Disorder Detection of Tomato Plant using IoT and Machine Learning

Abstract: India is an agricultural country and this sector accounts for 18 percent of India's GDP. This sector is the backbone of the country and focuses on better yield by using pesticides and fertilizers to prevent plant disorders which directly affects the yield. The primary method adopted for detecting disorders is through visual observation and other methods are quite expensive. Many authors have proposed solutions to this problem such as IoT for grapes, or system designed for accurate disorder detection using machine learning with limited scope. This paper showcase a prototype that uses multi-modal analysis through sensor data, computer vision. The main objective of this system is to accurately detect disorders in tomato plant using IoT, Machine Learning, Cloud Computing, and Image Processing.

Keywords: Plant Disorder, Feature Extraction, Segmentation, Deep Learning, Convolutional Neural Networks (CNN), IoT.

Reference to this paper should be made as follows: Saiqa Khan and Meera Narvekar (2019) 'Disorder Detection of Tomato Plant using IoT and Machine Learning', *Int. J. Data Mining, Modelling, and Management,*

1 Introduction

Global warming has been increased due to man-made activities over the past few decades. Due to such activities, it gave rise to uncertain climatic conditions. Moreover, these unusual climatic conditions influence all the major aspects of plant growth which includes soil fertility, temperature and cropping intensity. Infertile soil is not preferable for crops, because of which fertilizers are used as they contain all the much-needed nutrients such as potassium, nitrogen, and phosphorus. This paper showcase a system which can be used to detect disorders found in Solanum Lycopersicum plant commonly known as Tomato plant [3] which belongs to the nightshade family, Solanaceae [1]. According to research, leaves are the most affected parts of the plant in case of any disorder. Their properties provide important insight into the identification of the disorder and its current status. Our system focuses on both biotic and abiotic factors affecting the growth of the plant. In biotic disorders, the system is mainly concentrating on two disorders commonly seen in a tomato plant, early blight, late blight and one class of healthy. Temperature, humidity, and soil moisture are abiotic factors. Feature extraction is the major step in our system where patterns from image and sensor readings are learned by our machine learning algorithms [4]. The sensor's data has been collected using IoT and images are captured manually and processed further. Image data is collected from multiple sources like Plant Village Dataset, Real world images captured at the farm and some images downloaded from the internet. The system reads live sensor data and leaf image and predicts whether the plant is healthy or not and if not it also predicts which disorder is present among the three classes.

2 Related Work and Motivation

In the past, many researchers have worked on many techniques on identifying features in image data. Stephan Gang Wu et al. [5] used a probabilistic neural network along with leaf image and data pre-processing to implement leaf recognition system. Harish Velingkar et al. used feature extraction techniques like K-Means clustering algorithm for clustering important colors and then SVM for classification [6]. In [7] and [11] the authors used CNN for raw feature extraction from leaf images. Alvaro Fuentes et al [12] used multiple deep learning architectures like Faster Region-based Convolutional Neural Network (Faster R-CNN), Region-based Fully Convolutional Network (R-FCN), and Single Shot Multibox Detector (SSD). In [8] Mohammed A. Hussein Amel H. Abbas has considered various feature extraction methods like Texture-Based Features, Colour Moments, Shape-Based Features, etc. for classifying infected leaf images. Sanjay Mirchandani et al in [9] used simple Fully Connected layers (ANN) like feed-forward network with basic image processing. Jihen Amara et al used CNN architecture called LeNet [10] Architecture for Leaf classification. T. Rumpf et al. in [14] classified three diseases Cercospora leaf spot, leaf rust and powdery mildew in Sugar Beet Leaves using Support Vector Machines (SVM). The major drawbacks of all this work are that they only consider visual aspects of leaf i.e. leaf image. Only the visual characteristics of the leaf is not a suitable measure for determining the plant condition.

3 Materials and Methods

This paper describes the complete walkthrough from data collection to the building system. Different image processing, feature extraction, and dimensionality reduction techniques are used to get insights from both sensors and image data. All the steps carried out are discussed in details as follows.

Fig (1) Tomato Plants Setup at MHSS College, Mumbai, Maharashtra, India

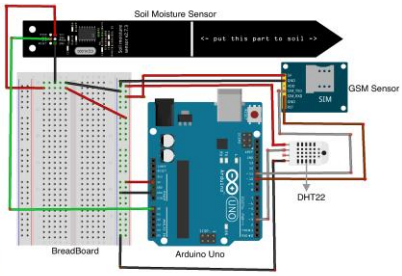


Fig (2) IoT Circuit for Data Collection

1. Sensor Dataset

Sensors data like temperature, humidity, and soil moisture are being collected from setup of 10 tomato plants located at M. H. Saboo Siddik College, Mumbai with latitude 18.9685103° N, longitude 72.8288362° E, temperature 29°C and Humidity: 69%. The setup of 10 tomato plants can be seen in fig (1). The IoT circuit diagram for data collection is shown in fig (2). The data plot in fig (3) shows the variation in abiotic parameters over a day.

Fig (3) Abiotic Parameters data plot for date 16th Feb, 2019(16:00pm to 23:00pm)

2. Image Dataset

The dataset consists of leaf images from Plant Village Dataset, internet images, and leaf images captured from Tansa Farm, Bhiwandi. There are three classes early blight, late blight and healthy. The comparison from the different sources is shown in fig (4).

Fig (4) Image Dataset Samples from Different Sources

A. Data Preprocessing

Data preprocessing is an important step, as real-world data comes with a lot of variation, outliers, and unexpected values. To make the predictions of the system accurate, data needs to be scaled down to a standard format.

1. Sensor Data Preprocessing

Real-world sensor data comes with some redundancies like missing values or NaN values, values over the threshold, etc. need to be handled before further processing. Upper bound values are clipped in this step and missing values or NaN are replaced with the mean value of that specific parameter.

1. Image Data Preprocessing

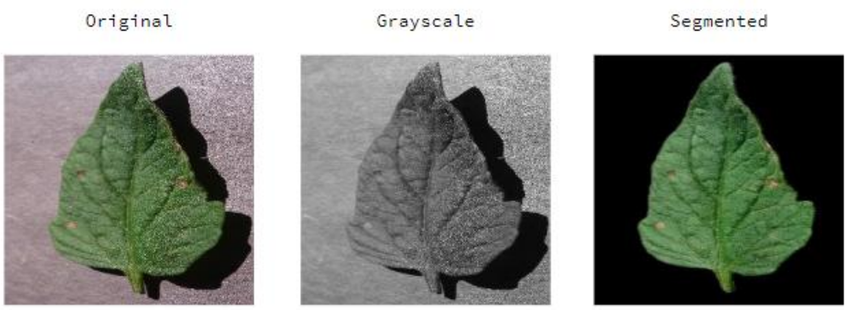


Fig (5) Segmentation of Leaf Image

In the case of leaf image data, leaf images are not always perfect as required by the model. The leaf image is pre-processed to remove the background and mainly concentrated on segmenting the green leaf part to train the deep learning model as shown in fig (5). The green leaf part is segmented by converting RGB image to HSV image and selecting green color Hue value range, to segment only the green leaf and rest with Black Colour as mentioned in [19].

B. Feature Extraction and Training

Most of the past and available feature extraction techniques only consider visual aspects of leaf and try to extract necessary information from them. The system considers both, sensor's data and leaf image. It maps the semantic representation between the visual properties and environmental parameters (Humidity, Temperature, Soil Moisture). Features are divided into two categories 1) Sensor-Based Feature Extraction, 2) Deep Learning Based Feature Extraction.

1. Sensor Based Feature Extraction and Training

The environmental conditions play a vital role in determining the health of the plant. The abiotic factors like temperature, soil moisture and humidity helps to determine whether the plant is growing in healthy conditions or not. The system uses two sensors soil moisture and temperature-humidity sensor. This data is gathered using IoT and stored on a cloud. All the data will be used in machine learning algorithms to predict new samples. It has two classes healthy and not healthy. Supervised learning algorithms like SVM and Random Forest Algorithms [13] are used as these algorithms have a good performance on statistical data. Unsupervised Learning Technique K-Means Clustering is also used to learn from the abiotic factors and form clusters. The block diagram for the sensor's based feature extraction and training is shown in fig (6).

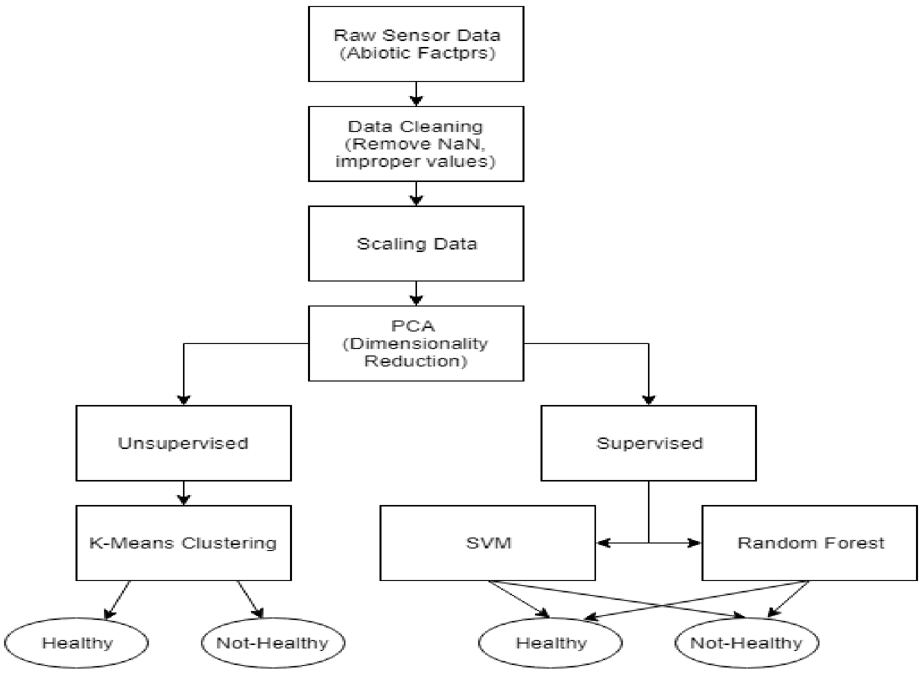


Fig (6) Block Diagram of Sensor Based Feature Extraction and Training/Prediction

2. Deep Learning Based Feature Extraction and Training

To train the model, images must be pre-processed by carrying out resizing, noise removal and segmentation. The image processing techniques are carried out using the OpenCV library in Python [20] which was developed by Intel. For training is performed using Keras Library with Tensorflow Backend [21] and this whole dataset is trained on Google Colab Platform. Keras [18] package supports various state-of-the-art pre-trained deep learning models ready for classical machine learning problems. For precise learning, using pre-trained models gives a huge boost in learning and prediction. Mohanty et al. [16] used deep learning for leaf image classification on Plant Village Dataset using AlexNet with 99.5% accuracy. Aravind et al. used pre-trained AlexNet and VGG16 [15] for leaf disorder classification with 86% accuracy on the testing set. Using Transfer Learning, the images are trained on pre-trained models by fine-tuning like VGG16 [17], VGG19. These are the CNN based architectures which are best for image classification based problems. These pre-trained models are trained on a huge dataset called "ImageNet" with 1000 classes and can be retrained by freezing some of its layers on a new dataset with a new number of classes. Both the architectures have been fine-tuned by adding one Convolution Layer and three Dense layers followed by softmax activation. The system classifies two biotic disorders in a tomato plant, Early Blight, Late Blight and

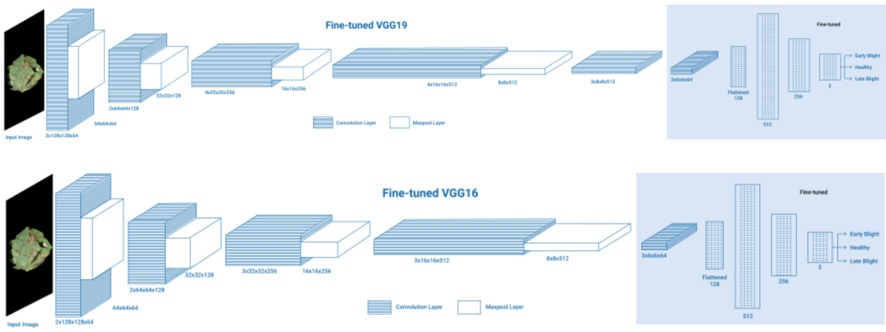


Fig (7) Fine-tuned VGG16 and VGG19 Model

4 Results

1. Tomato Health Detection Using Sensors

The dataset contains 5923 samples with timestamps, soil moisture value, temperature, and humidity. The dataset has been manually classified for training. The stats of dataset are given in Table 1. The performance of various algorithms is displayed in Table 2. The unsupervised algorithm, K-Means Clustering clusters visualization is shown in Fig (8). The dataset was also trained on multiple machine learning models like Support Vector Machines and Random Forest Classifiers. K-means required pre-processing steps like standard scaling and Principle Component Analysis (PCA).

Table 1 Dataset Stats

Target	*Training*	*Testing*	*Total*
Healthy	3427	381	3808
Not-Healthy	1903	212	2115

Table 2 Accuracy on Various Algorithms

Algorithm Accuracy	*Training Acc.*	*Testing Acc.*
SVM	99.45%	99.3%
Random Forest	99.56%	99.6%
K-Means	99.2%	99.5%

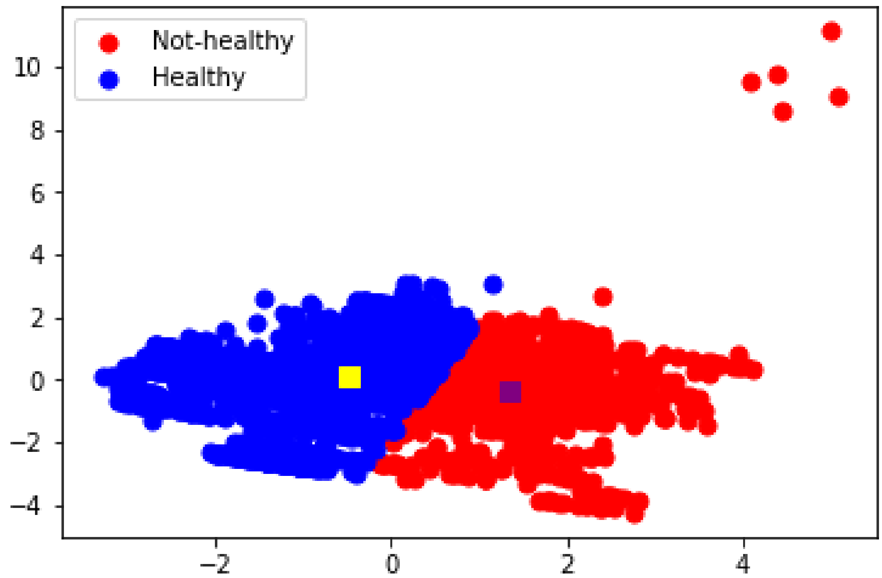


Fig (8) K-Means Clustering of Sensor's Data

2. Tomato Disorder Detection

The leaf image dataset contains 5,838 real-world images with unbalanced classes. The dataset has three classes Early Blight, Late Blight and Healthy. The stats of image dataset are given in Table 3. The fined tuned architecture of VGG16 and VGG19 is shown in fig (7). As shown in fig (10) VGG16 training accuracy is greater than VGG19 and training loss is minimum in VGG16 architecture. On validation set, VGG16 outperforms VGG19 architecture with high validation accuracy and low validation loss as compared to VGG19. Table 4 showcases the results of VGG16 and VGG19 on training and test set. The dataset was also trained over a Vanilla CNN Image Classifier which performs poor compared to VGG Architecture as shown in Fig (9).

Table 3 Leaf Image Dataset Stats

Class Name / Label	***Training Set***	***Testing Set.***	***Total***
Early Blight	1500	500	2000
Late Blight	1595	500	2095
Healthy	1401	342	1743

Table 4 Accuracy on different architectures

DNN Architecture	***Training Acc.***	***Testing Acc.***
VGG16 + fined tuned	93.24%	92.08%
VGG 19 + fined tuned	88.48%	86.27%
Vanilla Architecture	78.12%	80.24%

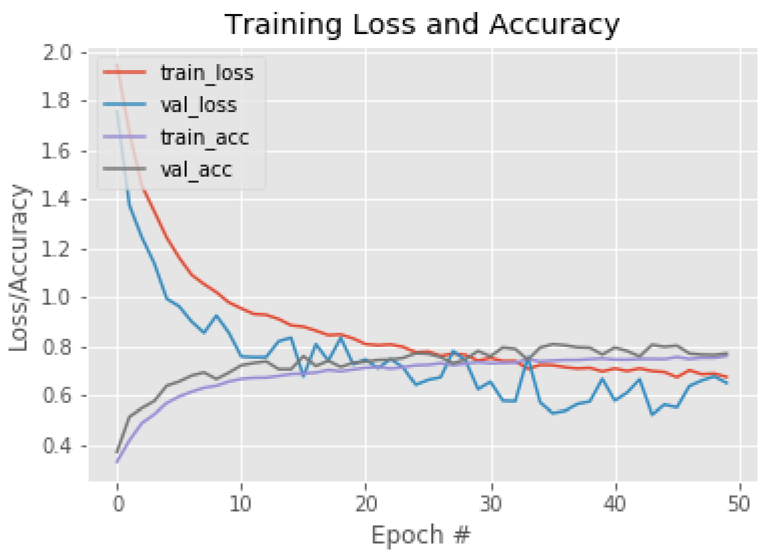
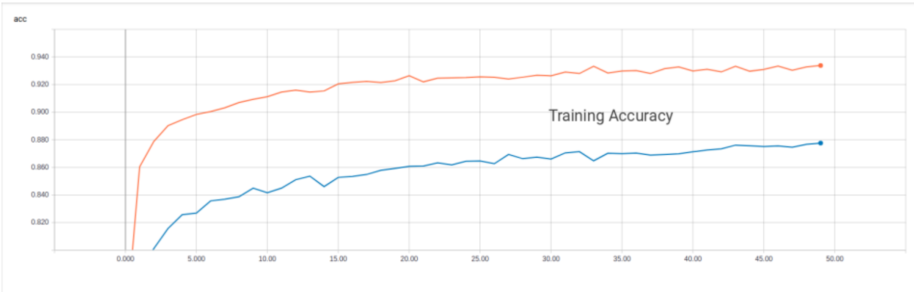
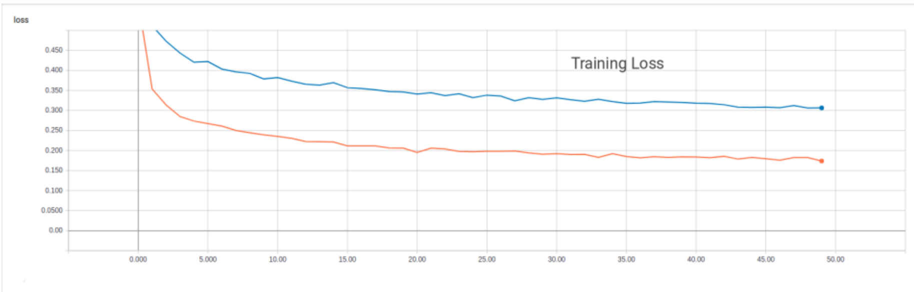


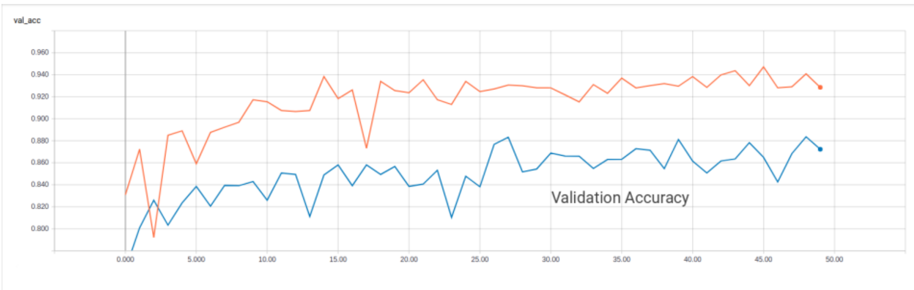
Fig (9) Vanilla CNN Network Performance



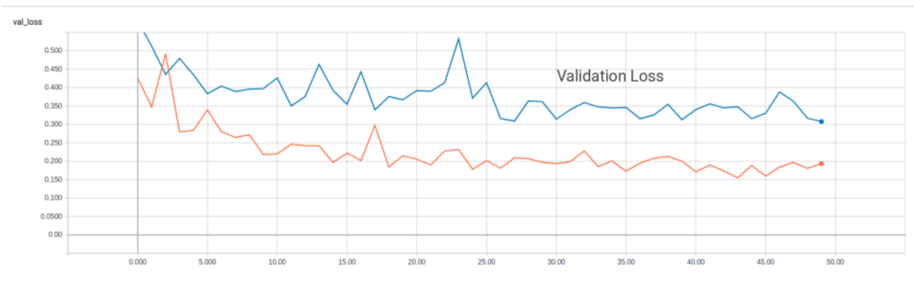
Training Accuracy



Training Loss



Validation Accuracy



Validation Loss

Fig (10) Benchmarks of VGG16(Orange) and VGG19(Blue)

For leaf image, dataset fined tuned VGG16 outperforms fined tuned VGG19 architecture.

Author

The screenshots of the Plant Monitoring Web App are shown in fig (11).

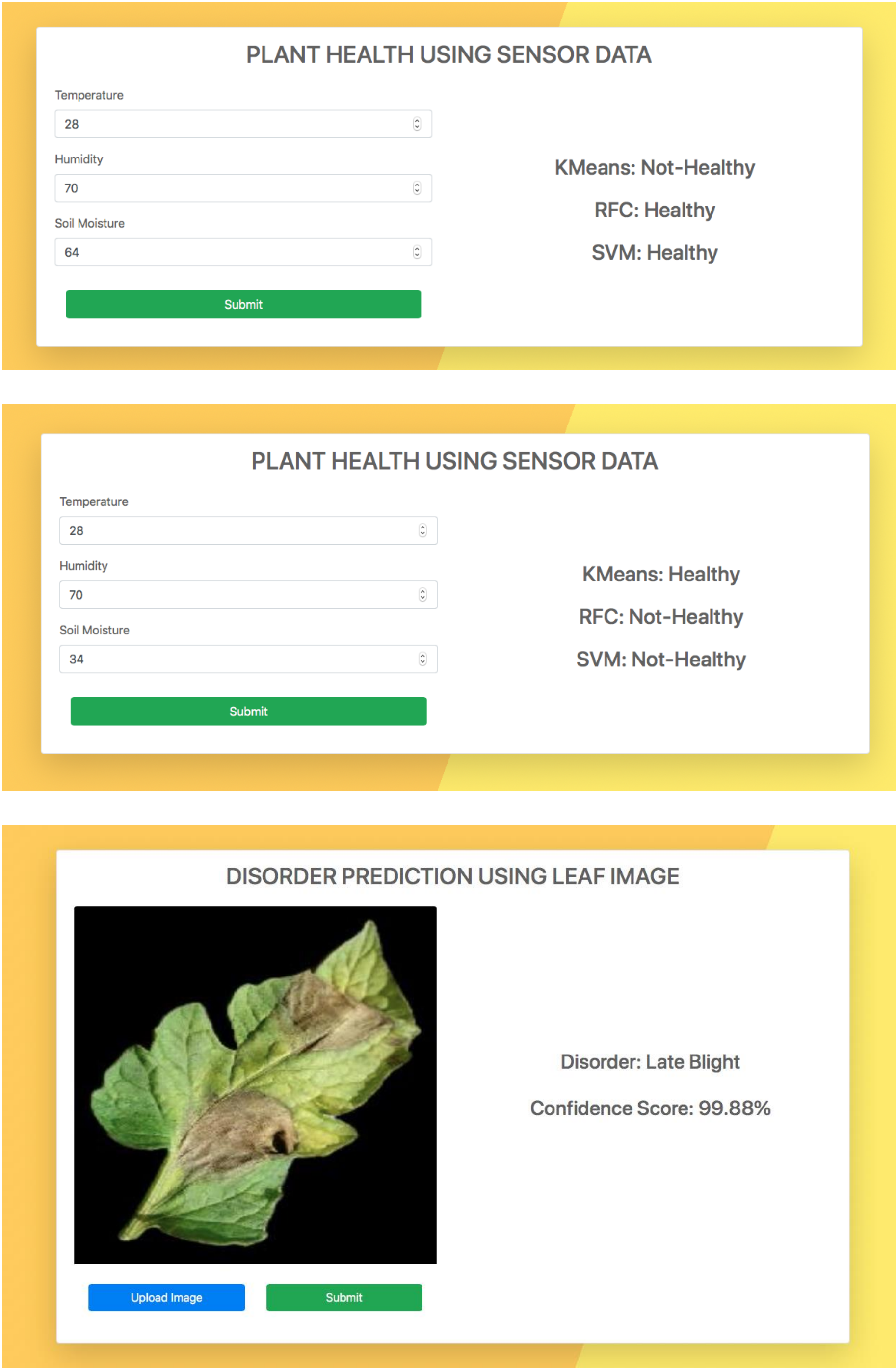


Fig (11) Screenshots of Plant Monitoring System WebApp

5 Conclusion

The past work on this field was mainly focused on classifying data using visual properties of leaf image using pattern recognition and Deep Neural Networks. The system is trained with real-world sensor data and image data. Both the results can be used for ensemble prediction for better results.

References

[1] M.M Peet, G Welles (2005). Green House Tomato Production. *CAB International 2005. Tomatoes (ed. E. Heuvelink).*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Tomatoes, Advisory Committee on Vegetable Crops. Vegetable Crops Production Guide. *Services Coordinating Committee*. [online] Available at https://www.faa.gov.nl.ca/agrifoods/plants/pdf/tomatoes.pdf.s

[4] Gaurav Kumar, Pradeep Kumar Bhatia (2014). A Detailed Review of Feature Extraction in Image Processing Systems. IEEE Explore. 2014 fourth international conference on advance computing and communication technology

[5] Gang Wu, S., Sheng Bao, F., & You Xu, E. (n.d.). A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network-IEEE Conference Publication. Retrieved from https://ieeexplore.ieee.org/ abstract/document/4458016/

[6] Shambhuraj Porob, et al. “Figure 2f from: Irimia R, Gottschling M (2016) Taxonomic Revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: e7720. Https://Doi.org/10.3897/BDJ.4.e7720.”

[7] “How Deep Learning Extracts and Learns Leaf Features for Plant Classification.” NeuroImage, Academic Press, 18 May 2017, www.sciencedirect.com/science/article/pii/ S003132031730198X.

[8] Hussein, Mohammed A., and Amel H. Abbas. “Comparison of Features Extraction Algorithms Used in the Diagnosis of Plant Diseases.” Ibn AL-Haitham Journal For Pure and Applied Science, www.jihcoed.com/ihj/ index.php/j/article/view/1785.

[9] Mirchandani, Sanjay, *et al.* “Figure 2f from: Irimia R, Gottschling M (2016) Taxonomic Revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: e7720. Https://Doi.org/10.3897/BDJ.4.e7720.”

[10] Amara, Jihen, *et al.* “Figure 2f from: Irimia R, Gottschling M (2016) Taxonomic Revision of Rochefortia Sw. (Ehretiaceae, Boraginales). Biodiversity Data Journal 4: e7720. Https://Doi.org/10.3897/BDJ.4.e7720.” A Deep Learning-Based Approach for Banana Leaf Diseases Classification, 2017, doi:10.3897/bdj.4.e7720.figure2f.

[11] Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. “Deep Learning in Agriculture: A Survey.” Computers and Electronics in Agriculture, Vol. 147, 2018, pp. 70–90., doi:10.1016/j.compag. 2018.02.016.

[12] Fuentes, Alvaro, *et al.* “A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition.” Sensors, Vol. 17,no.9,Apr.2017,p.2022.,doi:10.3390/s17092022.

[13] Ferdoush, S. and Li, X. (2014). Wireless Sensor Network System Design Using Raspberry Pi and Arduino for Environmental Monitoring Applications. *Procedia Computer Science*, 34, pp.103–110.

[14] Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectanceT.Rumpf, A.-K.Mahlein, U.Steiner, E.-C.Oerke, H.-W.Dehne, L.Plümer

[15] Rangarajan, Aravind Krishnaswamy, Raja Purushothaman, and Anirudh Ramesh. "Tomato crop disease classification using pre-trained deep learning algorithm." *Procedia computer science* 133 (2018): 1040-1047.

[16] Mohanty, Sharada P et al. “Using Deep Learning for Image-Based Plant Disease Detection” *Frontiers in plant science* vol. 7 1419. 22 Sep. 2016, doi:10.3389/fpls.2016.01419

[17] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

[18] Gulli, A., & Pal, S. (2017). *Deep Learning with Keras*. Packt Publishing Ltd.

[19] Sural, Shamik, Gang Qian, and Sakti Pramanik. "Segmentation and histogram generation using the HSV color space for image retrieval." *Proceedings. International Conference on Image Processing*. Vol. 2. IEEE, 2002.

[20] Howse, Joseph. *OpenCV computer vision with python*. Packt Publishing Ltd, 2013.

[21] Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 2016.