www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC
Gestão, Inovação e Tecnologias

Smart Proactive Warning System for Early Prediction of Potato Late Blight Via Machine Learning

Samer I. Mohamed[1]

[1]Associate Professor, October University for Modern Sciences and Arts MSA.

[1]dr.samer.ibrahim@gmail.com

Abstract

Precision agriculture is one of the applications which may use the Machine learning technology to maximize the crops productivity, optimize the quality of the crops, and minimize the negative environmental impact like crops diseases. Crop diseases are generally one of the most critical problems that threaten the worlds agriculture, causing large losses in agricultural production of about 25% per year. Some metrological data like relative humidity, temperature, wind speed, wind direction, and pressure play an essential role in the creation and spreading of late blight disease of potato. The proposed system aims to predict a potato late blight using a machine learning (ML) IoT based system trained on dataset collected from field sensors at crops environment through forecasting weather data from weather stations. The proposed system is designed based on Logistic regression and Neural Networks where model results show performance and accuracy improvements compared with the result of built in service created on Microsoft Azure cloud using the same data set.

Key-words: Machine Learning (ML), Potato Late Blight, Precision Agriculture, Binary Neural Networks, Logistic Regression, Warning System.

1. Introduction

Most of the field crops cultivated in Egypt suffer from a severe shortage in their production rates. Globally, plant diseases are generally one of the most critical problems that threaten the world's agriculture and food, causing large losses in agricultural production of about 25% per year, equivalent to the consumption of about 600 million people [5]. Potato is one of the most important crops locally in Egypt and globally because it is used for local consumption and exporting.

There are around 8961 thousand acres planted, 215 thousand acres of it cultivated with potato, producing around 2.2 million tons of potato with average 11.3 tons per acre, which makes Egypt the largest potato producer in Africa [1]. However, that is not enough for future need according to [2] the Egyptian consumption of Potatoes were increasing to 115.3% within the last 13 years.

Potato late blight is one of the most important diseases on potato, according to [3]. It is estimated that the yearly economic losses, through crop loss and the cost of control measures, in Europe due to this disease exceeds €1 billion. In order to increase the production and reduce the loss of potato it is important to manage the diseases that hit the potato plant by collecting data from the sensors in the field (relative humidity, temperature, wind direction, wind speed, and pressure) and forecasting data for the following five days for the prediction purposes [4]. This will help the grower to reduce the cost of losing and reduce the cost of the Pesticide by knowing the exact time and sector that will be infected by the potato late blight disease rather than losing much money in spraying the Pesticide in the whole farm [11].

2. Background

Dahikar and Rode [13] used Artificial Neural Network (ANN) as the proposed method for the prediction of some crops in a therein area based on weather parameters as well as soil parameters. The weather parameters inputs are relative humidity, temperature, rainfall, etc. the soil inputs are the type of soil, sodium, hydrogen, PH, carbon, potassium, etc. The simulation tool used is MATLAB. Feedforward and backpropagation algorithms are used. Gradient descent is the method used in backpropagation for error calculation. The proposed algorithm is proved for some crops like rice, wheat, cotton, soybean, and sugarcane [9].

Dhaka and Lamba [14] presented the technique of forecasting that has been used in wheat crop. It shows some of the development and research that has been done in the field of forecasting systems. It shows the difference between forecasting models like meteorological, simulation, remote sensing, and mathematical. The paper describes the models of Neural Networks (NN) also [10]. The study compares NN algorithm with other algorithms. The result of the paper shows that the NN algorithm is one of the most important tools for the prediction of non-linear data. A lot of systems have been studied in this paper, showing that NN can be more accurate when the data is nonlinear [7].

Henderson et al. [15] in this paper studied the useful weather parameter for forecasting potato late blight in southern Idaho. This paper used logistic regression for classification for late blight of potato (0 means no disease appears in the fields, and 4 means most of the fields get the disease). It also shows the favorable hours for the creation of the late blight and the amount of precipitation. In the binary Logistic regression, it predicts getting a disease with an accuracy of 67.5% using a leave-1-

year-out error estimate [6]. The mode was validated from “semi-arid Columbia Basin regions” the disease prediction was 80.8%, although the advancements achieved from the mentioned systems, the scalability, operability and maintainability of the solution still a challenge [8].

The most commonly applied system is a system called disease forecasting or disease management for plants [4], this system monitors the environment of the farm to see if there are any changes in some parameters that will affect the plant, this parameter is read from the farm by temperature, relative humidity, and rainfall sensors. This system is not accurate [12] due to the direct sunlight of the sun, which impacts the temperature sensor, the calibration of the relative humidity sensor may deviate, and the "rain gauge tipping" may be coagulated due to the spider web, and it is not well leveled, and data-loggers perhaps exhaust their batteries at crucial times.

The basis of the network in the field was weak; most of the systems work only in nearly flat terrain. Some systems do not have any prediction about the future that can help farmers to take safe action before any losses. For detection systems, there are different systems applied to detect the disease with different image processing techniques [20], but the accuracy of the system still needs to increase to be applicable, applied detection systems do not link the appearance of the disease with another environment automatically.

Late blight of potato is a potato disease that is caused by the water mold *Phytophthora* infectants or fungus. *Phytophthora* survives in stored tubers, dump piles, and files plants. Sporangia is dispersed by the wind to nearby plants where infection may occur within a few hours. When weather conditions are favorable, the fungus spread very rapidly especially during cool, wet weather. Spread of late blight occurs when spores are produced on infected potato leaves, then the spores are then carried to healthy tissue by rain or wind.

According to [19], “Temperature and moisture are the most important environmental factors affecting late blight development. Sporangia are formed on the lower leaf surfaces and infected stems shown in *Fig.1* when relative humidity is greater than 90%. Sporulation can occur from 3-26°C (37-79°F), but the optimum range is 18-22°C (64-72°F). Sporangia germinate directly via a germ tube at 21-26°C (70-79°F). Below 18° C (65°F), sporangia produce 6 to 8 zoospores which require water for swimming”.

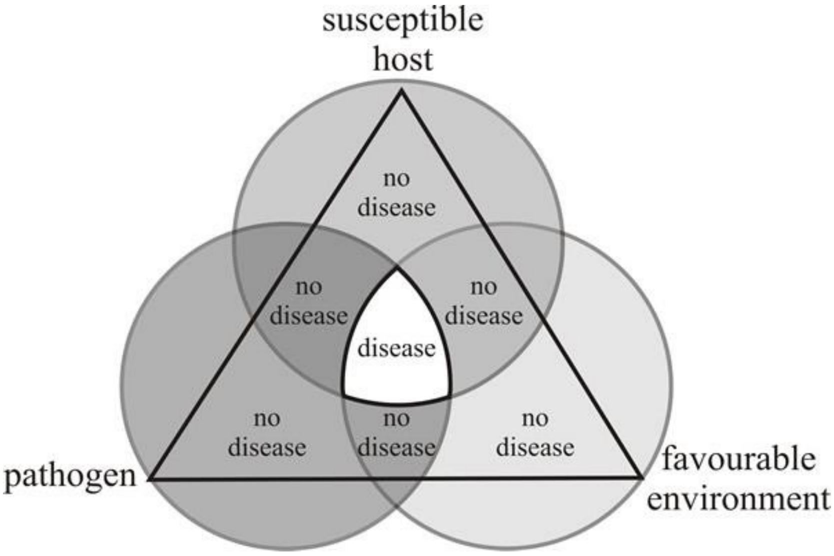
As [17] states “Each zoospore can initiate an infection, which explains why the disease is more severe in cool, wet conditions. Cool nights, warm days, and extended wet conditions from rain and fog can result in late blight epidemics in which entire potato fields are destroyed in less than two weeks. The pathogen can sporulate on contaminated tubers in ineffectively controlled storage zones where conditions are very humid [16].

Condensation produces water droplets on the surface of contaminated tubers which may at that point cause the pathogen to create sporangia and sully neighboring tubers, driving to an annihilation of the complete pile by delicate rot bacteria [18].

There are many different strategies for Late blight that vary between forecasting, varietal, culture, biological and chemical treatment. Disease forecasting is considered the best strategy for disease management as sort of proactive handling for preventing the disease.

Fig. 1 - Late Blight

Fig. 2 - Disease Triangle



The data from the cloud used for machine learning algorithms run locally on the gateway (Raspberry Bi) and externally on the cloud. Image processing and deep learning techniques were used to detect the disease and increase system accuracy [24].

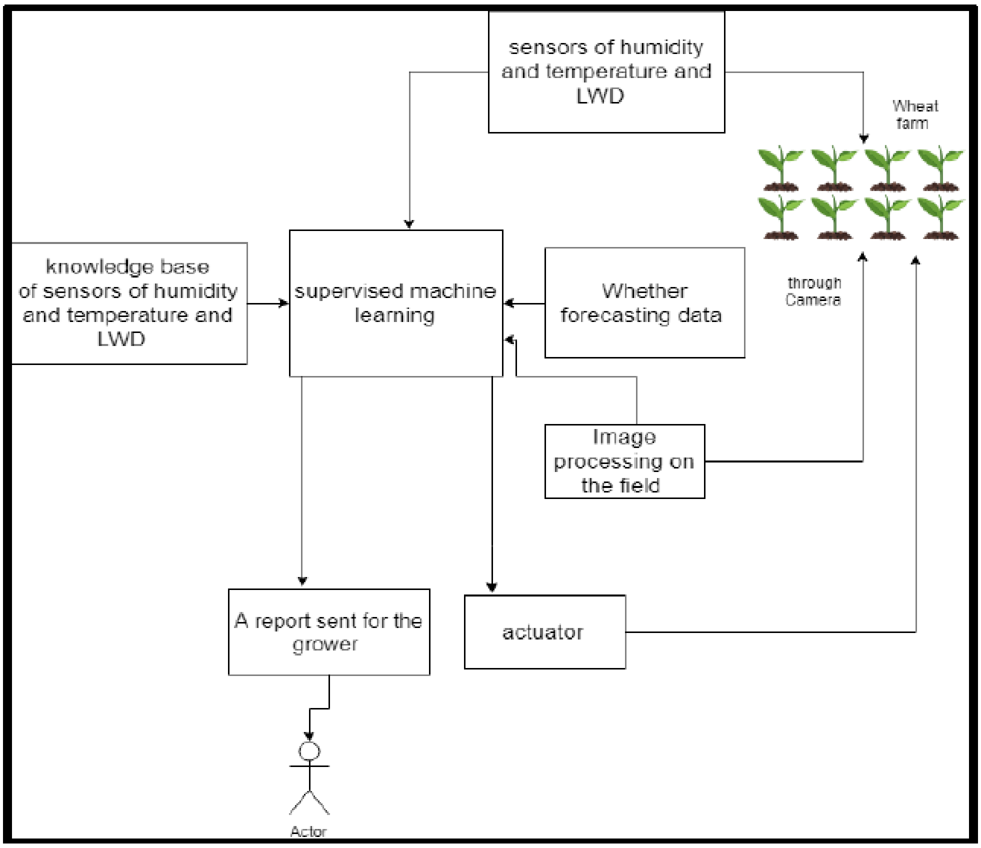
One of the most important factors in the creation of disease is the environment. The disease needs a perfect environment for the formation and spreading, so plant diseases can be analyzed using the concept of Disease triangle [23] shown in Fig.2. There are three conditions which create the disease. Firstly, it is necessary to have a host, and in this case, the host is the plant itself. Some plants can suffer from many diseases' others suffer from special ones. So, each plant has a range of sensibility to a range of diseases. Secondly to have the disease an active pathogen is needed. If there is not any active pathogen, there cannot be a disease. The third element is a favorable environment. The meaning of a favorable environment is the weather conditions like temperature and relative humidity in a therein range needed for a pathogen to develop. Disease happens only if the three conditions occur at the same time; if one or more of these conditions are not existing, the disease does not occur. Early warning of weather change, or the environment of the plant can play a vital rule in the prevention of the creation of the disease. This is the primary disease warning prediction standard [27]. Crop disease prediction systems are mainly based on the interactions between a pathogen, susceptible host plant and the favorable environment in terms of weather conditions.

Although all three parameters (Pathogen, Susceptible host, and favorable environment) are critical for the disease formation, we focused in our proposed disease warning system on sensing the meteorological data that are favorable for the disease formation where we can use a combination of technology and science to visualize what is happening in the atmosphere in each area. Since the pathogen itself depends on meteorological conditions for survival, spreading and ability to infect more plants, it will be with double value to use weather parameters to predict the disease severity. The other main challenge in the disease triangle which has the most effect on the disease formation is the host plant. Over the past decades, it was shown by [15] that host resistance is the best management option that controls the disease susceptibility. However due to drivers' virulence nature of *Phytophthora* infectants, the resistance of the varieties is wiped out over past decades. Many fungicides have been evaluated over different periods across the year; however, the pathogen has shown a remarkable capacity for change with respect to host genotype and fungicides [12]. Considering what mentioned we designed the proposed system based on weather or environmental parameters forecasting/sensing for two folds; first, is because Late blight pathogen itself is highly dependent on the environmental factors like temperature, and Relative Humidity for causing the disease, and second, is that susceptible host controllability is very challenging due changing of host genotype that makes fungicides ineffective.

Disease warning system provides the farmer or the grower relevant notification from the farm to deal and see what is happened in the farm from changes in weather and climate. One of the features that disease warning system gives is the time the farmer needs to use pesticides. Another feature of the system that will use weather forecasting Application Programming Interface (API), to predict if there will be a disease or not. According to [26] “Temperature and moisture are the most important environmental factors affecting late blight of potato development,” and according to [28] the most commonly used weather parameter as input for disease warning systems are relative humidity, temperature, rainfall, and Leaf Witness, and leaf wetness duration (LWD) [29]. Some additional variables used in only a few warning systems are solar radiation, wind speed, and wind direction. The parameters used in our proposed system are temperature, relative humidity, wind speed, wind direction, and solar radiation using Machine learning. The system does not use rainfall as an input because it is rare for the atmosphere of Egypt; the rain falls 3 or 4 times in the year or less in some parts in Egypt in some years [30].

The proposed system consists of two main components or sub systems as shown in Fig. 3, the first is a **proactive** subsystem, which uses Supervised Machine Learning (SML) to predict and avoid the favourite environment for disease using three inputs: 1) data gathered from the potato field using sensors; 2) the previous dataset of environment parameter; and 3) the weather forecasting data from weather station. Then if there is any action that needs to be done by an actuator or as an alert and or push notification for the farmer [33]. The second is **reactive** subsystem and is mainly based on sensing data and image processing to follow the crop if any of the early symptoms are shown in the leaf of the potato. This paper will focus on the design and implementation of the proactive subsystem of the overall Smart warning system. The SML takes these meteorological data extracted from the sensors or weather stations and process them in a way to have the best predicted and proactive analytics and send if applicable early alert/notifications or take proactive actions via some actuators on the failed of the crop to maintain the best productive environment for the crop [32].

Fig. 3 - Warning System Block Diagram



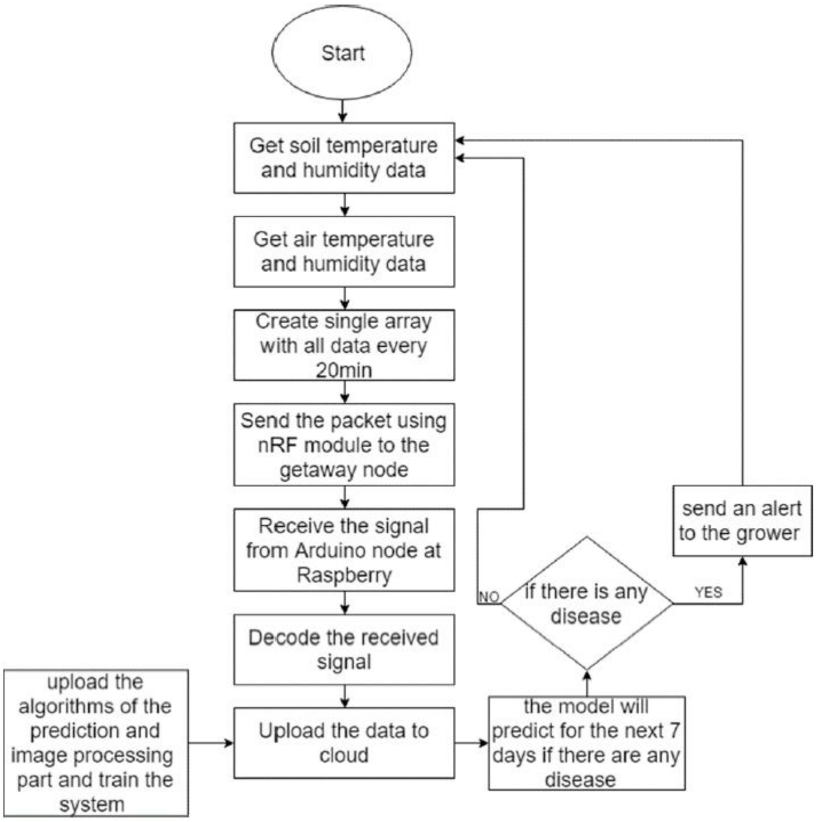
As shown in Fig. 4 the system will start by collecting the data from the sensors on the plant environment from the nearest weather station.

Fig. 4 - Sensors Dataset Structure

	temperature	humidity	pressure	wind_direction	wind_speed
count	44461.000000	44461.000000	44461.000000	44461.000000	44461.000000
mean	23.347276	52.942376	1002.039248	204.221722	3.452576
std	8.852984	27.456244	22.610353	144.899479	2.323312
min	-2.000000	0.000000	936.000000	0.000000	0.000000
25%	16.584000	31.000000	1002.000000	40.000000	1.000000
50%	23.000000	48.000000	1010.000000	264.000000	3.000000
75%	30.000000	75.000000	1016.000000	345.000000	5.000000
max	47.000000	100.000000	1040.000000	360.000000	44.000000

Then the system will check if there is any high risk or out of the safe range value. If there are no alerts about the disease and the environment is not favorable for the disease, the system will gather weather forecasting data from weather forecasting APIs (like Yahoo Weather, etc.). This data will be processed by prediction algorithm as shown in Fig. 5 which uses historical data for the disease appearance. Finally, the warning system will send to the farmer suggestion about proactive action to early tackle the disease before it even shows up on the crops [31].

Fig. 5 - Warning System Flow Chart



The dataset used for our supervised machine learning is collected from Cairo Over the course of 5 years consists of the main five parameters (relative humidity, temperature, wind speed, wind direction, and pressure). The Mean, Standard variation, maximum, minimum and percentage between min and max values of each parameter is shown in Fig 4. The formation of the disease has happened with specific conditions when relative humidity is greater than 90%. Sporulation can occur from 3-26°C (37-79°F), but the optimum range is 18-22°C (64-72°F). The spreading of the disease can occur in values of climate conditions. This paper proposed the usage of machine learning in warning systems for plant diseases. The warning system is a technology that predicts the possibility of getting a certain disease in a certain crop from some parameters of the atmosphere in a specific area [35].

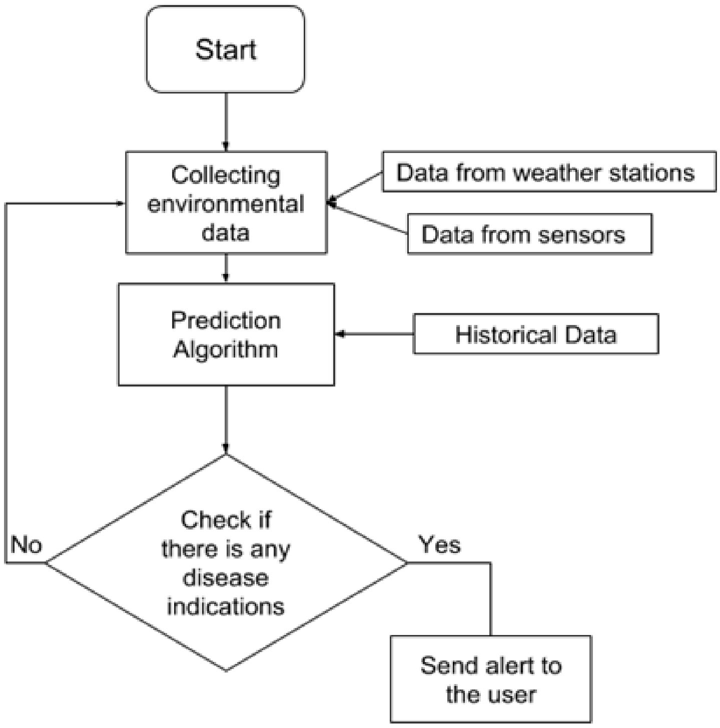
In order to build a system that makes a forecast for the late blight of potato to avoid the losses in the field of potato, we choose "logistic regression algorithm" that takes as its inputs a large dataset of weather forecasting datasets in Cairo of air temperature and relative humidity for more than 5 years. This large dataset of weather forecasting will be used for the training of the system and it will have other inputs like forecasting API's from google whether or yahoo weather, etc [36]. The logistic regression algorithm will process these inputs based on the conditions that the blight will appear and the best condition for the disease to spread in the field. The output of the algorithm will be seven days of forecasting and at least 48 hours of prediction of the symptoms of the blight of potato [39].

Logistic Regression (LR) [40] is a part of supervised machine learning. LR is a technique used for classification. When it comes to classification, the data needed to be expressed the probability should be a value between 0 and 1. To generate a value between 0 and 1 it can be expressed by using the Sigmoid function equation shown below, its boundary from 0 to 1, and it cuts across the y-axis at 0.5.

$$h_\theta(\boldsymbol{x}) = f(\theta^{\mathrm{T}} x) = \frac{1}{1 + e^{-\theta^{\mathrm{T}x}}} \quad \text{for } \vartheta\, z \in \Re$$

Binary neural network (BNN) [38] is a supervised machine learning. BNN is a technique used for binary classification. The BNN used consists of five inputs with two hidden layers and one output layer. The two hidden layers, each with four nodes. ReLu is the activation function used. For the output, the layer is built with one node, and the activation function used is sigmoid for the classification [35].

Fig. 6 - Proposed System Flowchart

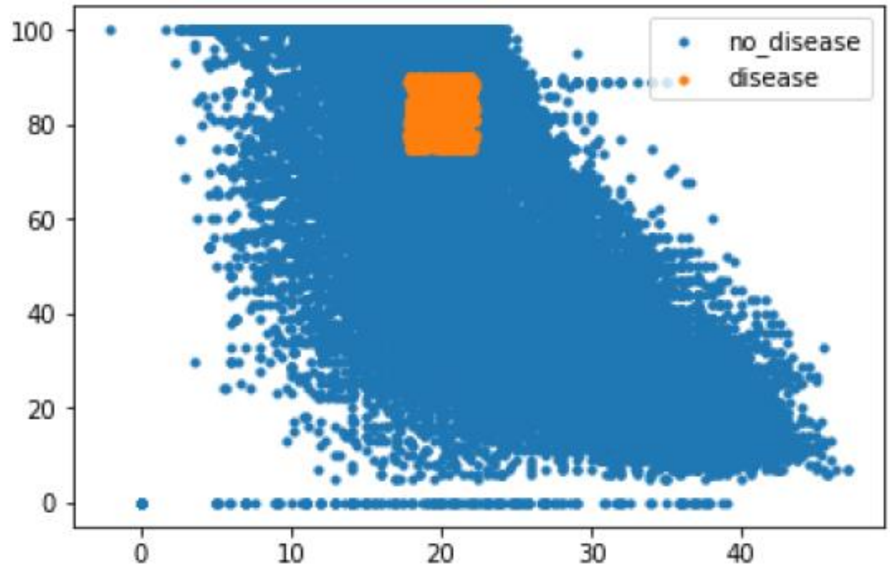


As shown in Fig. 6, the system will start by collecting the data from the sensors on the plant environment (relative humidity, air temperature, and soil moisture) while the other weather parameters are collected from the nearest weather station. Then the system will check if there is any high risk of disease favorable environment from historical data. If there are no alerts about the disease and the environment is not favorable for the disease, the system will gather weather forecasting data from weather forecasting APIs (like Yahoo Weather, etc.) [37]. This disease favorable environment data collected from historical systems will be used to label the meteorological weather parameters as 'disease' or 'no disease' for the binary classification for the BNN training and prediction algorithms. By using historical data about the disease appearance, the output data from this algorithm will be a prediction if there will be disease or not. Finally, it will send to the farmer suggestion about proactive action against the disease.

The dataset used from Cairo of air temperature and relative humidity for more than 5 years contains 5 classes of 44462 row, where each class (column) refers to relative humidity, temperature, wind speed, and soil moisture ending up with a binary classification problem. The data is split into three parts 70% for the training using labeled data and 20% for the testing, and 10% for model fine tuning and performance validation, for the training purposes, the 70% of labeled data is ended up with a binary classification and for the other unlabeled 20% of data, the binary classifier is removed to validate the model and see if the model can predict as expected. The remaining 10% of the data is used to fine tune the model parameters and to prevent any overly optimistic estimate of model performance on 'unseen data' when the model is deployed in the real field. The program used for running the algorithm is Jupiter and after the training the system that is running on is raspberry pi 3B+. Thus, the first model of BNN with 2 hidden layers was with accuracy of 98.94% and with adding the 3 hidden the first hidden layer is with 10 nodes, the second hidden layer is with 5 nodes, and the third layer is with 10 nodes with 4 input layer and one output layer. The accuracy of the model becomes 99.4% as shown in Fig. 7 [34]. The final design of the model was decided based on two phases, the first phase aims to get the optimum values for the model parameters. To achieve this objective, the model was running using different batch size with values [32, 64, 128..., 1024], different learning rates ranges between [0.001 and 0.1], different number of hidden layers ranges between 2 and 4 layers. The initial model of the proposed BNN was designed with 2 hidden layers only lead to accuracy of 98.94% and the one with the 3 hidden layers lead to accuracy of 99.4% where the first hidden layer is with 10 nodes, the second hidden layer is with 5 nodes, and the third layer is with 10 nodes and input layer with 4 input inputs and one output layer. The 4 hidden layers design consumes too much time for learning with no achievements in the overall system accuracy.

Thus, based on these trials, we selected the optimum model with 1024 as batch size, learning rate to 0.001, epoch size of 1000 and 3 hidden layers as detailed above. The objective of this fine tuning of model parameters was to avoid the overfitting and control the epoch size based on the error rates.

Fig. 7 - Training Dataset for Disease Classification



The proposed system includes the design of the weather station to collect the weather parameters from the crop field like temperature, relative humidity, wind speed, others along with the soil nodes to collect the soil parameters like soil temperature, others. The range of weather nodes communication is very wide, so we need one sensor for each parameter to cover very wide area, weather sensors can cover more than 40 acre but Soil sensors are limited range, so we need more than one soil sensors to achieve high accuracy reading for soil status, we need between 5 and 10 sensors for one acre to achieve high accuracy reading, Thus, the soil node will consist of small microcontroller to run one soil sensor and one nRF module to send data to the gateway [45].

3. Model Validation and Testing

The data set mentioned in section III is split into 70% for training, 20% for testing and 10% for model parameter tuning. We used K-folds cross validator to ensure cross validation for the model and used data set balance. The data was shuffled with a random state equal to 42. The data split into 5 parts/folds and we used Adam as optimizer with learning rate-0.001, beta 1=0.9, beta2=0.999 and with epsilon=1e-08**.**

After running the algorithms of the logistic regression and Binary neural networks (BNN) and comparing the proposed system results with the algorithms built with Microsoft Azure [41], the results are shown in Table 1. The reading that gives this prediction is shown in Fig.8 consists of (temperature, relative humidity, pressure, wind direction, and wind speed) respectively [44].

Fig. 8 - Reads used for the Prediction

```
([[22.5,81,1010,30,8],
  [21,85,1019,60,9],
  [20,78,1015,320,5],
  [20,77,1019,320,5],
  [20,88,1019,290,4],
  [30,40,1015,360,5],
  [20,77,1019,320,5],
  [40,50,1011,230,9]])
```

Table 1 - Comparing between the Two Algorithms

The output of Microsoft Azure		The output of the NN built algorithm	
Prediction %	Disease [0 or 1]	Prediction %	Disease [0 or 1]
32.47	0	30.28	0
96.91	1	93.35	1
98.21	1	50.81	1
96.87	1	99.58	1
93.33	1	94.79	1
1.56E-13	0	0.	0
96.87	1	99.58	1
1.36E-12	0	0.	0

The result was validated by using the “Confusion matrix”, the Total of true positive and true negative is 13267 out 13339 observations in the test dataset. so, the accuracy for test dataset is around 99.4% and the recall is 90.5% as shown in “fig.9”

Fig. 9 - Confusion Matrix Validation

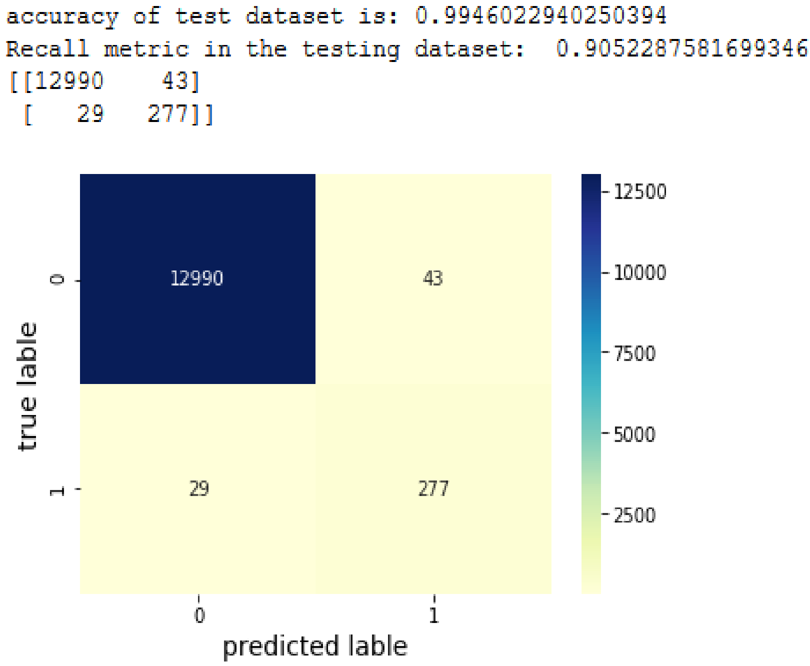


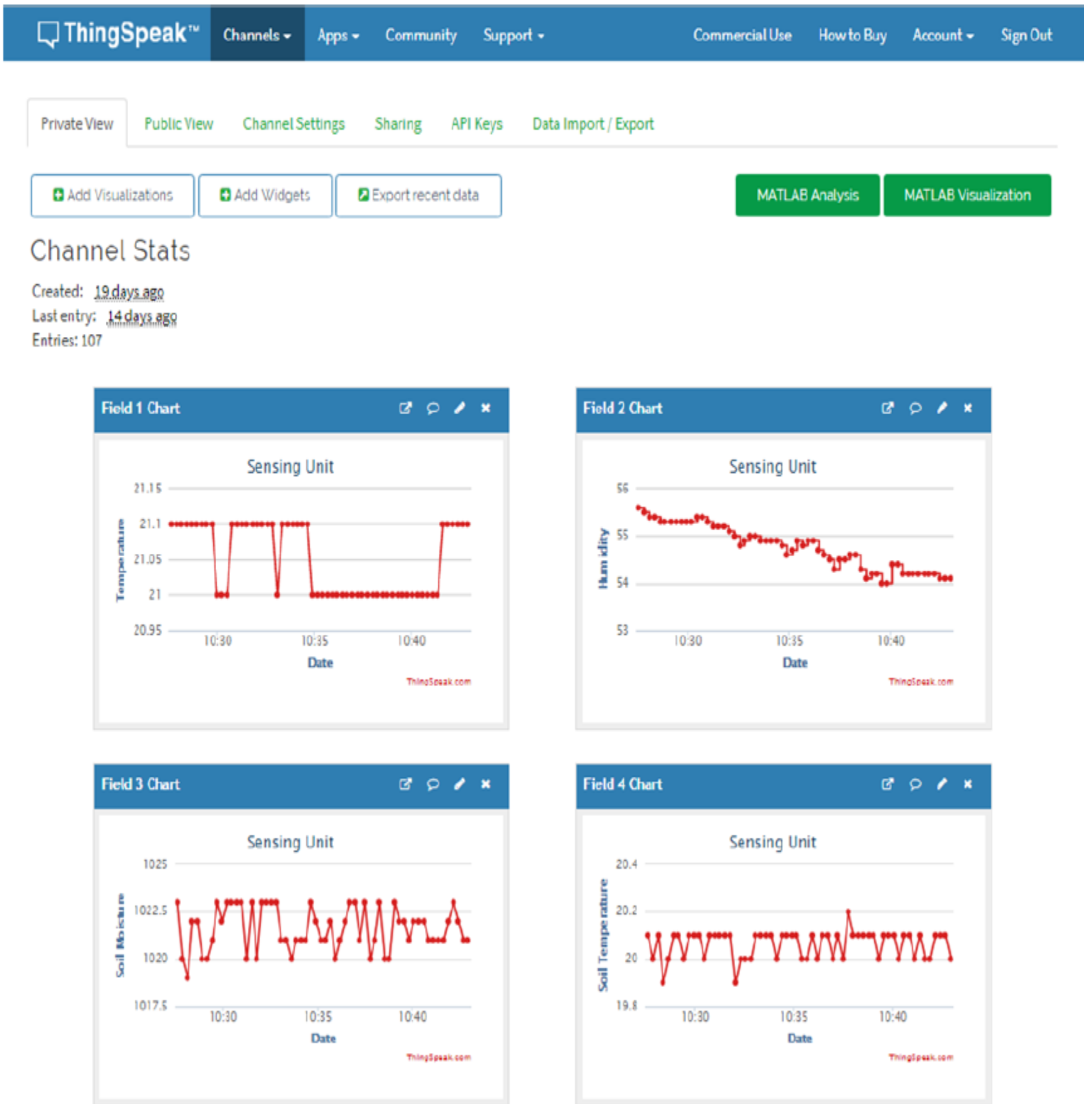
Table 2 illustrates the accuracy of the different models after the training for the same algorithm but with different implementations between Neural networks and Logistic regression for the two models/algorithms used from Azure [42] and our proposed algorithm.

Table 2 - The Accuracy of the Models

The accuracy of the model (%)	Neural networks	Logistic regression
Algorithm used from Microsoft Azure	99.4%	77%
The proposed algorithm	99.4%	76.92%

After building the first prototype, the sensing module can gather different types of data (soil parameters, air temperature and relative humidity) by sensors installed on the field connected with Arduino Uno then send this data as array using nRF24L01 module to the master node (Raspberry Bi) [43] with nRF module, Bi will decode the received data and send it to the cloud using Wi-Fi module connected with the Internet every 15 sec as shown in Fig. 10. The Fig shows the dashboard from the collected soil and meteorological data as generated by the cloud. This will help the farmer for easily handling data analytics for the diseases prediction and take the proactive actions to minimize any crop production losses [34].

Fig. 10 - Weather Dashboard for Diseases Prediction System



4. Conclusion

Precision agriculture is one of the approaches that predict and control it without wasting in using pesticides. Warning system or weather forecasting system is one of the systems that support Precision agriculture. Various parameters like relative humidity, temperature, wind direction, and wind speed are used for the prediction of a potato late blight. Analysis for the different meteorological parameters that lead to late blight of potato proved to be effective strategy for disease management. The proposed model is designed with different algorithms to predict such disease like logistic regression and neural networks. The proposed warning systems compares the two algorithms using same data set to show the effective and fit for purpose one. To show the value of the proposed against existing benchmark, we compared the proposed system with Azure predication system. The outcomes show high potential as shown in the paper with accuracy of 99.4%. The implementation and the realization of a warning system for precision agriculture for most of disease and pests in most of the plants will be the task of the future.

Acknowledgement

This paper and the research behind it would not have been possible without the engagement and support of MSA student Ahmed Nassar whom I would like to thank and appreciate all his efforts, insights and expertise that greatly assisted through the course of this research. Finally, it is with true pleasure that I pay my sincere gratitude and acknowledgement for Ahmed' contributions to achieve the research objectives as planned.

References

Hegazy, Eng Salah. *"Seed potato production in Egypt."* Agro-Food Co. Ltd., Egypt 2009.

Mark L. Gleason et al., "Obtaining Weather data for input to crop diseases-warning systems: Leaf wetness duration as a case study," *Science Agriculture*, V.65 special issue, 76-87, 2008.

“Economic study of Production and Consumption of Potato in Egypt,” *Assiut Journal of Agricultural Sciences*, 46(1), 58–67, 2015.

K. A.A., Y. Essa, F. Hashem, and K. Refaie, “Plant diseases for major crops in Egypt under future climate conditions”, *International Journal of Current Agriculture Science*, 6(12), 149–154. 2016.

J.G. Ramírez-Gil, G.O.G. Martínez, and J.G.M. Osorio, “Design of electronic devices for monitoring climatic variables and development of an early warning system for the avocado wilt complex disease,” *Computers and Electronics in Agriculture*, 153, 134–143, 2018.

V.K. Mishra, S. Kumar, and N. Shukla, “Image Acquisition and Techniques to Perform Image Acquisition,” *Samriddhi: A Journal of Physical Sciences, Engineering and Technology*, 9(01), 2017.

K. Thangadurai and K. Padmavathi, “Computer Visionimage Enhancement for Plant Leaves Disease Detection,” 2014 World Congress on Computing and Communication Technologies, 2014.

Gleason, M.L.; Macnab, A.A.; Pitlado, R.E.; Ricker, M.D.; East, D.A.; Latin, R.X. Disease-warning systems for processing tomatoes in eastern North America: Are we there yet? *Plant Disease*, 79, 113-121, 1995.

Gillespie, T.J.; Sristava, B.; Pitblado, R.E. Using operational weather data to schedule fungicide sprays on tomatoes in southern Ontario, Canada. *Journal of Applied Meteorology*, 32, 567-573, 1993.

“Economic study of Production and Consumption of Potato in Egypt,” *Assiut Journal of Agricultural Sciences*, 46(1), 58–67, 2015.

Vib.be. (2019). *VIB*. http://www.vib.be/en/Pages/default.aspx

S. Dahikar and S. Rode, Agricultural Crop Yield Prediction Using Artificial Neural Network Approach, 2nd ed. 2014.

Vikas Lamba, V.S. Dhaka, Wheat Yield Prediction Using Artificial Neural Network and Crop Prediction Techniques (A Survey), IJRASET, Vol.2Issue IX, ISSN: 2321-9653,2014

Donna Henderson, Christopher J. Williams, Jeffrey S. Miller, Forecasting Late Blight in Potato Crops of Southern Idaho Using Logistic Regression Analysis, 2007.

D. Moore, A. Trinci and G. Robson, *21st century guidebook to fungi.* Cambridge: Cambridge University Press, 2013.

"Late blight of potato and tomato", *Late blight of potato and tomato*, 2019. https://www.apsnet.org/edcenter/disandpath/oomycete/pdlessons/Pages/LateBlight.aspx

M. Gleason et al., "Obtaining weather data for input to crop disease-warning systems: leaf wetness duration as a case study", *Scientia Agricola, 65,* 76-87, 2008. 10.1590/s0103-90162008000700013

D. Corrales, "Toward detecting crop diseases and pest by supervised learning", *Engineering and University,* 19(1), 207, 2015. 10.11144/javeriana.iyu19-1.tdcd

A. Pandey and A. Mishra, "Application of artificial neural networks in yield prediction of potato crop". *Russian Agricultural Sciences, 43*(3), 266-272, 2017. 10.3103/s1068367417030028

Anandhi A, Srinivas VV, and Nanjundiah RS, Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine, *International Journal of Climatology, 28,* 401-420, 2008

Andrade-Sanchez, P, Gore MA, Heun JT, Thorp KR, Carmo-Silva AE, Andrew BD, French AN, Salvucci ME, and White JW. Development and evaluation of a field-based high throughput phenotyping platform, *Functional Plant Biology, 41,* 68–79, 2014.

Arthur WB (2011). The second economy, Mckinsey Quarterly, October 2011, 1-9. http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-second economy

Assunção MD, Calheiros RN, Bianchi S, Netto MAS and Buyya R Big Data computing and clouds: Trends and future directions, *Journal of Parallel and Distributed Computing,* 79-80, 3-15, 2015.

Antle JM, Jones JW, and Rosenzweig CE, Next generation agricultural system data, models and knowledge products: Introduction, *Agricultural Systems, 155:* 179–185, 2017.

Azmak O, Bayer H, Caplin A, Chun M, Glimcher P, Koonin S and Patrinos A, Using Big Data to Understand the Human Condition: *The Kavli Human Project, Big data, 3,* 173-188, 2015.

Benestad R, *Downscaling climate information, Oxford Research Encyclopedia, Climate Science* (climatescience.oxfordre.com), Oxford University Press USA, 2016, 37. DOI:10.1093/acrefore/9780190228620.013.27, 2016.

Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P and Jannic J, The triticeae toolbox: Combining phenotype and genotype data to advance small-grains breeding, *Plant Genome, 9*(2), 10. 2016. https://dl.sciencesocieties.org/publications/tpg/pdfs/9/2/lantgenome2014.12.0099

Bomgardner MM, Transforming agriculture, again, *Chemical and Engineering News, 94*(34). 32-38, 2016.

Butler D, When Google got flu wrong, *Nature, 494,* 155-56, 2013.

Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E, Loboguerrero AM, Ramirez-Villegas J, Rosenstock T, Sebastian L, Thornton P and Wollenberg E, Reducing risks to food security from climate change, *Global Food Security, 11,* 34-43, 2016.

Ekström M., Grose, M.R., and Whetton P.H, An appraisal of downscaling methods used in climate change research, *WIREs Clim Change,* doi: 10.1002/wcc.339, 2015.

Faghmous JH and Kumar V, Climate change: the case for theory guided data science, *Big Data,* 2(3), 155-163, 2015.

Fahlgren N, Malia AG and Baxter I, Lights, camera, action: high-throughput plant phenotyping is ready for a close-up, *Current Opinion in Plant Biology, 24*, 93-99, 2015.

Fan J, Han F and Liu H, Challenges of big data analysis, *National Science Review (China),* 1: 293-314, 2014.

FAO, Climate-Smart Agriculture: Policies, *Practices and Financing for Food Security, Adaptation and Mitigation,* FAO, 49, 2010.

Fischer EM and Knutti R, Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change, 5,* 560–564, 2015.

Ford JD, Tilleard SE, Berrang-Ford L, Araosa M, Biesbroekb R, Lesnikowskia AC, MacDonald GK, Hsu A, Chen C, and Bizikov L, Big data has big potential for applications to climate change adaptation, *Proceedings, National Academy of Sciences, 113* 10729–10732, 2016.

Gandomi A, Haider M, Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management 35*, 137–144, 2015.

Gilpin, L., How big data is going to help feed 9 billion people by 2050, 2014.
(http://www.techrepublic.com/resource-library/downloads/how-big-data-is-changing-farming-pdf-download/post/?skipAutoLoad=1)

Ginsberg J, Mohebbi MH, Rajan, Pate S, Brammer L, Smolinski MS, and Brilliant L, Detecting influenza epidemics using search engine query data, *Nature 457,* 1012-1014, 2009.

Girvetz, E.H., E.P. Maurer, P. Duffy, A. Ruesch, B. Thrasher, C. Zganjar, 2013, *Making Climate Data Relevant to Decision Making: The important details of Spatial and Temporal Downscaling,* The World Bank, March 27, 2013.
(https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1012&context=ceng)

Glendenning CJ and Ficarelli PP, *The relevance of content in ICT initiatives in Indian agriculture, International Food Policy Research Institute (IFPRI),* Discussion paper 01180, IFPRI, Washington DC, USA, 40 pp., 2012.

Global Harvest Initiative, The 2014 Global Agricultural Productivity Report, 65. http://www.globalharvestinitiative.org/gap-report-gap-index/2014-gap-report/, 2014.

Goly A, Teegavarapu RSV, and Mondal A, Development and evaluation of statistical downscaling models for monthly precipitation, *Earth Interactions, 18,* 1-20, 2014.