Machine Learning Approach for Agricultural IoT

Abraham Sudharson Ponraj, Vigneswaran T

Abstract: ***The rapid growth of Internet of Things (IoT) devices in cities, homes, buildings, industries, health care, automotive and also in agricultural farms have paved the way for deployment of wide range of sensors in them. In return IoT turns out to be the major contributor of new data in any of these fields. A data driven farm management techniques will in turn help in increasing the agricultural yield by planning the input cost, reducing loss and efficient use of resources. IoT on top of increasing the volume of data it also give rise to big data with varied characteristics based on time and locality. To increase the agricultural yield by smart farm management astute analysis and processing of the data generated becomes imperative. With high performance computing at machine learning has created new opportunities for data intensive science. Machine learning will help the farm management system to achieve its goal by exploiting the data that is continuously made available with the help of Agricultural IoT(AIoT) platform and helps the farmer with insights, decisive action and support. This article analyses various existing supervised and unsupervised machine learning techniques applied in agricultural domain and compares one technique with another with respects to accuracy and a confusion matrix is plotted for each***.

Index Terms: **Agricultural IoT, Big Data, Data Driven Farm Management, Internet of Things, Supervised Machine Learning; Unsupervised Machine Learning.**

I. INTRODUCTION

The One of the long-standing objectives of computing is to simplify and enrich human activities and experiences.The Internet of Things (IoT) is the talk of the town now and has made a massive inroad in the last decade in the field of modern wireless telecommunications. IoT is an integration of various embedded technologies such as the physical object, sensing and actuating, networking and computation connected to the internet. The strength of this technology will be its ability to have an impact in a simple decision making in everyday life to something as consequential as health monitoring and there by enriching human life and endeavors. It is predicted that by 2025 billion of devices ranging from a simple paper document to a complex industrial machine will be connected to an internet node [1]. This paves way to greater contribution to the economic growth by bridging the gap in various technologies given the advancement in technology. While making quite an impressive contribution in smart cities, industrial IoT, data driven Agricultural IoT(AIoT) is the next big thing. With the ever increasing population, upward mobility and depleting environmental condition has made feeding in 2050 a challenge [2]. By 2050 farm yield can be increased by at least by 67% while reducing the agricultural losses with data driven agricultural techniques as reported by International Food Policy Research Institute. A sensor based precision agriculture for sub-surface drip irrigation system indicated water saving by 35.7% while an increase in farm yield by at least 45% [3]. Furthermore data driven techniques can be equivalently applied to various other farm inputs like seeds, climate, soil nutrients, etc. to bridge the yield gap [4]. Precision agriculture has gain popularity with the rapid development of IoT techniques and it has made remarkable contribution in monitoring plant health, yield predication, irrigation planning, fertilizer usage, etc. there by contributing to increase in yield and reducing loss. Notably AIoT doesn't limit its support in the agriculture filed whereas it can make immense contribution to the agro-industrial production chain [5]. In the field it helps in monitoring and controlling field variables like soil condition, environmental conditions, and biomass of plants or animals. It further contributes during the produce transport by analyzing and managing temperature, humidity, disturbance and pest [6]. The shelf life and demand of the produce can be monitored and predicted based on the origin and properties of the produce, the end users or consumers will be greatly benefited by these information. The IoT based agro-farm and agro-industry as whole can bring in a smart rural community. To make the AIoT more intelligent system it's important to apply data science to the data generated from the various parts of the Agro-farm and Agro-industry. Data science is the field which is a combination of data mining, artificial intelligence and other techniques to analyze the data and to come out with new inference and predictions [7]. While using various data analytic techniques it is important to understand the various data characteristics and apply a suitable algorithm to make an efficient AIoT data analysis. In this article the need for data driven farm management system with the help of AIoT is underlined and the architecture of AIoT is described in detail. Taxonomy of machine learning is neatly illustrated and how the different types of machine learning algorithms are made used in an agrarian framework.

Revised Manuscript Received on 30 March 2019.

* Correspondence Author

Abraham Sudharson Ponraj*, Assistant Professor at Vellore Institute of Technology, Chennai, India.

Dr.Vigneswaran. T Professor, Vellore Institute of Technology, Chennai, India.

Finally with the help of wheat kernel dataset the machine learning models were analysed in terms of accuracy and a confusion matrix was plotted for them using python. The rest of the paper is organised as follows: section 2 deals about agricultural IoT. Machine learning for AIoT is described in section 3 and section 4 gives the review of machine learning models in agricultural framework. Section 5 presents the analysis of machine learning algorithms and the conclusion is given in section 6.

II. AGRICULTURE IoT

A basic structure of AIoT is shown in Fig 1. AIoT can be broadly classified into four layers Data Collection and Transfer Layer, Network Layer, Service Layer and Application Layer [8]. The data collection layer houses the sensor nodes or WSN, each sensor node comprises of an embedded controller, various sensors ranging from a simple temperature sensor to camera, actuators like motors, sprinklers, etc. and any wireless communication interface which may can be WiFi, LoRaWAN, Zigbee, etc. The various types of data collected should be made be available in the internet to make this possible, the local WSN gateways transmits them through an internet gateway which may be either a mobile network or an Ethernet based connection and this constitutes the network layer. To make sense of the collected data in the service layer it is necessary to do some data processing like data visualization, data analytics, data storage and protection, etc. Finally the application layer is where it all maters here is the end user can monitor and control the various process in the agro-farm and also make important decisions based on the predictions, market trends and local agricultural departments [9]. AIoT leads to an increase in wide variety of data generation from different sources in and around an agricultural farm in the form of voltage values to images, status of actuator to a position of robot, etc. [10]. The data generated may be of continuous value which will contribute to a greater increase in data volume, continuous data from GPS to hourly soil moisture and temperature value update. The data collected will be dynamic in nature for example a fertilizer spraying quad copter location varies instantaneously. The factors that affect the data quality are the redundancy of data, data accuracy, dynamic and crude nature of data. It is imperative that at all times the quality of data is maintained though the data generated are from heterogeneous sources. The factors that most likely influence the quality of the data are the noise from the environment, measurement errors, heterogeneous data scalability, data stream processing and sensor node failures [11]. It is understandable that a quality data leads to quality information, without quality data it's not possible to apply data analytics or any machine learning algorithm to create a prediction model. Many techniques can be used to improve the data quality from semantic data annotation to data quality techniques like outlier detection, interpolation, data integration, data duplication and data cleaning [12]. Data quality improvement should be accomplished without losing the data integrity and or its original attributes. Now applying machine algorithms to these improved data sets will result in better analysis and accurate prediction.

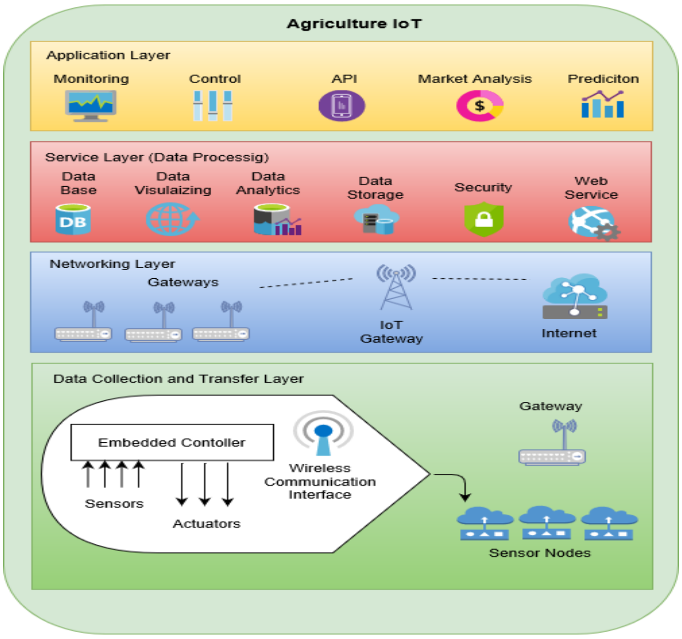

Fig 1. Architecture of AIoT

III. MACHINE LEARNING FOR AIoT

Machine learning is a part of Artificial Intelligence and can be classified under computer science. It has the ability to render machine to learn without definite computer programming and thereby enhancing machine performance with detection and characterise the consistencies and patterns in trained data. Machine learning can be classified into three different categories based on their learning method [13, 14].

A. Supervised learning

When data is available they can always be trained. For example let 'x' be the input data and 'y' be the output data, the mapping function from the output to input is learned using an algorithm [15].

$$y = f(x) \tag{1}$$

The mapping function is approximated such that when a new set of input 'x' is given the output 'y' can be predicted. Supervised learning can be subsequently categorised into regression and classification problems.

- Classification: A classification problem is when the output variable is a category, such as "Mango" or "Apple" and as "disease" or "no disease".
- Regression: A regression problem is when the output variable is a real value, such as "dollars" or "weight".

B. Unsupervised learning

Here only input data 'x' is available and no corresponding output data available. The algorithm learns more about the data by modelling the underlying structure of the data. Since there is no definite output and no supervisor it's called unsupervised learning.

They can be subsequently categorised into clustering and association problems [16].

- Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
- Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

C. Reinforcement learning

For a given set of data ‘x’ the algorithm learning problem is to find the best possible actions based on the best possible action to be taken by which you get to maximize expected turnouts. Here the learning algorithm uses trial and error method to find the optimal output. A general feature of reinforcement learning is the trade-off between exploration, in which the system tries out new kinds of actions to see how effective they are, and exploitation, in which the system makes use of actions that are known to yield a high reward [17].

Table 1 Taxonomy of machine learning

	Learning Process	Data Processing Task	Reference
Support Vector Machine (SVM)	Supervised Learning	Classification	[[8 – 21]
K-Nearest Neighbour (KNN)	Supervised Learning	Classification	[22]
Naive Bayes (NB)	Supervised Learning	Classification	[23,24]
Support Vector Regression (SVR)	Supervised Learning	Regression	[25,26]
Logistic Regression (LR)	Supervised Learning	Regression	[27]
K-Mean	Unsupervised Learning	Clustering	[28,30]
Random Forest (RF)	Supervised Learning	Classification/Regression	[31,32]
Decision Tree (DT)	Supervised Learning	Classification/Regression	[33,34]
Artificial Neural Network (ANN)	Supervised Learning	Classification/Regression/Clustering/Feature Extraction	[35,36]
Convolution Neural Network (CNN)	Supervised Learning	Classification/Regression/Clustering/Feature Extraction	[37,38]

IV. MACHINE LEARNING MODELS IN AGRICULTURAL FRAMEWORK

Machine learning application in agro-farm can be widely found in areas like yield detection, disease detection, weed detection, irrigation planning, soil condition, quality of crop and weather prediction. After yield one can find machine learning used in analysing the produce freshness (fruit and vegetable freshness), shelf life, produce quality, market analysis, etc.

A. Crop and Yield Management

Yield monitoring and prediction in agriculture plays vital role by giving information to the user to make decisive action and thereby reducing loss. [39] Proposes a SVM based prediction of rice development process with the assistance from the Chinese weather station basic geographical information. [40] Describes an efficient automatic coffee fruit counting technique, this information help the farmer to plan the agriculture process and efficiently reduce loss.[41] Presents a crop yield prediction and pest control analysis system with neural network based on the environmental condition. An Artificial Neural Network (ANN) based generalized agriculture yield forecast was presented in [42]. [43] States a K-Means algorithm which analyses the measure of extreme precipitation due to spatiotemporal changes, and its role in crop yield, which in turn helps in improving the crop yield. Decision tree based assessment of the crop production under different types of edaphoclimatic environmental conditions is inferred from [44]. Study of grain loss and predicting the influence of several agricultural parameters in the extent of loss incur by applying decision tree algorithm are discussed in [45]. A convolution neural network (CNN) to recognize and categorize legume crop based on vein leaf image of red, white and soya bean [46].

B. Soil Management

A K-Nearest Neighbour based prediction of soil drying with the help of hydrologic data of precipitation and evaporation showed an accuracy of 94% [47]. [48] Recommends required quantity and type of fertilizers and suitable crop based on the soil chemical parameters. [49] Describes an irrigation scheduling method based predicted soil moisture data using Support Vector Regression (SVR). [50] Presents a classifier of soil datasets which is based on Naive Bayes and achieves an accuracy of 100%. Soil condition forecasting of soil organic compound (OC), soil moisture (M) and total nitrogen (TN) based on various regression models was illustrated in [51].

Soil moisture prediction with the help of ANN model on data from no-till chisel opener mounted force sensor. [52] Describes an ANN based prediction of the density and variety of bacteria and microbial present in the soil. [53] Presents everyday soil temperature estimation for various depths in different climatic areas with the help of a live weather data input and self-adaptive evolutionary-extreme learning machine (SaE-ELM) model.

Table 2 Machine Learning in Yield Prediction

Model/Algorithm	Functionality	Reference
SVM	Prediction of rice development process	[39]
SVM	Automatic coffee fruit counting technique to reduce loss	[40]
ANN	Agriculture yield forecast and pest control	[41]
ANN	Generalized crop yield prediction	[42]
K-Mean	Measure extreme precipitation to help in crop yield increase	[43]
Decision Tree	Analysing crop production under different environment conditions.	[44]
Decision Tree	Grain loss and yield loss influenced by agricultural parameters.	[45]
CNN	Recognize and categorizing legume crop	[46]

Table 3 Machine Learning in Soil Management

Model/Algorithm	Functionality	Reference
K-NN	Predicting soil drying from precipitation and evaporation hydrologic data	[47]
RF and SVM	Quantity and type of fertilizers for suitable based on soil	[48]
SVR	Soil moisture based irrigation scheduling	[49]
Naive Bayes	Soil based agricultural land classification	[50]
Least Square SVM	Soil condition forecast of OC, MC and TN	[51]
ANN	Soil moisture prediction from force sensor	[52]
ANN	Everyday soil temperature estimation for various depths.	[53]
ANN	Soil drying prediction	[54]

C. Disease Management

With the help of 79 and 75 sample data of okra and bitter gourd and by applying navie bayes classifier Yellow Vein Mosaic Virus (YVMV) disease was identified in them [55]. Bacterial Blight, Alternaria, Gray Mildew, Cereospra, and Fusarium wilt are cotton leaf diseases which were observed and classified with Support Vector Regression [56].Another SVR based forecasting of every day air temperature, relative air humidity and wind speed based on the data collected and thereby forecasting the dissemination and existence of fungal diseases via the focused crop field [57]. Using data from various leaf images with enough quality of information to distinguish healthy leaves form infected leaves based on CNN algorithm [58]. [59] Describes a K-Means clustering was used for identification of crop disease and the crop health status which in turn contributes in forecasting the loss in crop yield. An automated Bakanae disease diagnosis in rice seedlings was presented in [60], this system enabled in reducing loss and saves examination time when compared with human examination. [61] Presents an early diseases diagnosis and classification for various crop method based on SVM. With inputs from spectral reflectance features, an ANN model was developed to differentiate yellow rusted infected wheat from healthy wheat. [62]. A novel method using Genetic Algorithm (GA) together with Multi-Layered ANN(ML-ANN) was used to identify viruses in plant[63].

Table 4 Machine Learning in Disease Management

Model/Algorithm	Functionality	Reference
Naive Bayes	Yellow Vein Mosaic Virus (YVMV) disease was identified in bitter guard and okra	[55]
SVR	Cotton leaves diseases were observed and classified.	[56]
SVR	Forecasting the dissemination and existence of fungal diseases	[57]
CNN	Distinguishing healthy leaves from infected leaves	[58]
K-Mean	Identification of crop disease and the crop health status	[59]
SVM	Bakanae disease diagnosis in rice seedlings	[60]
SVM	Early diseases diagnosis and classification for various crop	[61]
ANN	Differentiate yellow rusted infected wheat from healthy wheat.	[62]
GA with ML-ANN	Identification of virus infected plant	[63]

D. Weed Detection

Ridolfia segetum is a common weed in sunflower crop field. Logistic regression was used on the data collected from multitemporal remote sensing to spot this weed infestation [64]. Another weed infestation identification based on random forest algorithm in sugarcane filed using the data from UAV images is described in [65] In another study [66] of weed detection which contributed to major yield loss is counter propagation (CP)-ANN based system which detects Silybum marianum with the image inputs from unmanned aircraft system.

Table 5 Machine Learning in Weed Detection

Model/Algorithm	Functionality	Reference
Logistic Regression	Ridolfia segetum identification from mutitemporal remote sensing data	[64]
Random Forest	Weed detection in sugarcane field from UAV images	[65]
CP-ANN	Identification of Silybum marianum from UAV images	[66]

E. Water Management

Water being a depleting resource in most part of the world it is important to store and use the available water resources optimally. [67], An ANN based weekly evapotranspiration forecast with at least the data of minimum and maximum air temperature from local meteorological department for better water management in arid region. Another water management model is based on the input of various weather data like the average air temperature, relative humidity, atmospheric pressure, etc. to forecast the daily dew point temperature with the help of ANN [68].

Table 6 Machine Learning in Water Management

Model/Algorithm	Functionality	Reference
ANN	Weekly evapotranspiration	[67]
ANN	Daily due point temperature	[68]

In this study 30 journal articles in which machine learning was applied to agricultural data were reviewed. Out of the 30 articles, 8 of them discussed yield prediction, another 8 of them were about soil management, 9 of them dealt about disease detection, 3 about weed detection and lastly 2 articles discussed water management. A total of 10 different machine learning algorithms were used in these 30 articles. About 5 machine learning models were used in yield prediction ANN and SVM were the most used, in soil management yet another 5 machine learning were made used here ANN is the most considered one. Disease detection sees about 6 different machine learning models being used, while the weed detection and water management makes use of 3 and 1 machine learning models respectively. From Fig.2, ANN seems to be the most widely used to model in the entire study with 33.3% , SVM follows by 20%, SVR come next with 10%, Naïve Bayes, K-Mean, CNN, Random Forest and Decision Tree follows with 6.6% and finally with 3.3% K-NN and Logistic Regression are the least used.

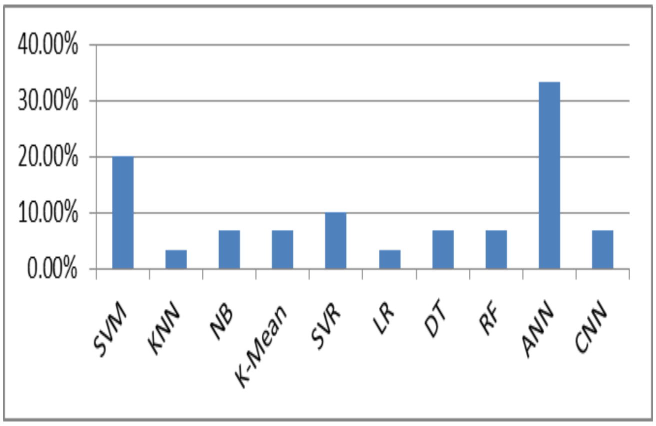

Fig. 2 Machine learning model used in agricultural framework

V. Methodology

To deploy machine learning effectively, it has to go through series of steps as in fig. 3. First and foremost is to identify what is expected from the ML model, for example should it classify or predict a new outcome. Collecting data is an important part, as without data it is not possible apply ML model. In order to enhance the quality of the gathered data various technique like data preparation, data cleansing, etc. are used. Then the enhanced data is passed on to the ML model to be trained. This process is called the learning phase, where the labeled or unlabeled are processed using ML model to acquire knowledge using the various attributes of the given data. To evaluate the ML model new unseen data are given as inputs and various mathematical and statistical techniques are used to validate the predicted output. If performance is still a concern, certain performance enhancement techniques can be considered and the process is repeated again.

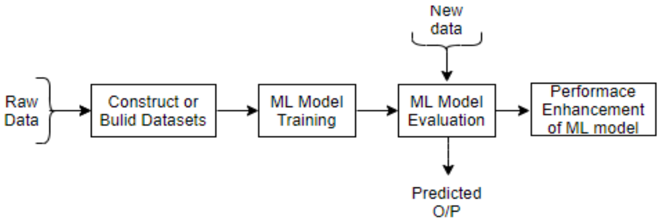

Fig. 3 Machine learning approach

VI. Result and Discussions

The survey uses dataset of three varieties of wheat kernels and they are Kama, Rosa and Canadian. 70 samples of each variety were arbitrarily selected for this survey work. A non-destructive method which uses soft X-ray for better visualization is used to identify the interior structure of kernel. Studies were conducted using combine harvested wheat grain originating from experimental fields, explored at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin carried out various studies on the harvested wheat grain from the observational field. The wheat data sets comprise of seven attributes which are continuous value. They are area A, perimeter P, compactness C = 4*π*A/P2, length of kernel, asymmetry coefficient and length of kernel groove.

The wheat datasets were collected from the UCI machine learning repository [69]. In the confusion matrix the wheat kernels are named like, 1 is Kama, 2 is Rosa and 3 is Canadian.

A. Support Vector Machine

SVM is another dynamic supervised learning algorithm for precise solving of classification problems. It aims at creating a model to predict the output data value with the given input data sets [18-20]. A hyperplane splits the input sets with a wide margin on either sides of the binary class to improve the accuracy. Now the predicted output data value of the new unseen input datasets is determined by finding on which of the hyperplane it falls. If the data are non-linear in nature which is likely in most cases, kernel is used to map them to high dimensional space [21]. SVM algorithm was applied to the wheat dataset and it resulted with an accuracy of 88.57%. Fig.4 shows the confusion matrix of the SVM model.

Fig. 4 Confusion matrix of KNN

	KNN			
True Label	1	0.56	0.06	0.39
	2	0.00	1.0	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

B. K-Nearest Neighbour

In KNN the classification is based on the data point closest to the training set in the input data set. Therefore to come out with K nearest neighbours of a new sample point, a Euclidean distance metric is used. Let X be the input dataset and Y be the training set, then the distance between them is given by the following equation (2),

$$d(X,Y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (2)$$

When the KNN algorithm was applied on the wheat data set and resulted in an accuracy of 87.14%, fig. 5 shows the confusion matrix. The limitation of KNN algorithm is that for large datasets computation time will be high as it needs to sort through the entire training sets [22]. Though KNN algorithm can be used for regression task it is predominantly used in classification application.

Fig. 5 Confusion matrix of SVM

	SVM			
True Label	1	0.61	0.06	0.33
	2	0.00	1.0	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

C. Naive Bayes

Naive Bayes is a probabilistic classifier algorithm based on the Bayes' theorem. It makes the naive assumption of independence among predictors, such that a presence or absence of an attribute in a class is unrelated to another attribute [23]. Let z be the new unseen input data z = (z_1 , z_2 , …, z_n) by applying Bayes' theorem,

$$P(c|z) = P(z|c)P(c)/P(z) \qquad (3)$$

$$P(c|Z) = P(z_1|c) * P(z_2|c) * \ldots\ldots * P(z_n|c) * P(c) \qquad (4)$$

Where,

P(c|z) is the posterior probability of class c
P(c) is the class prior probability.
P(z|c) is the possibility which is the predictor probability.
P(z) is the predictor prior probability.

Naïve Bayes classifier can be effectively used to train small data sets and a classifier which can manage high dimensional data samples [24]. When the Naïve Bayes algorithm was applied on the wheat data set and resulted in an accuracy of 86.48% and the confusion matrix is shown in fig. 6.

Fig. 6 Confusion matrix of Naive Bayes

	Naïve Bayes			
True Label	1	0.87	0.04	0.09
	2	0.17	0.83	0.00
	3	0.11	0.00	0.89
		1	2	3
	Predicted Label			

D. K-Mean

This algorithm is used to classify a given dataset into subsets in such a way that each element of a given data set is mapped to a cluster defined by a randomly selected seed element out of the given dataset [25]. The objective function which the algorithm minimizes is given in equation (5), where $||x_i^{(j)} - c_j||^2$ is the Euclidean distance between the data point $x_i^{(j)}$ and the centroid c_j, k is the number of clusters and n is the number of data points [26,27]. After applying the K-Mean model to the wheat kernel dataset an accuracy of 91.42% was obtained, the confusion matrix can be seen in fig. 7.

$$S = \sum_{j=1}^{k}\sum_{i=1}^{n}|| x_i^{(j)} - c_j||^2 \qquad (5)$$

Fig. 7 Confusion matrix of K-Mean

	K-Mean			
True Label	1	0.72	0.00	0.28
	2	0.05	0.95	0.00
	3	0.00	0.00	1.0
		1	2	3
	Predicted Label			

E. Support Vector Regression

Support Vector Regression aims at reducing the generalized error, which includes the training error and a regularization term [28]. SVR is derived from the SVM which is predominantly used for classification. To get a predicted output, SVR uses a threshold value 'ε' as the maximum tolerated deviation from the true value [29]. This parameter 'ε' is used to calculate the loss function. Instead of using the entire training data, SVR applies the parameter to a subset derived from the training datasets this helps in achieving the goal of reducing the generalized errors. When using the SVR model in the wheat kernel datasets an accuracy of 95.31%, the confusion matrix is shown in fig. 8.

Fig. 8 Confusion matrix of SVR

True Label	SVR			
	1	0.91	0.00	0.09
	2	0.05	0.95	0.00
	3	0.00	0.00	1.0
		1	2	3
	Predicted Label			

F. Logistic Regression

In logistic regression analysis of data is carried out using logistic functions in such a way that there are only two possible outcomes. In logistic regression, a best fit model is found for the probability that the given data set will give binary output and the input independent variables [30]. Once the model is defined the values for the coefficients will be available in the below given equation (6), where p is the probability of getting binary outputs. After applying the logistic regression model to the wheat kernel datasets an accuracy of 92.85% was obtained and fig. 9 shows the confusion matrix obtained.

$$\text{logit}(p) = b_0 + b_1X_1 + \ldots + b_kX_k \qquad (6)$$

Fig. 9 Confusion Matrix of Logistic Regression

True Label	Logistic Regression			
	1	0.78	0.00	0.22
	2	0.00	1.0	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

G. Decision Tree

For a given problem a decision tree is generated which goes from a rote node to the leaf node. The leaf node gives the final predicted outcome, whereas there are a lot of test case generated between the root node and the leaf node. Entropy and Gini Index (6&7) are the two most widely used methods to prune away the impurity in the datasets [31].

$$\text{Entropy} = -\sum_{i=1}^{m} p_i \log p_i \qquad (6)$$

$$\text{Gini Index} = 1 - \sum_{i=1}^{m} p_i^2 \qquad (7)$$

They are also more flexible in the type of data they handle [32]. One major disadvantage is that a small change in input contributes to reconstructing the entire decision tree model. 91.89% was acquired using decision tree algorithm on the wheat kernel datasets and the fig. 10 shows the confusion matrix of the decision tree model.

Fig. 10 Confusion matrix of Decision Tree

True Label	Decision Tree			
	1	0.91	0.09	0.00
	2	0.08	0.92	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

H. Random Forest

Unlike decision tree, random forest is made up of many decision trees to make a forest, which contributes to an improved accuracy [33]. Training for each tree is done separately and the outcome is averaged to get the final predicted output. This can be accomplished by a two stage process. Firstly a random tree is created, which involves splitting, prediction in each node and finally introducing randomness. The second step is to come out with final prediction of the outcome from the random forest created. Choice of number of input variables plays an important role in validating the model generated [34]. As mentioned earlier they are more reliable but are not very comprehendible. After applying random forest on the wheat kernel datasets a result of 87.83% accuracy was achieved, the confusion matrix for the random forest model is shown in fig. 11.

Fig. 11 Confusion matrix of Random Forest

True Label	Random Forest			
	1	0.87	0.04	0.09
	2	0.17	0.83	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

I. Artificial Neural Network

To solve highly analytic models Artificial Neural Network is used, which mimics the human neural system [35]. ANN consists of three layers namely Input layer, Hidden layer, Output layer. In ANN weights and learning rates are initialized with some small values near to zero [36]. Input is fed to the input layer and output is calculated by feed forward through the hidden layer.

The output is then compared to the expected output, and the error is calculated. This error is feedback to the hidden layers and the weights are updated accordingly. The decrease in the error can result in the convergence of the algorithm with the termination of learning process. With slower learning rate it is possible to get accurate results. An accuracy of 94.8% was obtained when applying ANN to \the wheat kernel dataset, the confusion matrix for the ANN model is show in fig. 12.

Fig. 12 Confusion Matrix of ANN

True Label	Random Forest			
	1	0.92	0.01	0.07
	2	0.05	0.95	0.00
	3	0.05	0.00	0.95
		1	2	3
	Predicted Label			

J. Convolutional Neural Network

In CNN there are three elements namely input image, feature detector, feature map [37]. To increase the nonlinearity in the data rectified linear unit function is applied. In pooling function a smaller matrix (window) is moved throughout the data and maximum value (maxpooling) is taken out of all the elements of the window [38]. The data thus generated is called pooled feature map. Then flattening is done to generate a single array of the elements out of two dimensional data to further process it using a neural network. This array is passed through a fully connected artificial neural network. The output layer of this neural network represents the predicted classes. Also the error is back propagated to improve accuracy. While applying the CNN model to the wheat kernel dataset an accuracy of 90.46% was obtained. Fig. 13 shows the confusion matrix of CNN algorithm.

Fig. 13 Confusion matrix of CNN

True Label	CNN			
	1	0.81	0.00	0.19
	2	0.04	0.96	0.00
	3	0.07	0.00	0.93
		1	2	3
	Predicted Label			

Fig. 14 shows the performance analysis of the various machine models after applying the wheat kernel datasets. SVR and ANN have a better performance when compared to all the other machine learning models used.

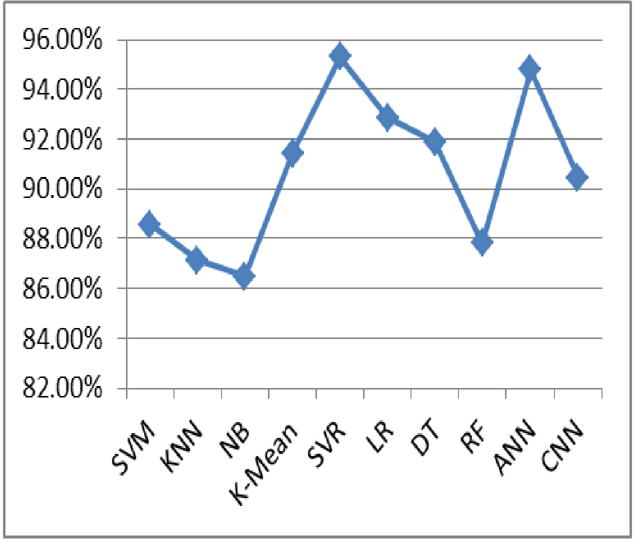

Fig. 14 Performance analysis of ML models

VII. CONCLUSION

While AIoT is the next big thing in the modern agricultural farm management system, applying machine learning algorithm to the data generated from the different inputs of a farm set up with the help of AIoT makes the system more intelligent, provides decisive information and predicts the upcoming outcome. In this study various machine learning algorithm were analysed, each have their own pros and cons from the process to the outcome. This means that the user have to understand each models before applying them to their application to get best out of the model used. For example decision tree though being precise can show a decrease in accuracy when there may be missing and outliers in the datasets but random forest edges here with better accuracy. ANN though being very complex and has a very high training time, but is probably the most stable model for uncorrelated data. SVM has a very long training time though it boost of better performance, the training period can be reduced by improving the data quality.

The various machine learning models can be improved using appropriate performance enhancing algorithms. With depleting resource it is evident that artificial intelligence in the field of agriculture is the future for decision making and production improvement. Artificial intelligence together with AIoT can be called data driven farm management system.

REFERENCE

1. L. Atzori, et al., “The internet of things: a survey,” Comput. Netw. 54 (15) 2787–2805, 2010.
2. H. C. J. Godfray, et al., “Food Security: The Challenge of Feeding 9 Billion People,” Science, 2010.
3. Mohammed H. Almarshadi and Saleh M. Ismail, “Effects of Precision Irrigation on Productivity and Water Use Efficiency of Alfalfa under Different Irrigation Methods in Arid Climates,” Journal of Applied Sciences Research, 7(3): 299-308, 2011.

4. Nathaniel D. Mueller, et al., “Closing yield gaps through nutrient and water management,” Nature, Vol 490, 2012.
5. Medela, et al., “IoT Multiplatform networking to monitor and control wineries and vineyards. In: Future Network and Mobile Summit,” IEEE, pp. 1–10, 2013.
6. Pang, Z., et al., “Value-centric design of the internet-ofthings solution for food supply Chain: value creation, sensor portfolio and information fusion,” Inform. Syst. Front. 17, 289–319, 2015.
7. Mohammad Saeid Mahdavinejad, et al., “Machine learning for internet of things data analysis: a survey,” Digital Communications and Network, 4, 161–175, 2018.
8. Jesús Martín Talavera, et al., “Review of IoT applications in agro-industrial and environmental fields,” Computers and Electronics in Agriculture 142, 283–297, 2017.
9. Andreas Kamilaris, et al., “Agri-IoT: A Semantic Framework for Internet of Things-enabled Smart Farming Applications,” European Union, 2016.
10. Prem Prakash Jayaraman, et al., “Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt,” Sensors, 16, 1884, 2016.
11. Sathe. S et al., “A survey of model-based sensor data acquisition and management, in: Managing and Mining Sensor Data,” Springer, pp.9–50, 2013.
12. Aimad Karkouch, et al., “Data quality ininternet of things: A state-of-the-art survey,” Journal of Network and Computer Applications, 73, 57–81, 2016.
13. D. Barber, “Bayesian Reasoning and Machine Learning,” Cambridge University Press, 2012.
14. K.P. Murphy, “Machine Learning: a Probabilistic Perspective,” MIT press, 2012.
15. Castelli, Mauro, et al., “Supervised Learning: Classification, Reference Module in Life Sciences,” Elsevier, 2018.
16. Serra, et al., “Unsupervised Learning: Clustering, Reference Module in Life Sciences,” Elsevier, 2018.
17. Evans, et al., “International Encyclopedia of the Social &Behavioral Sciences”, second ed., Elsevier, pp. 207–210, 2015.
18. C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn. 20 (3) 273–297, 1995.
19. N. Cristianini and J. Shawe-Taylor, “An Introduction to Support Vector Machines and Other Kernel-based Learning Methods,” Cambridge university press, 2000.
20. B. Scholkopf and A.J. Smola, “Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond,” MIT press, 2001.
21. Zoppis, Italo, et al., “Methods, Kernel, Kernel Methods: Support Vector Machines, Reference Module in Life Sciences,” Elsevier, 2018.
22. H.V. Jagadish, et al., “I distance: an adaptive b+-tree based indexing method for nearest neighbor search,” ACM Trans. Database Syst. (TODS) 30 (2) 364–397, 2005.

A. McCallum, et al., “A comparison of event models for naive bayes text classification, in: AAAI-98,” Workshop on Learning for Text Categorization, vol. 752, Citeseer, pp. 41–48, 1998.

23. H. Zhang, “The optimality of naive bayes,” AA 1 (2), 3, 2004.
24. Kanungo, T., et al., “An efficient k-means clustering algorithm: analysis and implementation.” IEEE Trans. Pattern Anal. Mach. Intell. 24, 881–892, 2002.

A. Likas, et al., “The global k-means clustering algorithm,” Pattern Recognit. 36 (2) 451–461, 2003.

25. Coates and A.Y. Ng, “Learning feature representations with K-Means, in: G. Montavon, G.B. Orr, K.R. Müller (Eds.), Neural Networks: Tricks of the Trade,” Lecture Notes in Computer Science, vol. 7700, Springer, Berlin, Heidelberg, 2012.
26. Debasish Basak, et al., “Support Vector Regression,” Neural Information Processing – Letters and Reviews, Vol. 11, No. 10, 2007.
27. Alex J. Smola and Bernhard Scholkopf, “A tutorial on support vector regression,” Statistics and Computing 14: 199–222, 2004.
28. Hosmer, D.W and Lemeshow, S, “Applied Logistic Regression.” John Wiley & Sons, 2004
29. L. Breiman, “Random forests,” Mach. Learn. 45 (1) 5–32, 2001.
30. Daniel F. Polan, et al., “Tissue segmentation of Computed Tomography images using a Random Forest algorithm: a feasibility study,” Phys Med Biol. 61(17): 6553–6569, 2016.
31. Saher Esmeir and Shaul Markovitch, “Anytime Learning of Decision Trees,” Journal of Machine Learning Research 8, 891-933, 2007.
32. S. Suthaharan, “Decision tree learning, in Machine Learning Models and Algorithms for Big Data Classification.” Springer, pp. 237–269, 2016.
33. McCulloch W.S. and Pitts W, “A logical calculus of the ideas immanent in nervous activity.” Bull. Math. Biophys, 5, 115–133, 1943.
34. Zafer CÖMERT and Adnan Fatih KOCAMAZ, “A study of artificial neural network training algorithms for classification of cardiotocography signals,” Journal of Science and Technology 7(2) 93–103, 2017.
35. Krizhevsky. A et al., “Imagenet classification with deep convolutional neural networks.” In Advances in neural information processing systems (pp. 1097-1105), 2012.
36. Sakshi Indolia, et al., “Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach,” Pooja Asopa, International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science 132, 679–688, 2018.
37. Ramos, P.J et al., “Automatic fruit count on coffee branches using computer vision.” Comput. Electron. Agric, 137, 9–22, 2017.
38. Su, Y.; et al., “Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J,” Biol. Sci, 24, 537–547, 2017.
39. Kung, H.-Y et al., “Accuracy Analysis Mechanism for Agriculture Data Using the Ensemble Neural Network Method.” Sustainability, 8, 735, 2016
40. Daniel, Jiménez, et al., “A survey of artificial neural network-based modeling in agroecology.” Soft Comput. Appl. Ind. 247–269, 2008.
41. Huang, et al., “A multiple crop model ensemble for improving broad-scale yield prediction using Bayesian model averaging.” Field Crops Res. 211, 114–124, 2017.
42. Neto, João Rossi, et al., “Use of the decision tree technique to estimate sugarcane productivity under edaphoclimatic conditions.” Sugar Tech 19 (6), 662–668, 2017.
43. Liu, Xueli, et al., “Analysis of grain storage loss based on decision tree algorithm.” Procedia Comput. Sci. 122, 130–137, 2017.
44. Grinblat, G.L. et al., “Deep learning for plant identification using vein morphological patterns.” Comput. Electron. Agric., 127, 418–424, 2016.
45. Evan J. Coopersmith, et al., “Machine learning assessments of soil drying for agricultural planning,” Comput. Electron. Agric. 104, 93–104, 2014.
46. M.S. Sirsat, et al., “Classification of agricultural soil parameters in India,” Comput. Electron. Agric. 135, 269–279, 2017.
47. Amarendra Goapa, et al., “An IoT based smart irrigation management system using Machine learning and open source technologies,” Computers and Electronics in Agriculture 155, 41–49, 2018.
48. Ramesh Vamanan and K.Ramar, “Classification of Agricultural Land Soils A Data Mining Approach,” International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 1, 2011.
49. Morellos, A.; et al., “Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy.” Biosyst. Eng., 152, 104–116, 2016.
50. Johann, et al., Soil moisture modeling based on stochastic behavior of forces on a no-till chisel opener. Comput. Electron. Agric, 121, 420–428, 2016.
51. Nahvi, B, et al., “Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature.” Comput. Electron. Agric., 124, 150–160, 2016.
52. Oluseun Adetola Sanuade, et al., “Using artificial neural network to predict dry density of soil from thermal conductivity”, RMZ – M&G, Vol. 64, pp. 237–012, 2017.
53. Dhiman Mondal, et al., “Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy based binning and Naive Bayes classifier after identification of leaves,” Comput. Electron in Agric 142, 485–493, 2017.
54. Adhao Asmita Sarangdhar and V. R. Pawar, “Machine learning regression technique for cotton leaf disease detection and controlling using IoT,” International conference of Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2017.

55. Thomas Truong, et al., “An IoT Environmental Data Collection System for Fungal Detection in Crop Fields,” IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017.
56. Ferentinos, K.P. “Deep learning models for plant disease detection and diagnosis.” Comput. Electron. Agric, 145, 311–318, 2018.

57. Han, Liangxiu, et al., “Automatic detection and severity assessment of crop diseases using image pattern recognition.” Emerging Trends and Advanced Technologies for Computational Intelligence 283–300, 2016.
58. Chung C.L, et al., “Detecting Bakanae disease in rice seedlings by machine vision.” Comput. Electron. Agric. 121, 404–411, 2016.
59. Huang, Tisen, et al., “Detecting sugarcane borer diseases using support vector machine.” Inform. Process. Agric. 5 (1), 74–82, 2018.
60. Moshou, D. et al., “Automatic detection of “yellow rust in wheat using reflectance measurements and neural networks.” Comput. Electron. Agric., 44, 173–188, 2004.
61. Glezakos, et al., “Plant virus identification based on neural networks with evolutionary preprocessing,” Comput. Electron. Agric, 70 (2), 263–275, 2010.
62. P.A. Gutiérreza, et al., “Logistic regression product-unit neural networks for mapping Ridolfia segetum infestations in sunflower crop using multitemporal remote sensed data,” Comput. Electron. Agric, 64, 293–306, 2008.
63. Yano, et al., “Identification of weeds in sugarcane fields through images taken by UAV and Random Forest classifier.” IFAC-PapersOnLine 49 (16), 415–420, 2016.
64. Pantazi, X.E, et al., “Evaluation of hierarchical self-organising maps for weed mapping using UAS multispectral imagery.” Comput. Electron. Agric., 139, 224–230, 2017.
65. Patil, A.P and Deka, P.C. “An extreme learning machine approaches for modeling evapotranspiration using extrinsic inputs.” Comput. Electron. Agric., 121, 385–392, 2016.
66. Mohammadi, K et al., “Extreme learning machine based prediction of daily dew point temperature.” Comput. Electron. Agric., 117, 214–225, 2015.
67. Małgorzata Charytanowicz, et al., “Complete Gradient Clustering Algorithm for Features Analysis of X-ray Images,” 2010

AUTHORS PROFILE

Abraham Sudharson Ponraj is currently a PHD student and working as an Assistant Professor at Vellore Institute of Technology, Chennai, India.

Dr.Vigneswaran. T received his PHD degree from SRM University, India in 2009. He is currently working as a Professor in Vellore Institute of Technology, Chennai, India.

.