoftAP+station
- Security: WPA/WPA2
- Encryption: WEP/TKIP/AES

*Thingspeak Server[17]*

- Collect data in private channels
- Share data with public channels
- RESTful and MQTT APIs.
- MATLAB analytics and visualizations
- Alerts
- Event scheduling
- App integrations
- Worldwide community

*Temperature Sensor -DHT11[12][13]*

- PCB size: 22.0mm X 20.5mm X 1.6mm
- Working voltage: 3.3 or 5V DC
- Operating voltage: 3.3 or 5V DC
- Measurement range: 20-95
- Supply Voltage: +5 V
- Temperature range :0-50 C error of 2 C
- Humidity :20-90
- Interface: Digital

*Soil Moisture Sensor -Kitsguru KG003[14]*

- Operating Voltage: 3.3V or 5V
- Sensing Probe Dimensions: 60x30mm
- Panel PCB Dimensions: 30 x 60mm
- On-board LM393 comparator
- On-board power indicator LED
- On-board digital switching indicator LED

## VI. IMPLEMENTATION DETAILS

Implementation of this project is done on small scale currently. Once all the functionalities are implemented the project will be expanded for large scale. Decision Tree Classifier algorithm is the heart of the proposed system. It is a simple and widely used classification technique. The benefit of using this algorithm is it's simplicity as it applies a straightforward idea to solve the classification problem. It implies a set of questions about the attributes of the test record. Every time it gets an answer a subsequent question is asked prior to the levels of class labels is achieved in the database. The Anaconda software alongwith Jupyter Notebook is used to predict the suitable output. The main steps of the project are as follows:

- Acquire farming data using sensors and human input through mobile applications.
- Push the collected data into Thingspeak server.
- Analyse and predict the collected data using Decision Tree Classifier.
- Develop mobile app for farmers for crop related predictions and alerts

*A.Steps Involved in Decision Tree Classifier*

Training Data is selection of certain percentage of total database alongwith testing data which is used to train the model. The better the training data higher the efficiency of model. Fig 4 shows Training Data

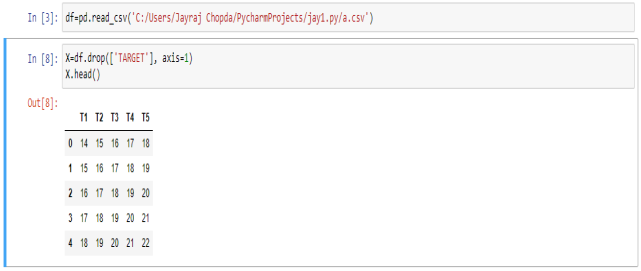



Fig.4. Training Data

After the model is trained, it is analyzed on test set. Generally, these sets are taken from the same total database, though the training set should be labeled or enriched to increase an algorithm's confidence and accuracy. Fig 5 shows Testing Data:


Fig.5. Testing Data

Further testing data, next phase comes of prediction of required output which is achieved by Decision Tree Classifier model.

Table II: Cotton Crop Disease Labels

| Type | Label |
|---|---|
| Wilt | 0 |
| Anthracnose | 1 |
| None | 2 |

## VII. SAMPLE OUTPUT

In previous section, training data which is shown in Fig 4 is used to predict the sample output. Training data is needed to make the Decision Tree Classifier to learn from it and evaluate experience. Testing Data shown in Fig 5 is used to assign labels in Decision Tree Classifier for type of disease. Prediction occurs with the help of Decision Tree Classifier by taking suitable temperature range from the user, analyze it with the Training data and provides significant output. Following shows the sample output of the prediction of Decision Tree Classifier:

clf_gini.predict([29,30,31,32,33])
array([1], dtype=int64)
The array([1]) indicates that there is a risk of Anthracnose.

clf_gini.predict([25,26,27,28,29])
array([0], dtype=int64)
The array([0]) indicates that there is a risk of Wilt.

clf_gini.predict([33,34,35,36,37])
array([2], dtype=int64)
The array([2]) indicates that there is no risk of any disease.

## VIII. CONCLUSION

The proposed system can predict the cotton crop diseases using decision tree with the help of the parameters like temperature, soil moisture, etc. based on the previous year data and through sensors. The future scope of the project would be focusing on building an Android Application.

## REFERENCES


[1] A Survey on Detection and Classification of Cotton Leaf Diseases,Bhumika S. Prajapati, Vipul K. Dabhi, Harshadkumar B. Prajapati

[2] Advance computing enrichment evaluation of cotton leaf spot disease detection using Image Edge detection,P. Revathi, M. Hemalatha

[3] eAGROBOT- A Robot for Early Crop Disease De- tection using Image Processing,Sai Kirthi Pilli, Bharathiraja Nallathambi, Smith Jessy George, Vivek Diwanji

[4] Detection and classification of diseases of Grape plant using opposite colour Local Binary Pattern feature and ma- chine learning for automated Decision Support System,Harshal Waghmare, Radha Kokare, Yogesh Dandawate

[5] SVM-based Multiple Classifier System for Recognition ofWheat Leaf Diseases,Yuan Tian,Chunjiang Zhao,Shenglian Lu

[6] http://www.mdpi.com/journal/agriculture/special_issues/ plant_disease

[7] http://ikisan.in/mh-cotton-disease-management.html

[8] http://i0.wp.com/thomascountyag.files.wordpress.com/ 2015/07/angularleafspot-009.jpg?ssl=1

[9] https://www.kisansuvidha.com/wp-content/uploads/ 2017/07/grey-mildew-of-cotton.jpg

[10] http://www.hobbytronics.co.uk/arduino-uno-r3

[11] https://cdn-shop.adafruit.com/product-files/ 2471/0A-ESP8266__Datasheet__EN_v4.3.pdf

[12] https://www.itead.cc/wiki/ DHT11_Humidity_Temperature_Sensor_BrickSpecifications

[13] https://www.dfrobot.com/wiki/index.php/ DHT11 Temperature_and_Humidity_Sensor_(SKU:_- DFR0067)

[14] https://www.amazon.in/KitsGuru-Moisture- Sensor-Module-Arduino/dp/B00XU8MJ4E/ref=sr_1 1?ie=UTF8qid=1509719422sr8keywords=soil+moisture +sensor

[15] https://www.crowdflower.com/what-is-training-data/

[16] Farm Data Analytics Using Splunk, International Journal of Advanced Studies in Computer Science and Engineering, UGC (India) approved Journal , Volume 6 , Issue 9, 2017, 83-87. Available at http://www.ijascse.org/volume-6-theme-based-issue-9/Splunk.pdf

[17] https://thingspeak.com/