International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Machine Learning in Agriculture Application: Algorithms and Techniques

Meeradevi, Sindhu N, Monica R Mundada

Abstract: ***Machine learning techniques with high performance computing technologies can create various new opportunities in the agriculture domain. This paper does comprehensivereview of various papers which are concentrating on machine learning (ML) and deep learning application in agriculture. This paper is categorized into three sections a) Yield prediction using machine learning technique b) Price prediction c) Leaf disease detection using neural networks. In this paper we study the comparison of neural network models with existing models. The findings of this survey paper indicate Deep learning models give high accuracy and outperform traditional image processing technique and ML techniques outperforms various traditional techniques in prediction.***

I. INTRODUCTION

Using DL and ML will help to tackle lot of challenges in agriculture which may be related to production or food security. In this review paper the following research questions are analyzed.

1. Which crop related problem it addresses?
2. Which deep learning models were used for analysis?
3. Which models were compared with their proposed model?
4. Which attributes in the dataset were considered for preprocessing?
5. What is the performance of model they employed and did they test their dataset with other models?

DL is complex model which we used for leaf disease detection. Because of its complexity it can solve more complex problems with good accuracy and less error in the model. DL has feature extraction technique, extracts features from raw data. DL consists of various features like pooling, activation, convolutions, fully connected layers, gates, memory cells and many more. DL is hierarchical in structure and performs classification and prediction well. We have employed CNN for leaf disease detection. Architectures employed in CNN are AlexNet, VGG16, GoogleLeNet and so on.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Meeradevi*, Dept. of CSE M S Ramaiah Institute of Technology Bangalore,India.
Sindhu N, Dept. of CSE M S Ramaiah Institute of Technology Bangalore,India.
Monica R Mundada, Dept. of CSE M S Ramaiah Institute of Technology Bangalore,India.

Machine Learning (ML) helps to manage the data with heterogeneous information coming from real time data which is collected using sensors. The focus of this survey paper is to concentrate on design and deployment of time-series dataset for forecasting of crop yield, crop price and leaf disease of plant. The data is preprocessed and applying efficient ML and DL algorithms to check the accuracy by comparing with various algorithms.

II. RELATED WORK

In [1] paper investigates the author discuss about Zambia's Crop Forecast Survey (CFS) and its limitations. In this proposed method at the end of season harvest the data is collected. The data about crop yield and area of production is collected from the actual harvest from households in real time. Using this data a forecast model for yield is developed using Extreme Gradient Boost. To test the accuracy values RMSE is used. From results it is found that CFS accuracy is less compared to gradient boost. In [2] this paper, the corn yield prediction is made using deep neural network model from environment and genotype data. The dataset consist of 148,452 samples for training. Yield performance is calculated using yield and check yield using deep neural networks. Neural network in his paper is also used for weather prediction. The model accuracy is tested using RMSE 11 % error for single hidden layer with 300 hidden neurons. In [3] The paper compares the accuracy in forecasting the price of agricultural products using Auto ARIMAmodel and back propagation (BP) network method. A web crawler technology is also used to get prediction of price for the crop commodities. . The data is collected about the cucumber crop. The data obtained is normalized in order to scale it down to (0, 1) range. The dataset is consist of linear data on which ARIMA algorithm is applied for forecasting. The ARIMA model does not consider the rapid changes in the prize of the crop that arises due to instability in seasons. In order to consider such changes BP network model is applied to the data. BP network model contains set of input, hidden and output layers. The number of output layers is increased in order to check whether the price prediction for long-term is accurate or not.RNN model used to process sequential data is applied on the normalized data to obtain the price. The results show that the foresting for monthly, weekly and daily price of crop ARIMA does well if it is for short term but accuracy is not proper for long term. So, BP neural network is used for such issues which gives high accuracy.

In [4] Agribusiness is the foundation of Indian economy. The agrarian yield is fundamentally relies uponclimate conditions like temperature, rainfall, biomass. The agriculturists essentially require an advantageous incite to anticipate the future yield profitability and an investigation to help the farmer's to enlarge harvest creation in their products.

Despite the fact that a great deal of research has been led for building up the choice emotionally supportive network for agriculturists, the greater part of the examination centre around the harvest administration, edit illness administration and product yield estimating. The agriculturist's harvest determination at the prior stage is a standout amongst the most critical components since proper product choice at the prior stage will help ranchers to enhance edit administration and product yield. The results of Fuzzy C-Means is used for implementation which is using clustering method. It gives less error rate in terms of degree of membership means which data points have higher probability or how much similarity is there in between the data points.

In [5] the study was done for Maharashtra zone for wheat, rice, jowar, bajra, pulses the main aim was to achive increased crop yield. The dataset parameters include min and max temperature, evop-transpiration, area under production, crop yield for previous years. Artificial neural network is used for prediction with multilayer preceptron. For the analysis of dataset Weka tool is used and confusion matrix was generated using multilayer preceptron. The three layer feed forward network is used for training the dataset and authors could achieve the accuracy of 97.54% An ANN with 10 fold cross validation function is used for subset of dataset. For training and testing the linear data linear regression is used and for non linear data ANN is used. The accuracy of the model is obtained using RMSE, RAE.

In [7] In this paper, price forecasting for Mysore district is done. As farmers are not getting the expected price to improve this problem price forecasting is done using ANN with feed forward back propogation scheme. The attributes include pressure, temperature,soil type, humidity, seed variety. MSE is used to evaluate the accuracy of the model. The proposed model is compared with decision tre, genetic algorithm. The accuracy with proposed model is high.

In [8] the proposed method mainly aims at creating a system which uses the new generation high computing technologies for detection of plant diseases. By using the fast GPU's and embedded processors we can considerably increase the accuracy and the rate of output results in image classification. Also, by making this image classification available locally on mobile phones, it will be very useful for farmers. The main method the paper focuses on is using the CNN with 'n' number of hidden layers, which would normally take days to train, and then training it on the new faster processors and CPU's. We are also able to observe that the model does considerably better than using other conventional methods.

In [9] the proposes a deep learning technique using a jaya algorithm for classification of paddy leaf diseases, which uses a feedback loop system in the post processing step. The main methodology mainly includes five steps-Image acquisition, Pre-processing, Segmentation, Feature Extraction and Classification. And the feedback loop system takes the values from the classification back to the Segmentation unit. The Segmentation is based on k-means clustering. This gives an enhanced deep learning method where the error is reduced considerably. The method is formally called the DNN_JOA method of image classification.

In [10] it proposes a method of 10-fold cross validation method strategy which will give a higher accurate output then one with the straightforward CNN with 10 or more hidden layers. This also saves computational time and memory to save the 'n' weights achieved from training the image data. They use the standard BP-algorithm and then a SVM and then to finally decrease the SME by using the particle swarm optimization (PSO). CNN can be designed using AlexNET or GoogleNET or the standard libraries in Python like Tensorflow and Keras or in MATLAB or SimuLink. In [11] K-Means Clustering method is used for the segmentation of pomegranate leaf disease. The image of the leaf is segmented using K-Means Clustering method. The dataset is a mixture of different healthy and unhealthy images of leaves. The image data is first preprocessed and then segmented into three sub-images each containing different features. The diseased leaves are detected after this step. Image processing and segmentation of images into three sub-images each makes it easy for the detection of diseases associated with it. The process used in this paper does not provide a great accuracy rate. The sub-images obtained after segmentation are not clear enough to help detect diseases.

In [12] potato leaf diseases detection and classification are based on neural networks and make use K-Means Clustering for segmentation. Various leaves of healthy and unhealthy types are used as the dataset. Algorithms are developed to extract more than 24 features. The gray level co-occurrence matrix (GLCM) is used to extract texture features. To identify whether the leaf is diseased or not and to classify the diseases a back propagation neural network (BPNN)- based classifier is made use of. The features are used to train the network which is hence used to detect and classify the leaf samples. The classification has an accuracy of more than 92%. This method is quick, automatic and accurate for leaf disease detection and classification. In [13] proposes a method for the detection of diseases in cotton crop leaves. The image data is pre-processed, segmentation is carried out using thresholding method which takes RGB input image resulting from K-Means Clustering method. Pixels values of red, green and blue bands are read separately. Feature extraction makes use of GLCM methodology. In [14] The diseased portion of the plants are identified clearly in the green band in the thresholding method. Using this method, only the region of interest (diseased area) can be extracted. Diseases can be classified using Euclidean classifier.

Table 1: Applications of Machine Learning in Agriculture

Reference	Problem Description	Crop	Algorithm /Models	Attributes used for data analysis and prediction
[15]	Optimize the wheat production using existing soil and climate data.	Wheat, Rice, Jowar, Groundnut,Cotton	PAM, CLARA and DBSCAN,MLR	crop, district, year, production, area, yield
[16]	Crop yield estimation is done using a dat mining techniques.	Rainfall as key parameter	k-means, Multiple linearregression, KNN,MLR,ANN	Year, Rainfall, Areaof Sowing and Production
[17]	Identifying the sowing time for farmers and seed quality soil based on requirements which crop can grown. Proposed methods help farmers to increase crop yield and profit.	millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander	ANN ,Naïve Bayes, , Random Tree CHAID	Soil Ph, soil color, depth, soil texture, water holding capacity etc.,
[18]	Precision agriculture for vegetable production	Beijing vegetable	stochastic frontier production function	Seed quality, soil properties (NPK), temperature
[19]	Identify different irrigation schedules for rice crop and recommendation of NPK for crop. study investigated the interactive effects of rice cultivation methods with different irrigation schedules and plant density on the uptake and concentration of micronutrients in India	Rice	Recommendation of uptake of nutrients	Nitrogen Potassium, Phosphorus, sulphur, iron, aluminium
[20]	a comparative analysis is done with various classification technique and concluded as Bagging algorithm of classification technique is best suited for soybean crop prediction.	Soya bean	SVM, Random Forest, Neural Network, REPTree, Bagging, and Bayes.	Soil properties and weather data
[21]	The proposed method uses eXtensible Crop Yield Prediction Framework (XCYPF). This framework has provision for crop selection and other variables dependent on crop for precision agriculture.	Rice and Sugarcane	eXtensible Crop Yield Prediction Framework (XCYPF) is proposed	NDVI, Temperature
[22]	Crop yield prediction using historical data	Paddy, wheat	Self-organizing maps and multi-dimensional scaling techniques (Sammon's mapping)	Area under cultivation,total production , productivity

[23]	Maximize the crop yield suing data mining techniques.	Rice, Wheat, Mung	Crop selection algorithm	weather, soil type, water density, crop type, sowing time, plantation days and predicted yield rate
[24]	Nutrition System for desired yield targets of maize has been developed for deep black calcareous, Pilamedu soil series (Typic Haplustert) of Tamil Nadu	Maize	Soil type recommendation using real time testing of soil	N, P and K, fertilizer doses of N, P2O5 and K2O and FYM levels

III. METHODOLOGY

Figure 1 shows the classification data mining algorithms were applied on a data set containing different crops, rainfall, production, area and pH values. The dataset includes rainfall, soil pH, soil attributes.

The raw dataset is collected and data is preprocessed to remove noise. On the processed dataset data mining algorithms are applied and accuracy of the dataset is calculated using F-measure. Comparative study is done to analyze and understand various other algorithms. And finally best algorithm is found based on accuracy Weka tool is used for analysis of dataset.

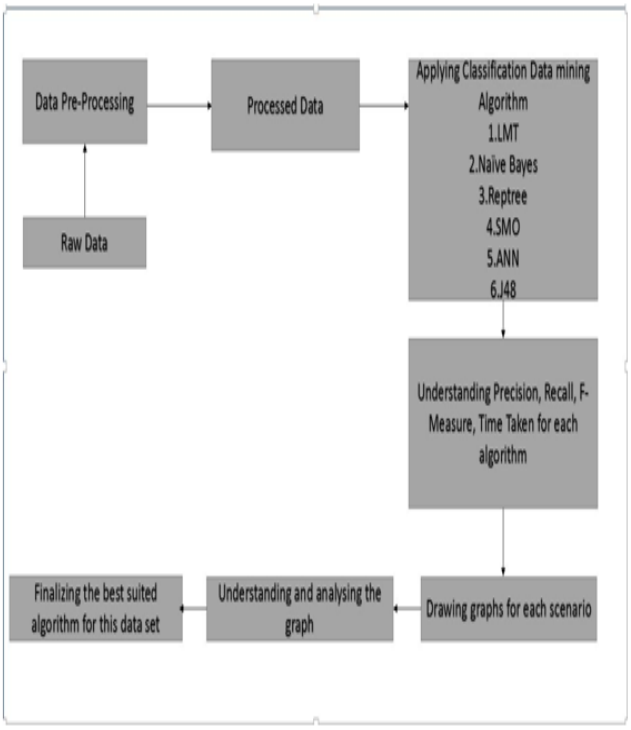

Fig 1: System Architecture

IV. IMPLEMENTATION

A. Artificial neural networks

As shown in figure 2 the WEKA tool is used analysis for the dataset with attributes NPK values of soil,crop yield, district, are of production, rainfall, pH soil type for yield prediction using ANN model for multilayer preceptron. The districted names are shown in the figure for which the prediction is done.

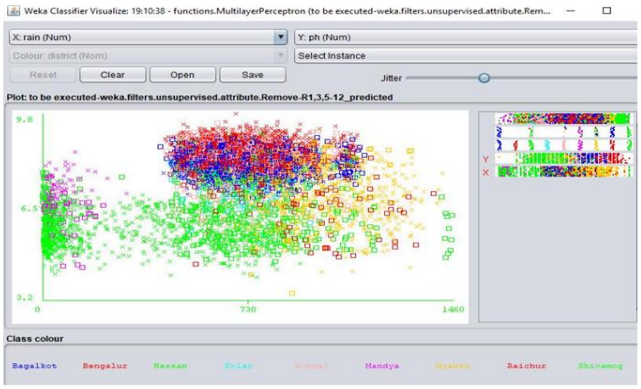

Fig 2: Analysis using ANN multilayer perceptron

B. Naive Bayes

As shown in figure 3 Naive bayes is an algorithm for prediction. Naïve Bayes algorithm totally depends on the probability model's precise nature.

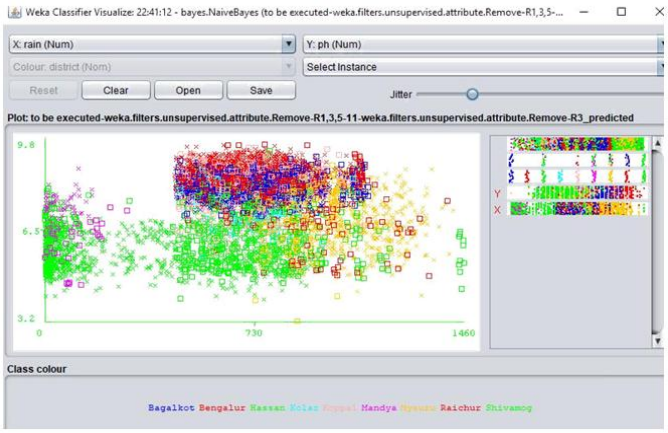

Fig 3: Analysis using Naïve bayes algorithm

C. J48

Figure 4 shows the implementation using J48 algorithm. Whichis an decision tree algorithm. This Algorithm is continuation of Quinlan's ID3 algorithm which was enhanced as C4.5 by himself. An ID3 algorithm being re utilized for diminishing of decision trees, derivations of rules, value extents and so on. J48 algorithm is used for statistical analysis.

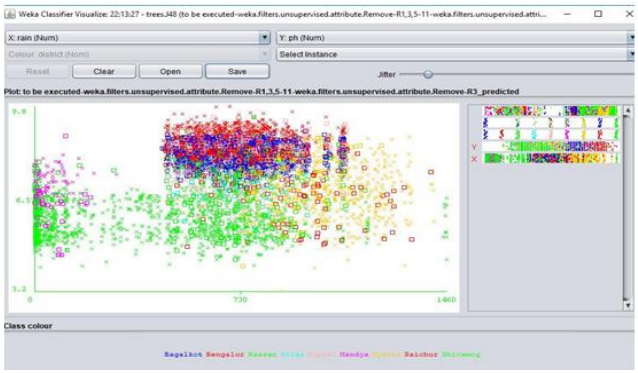

Fig 4: Analysis using J48 tree algorithm

D. The Sequential Minimal Optimization (SMO) algorithm

Figure 5 shows SMO algorithm implementation using Weka tool. The power of this technique resides in the fact that the optimization problem for two data points admits an analytical solution, eliminating the need to use an iterative quadratic programming optimizer as part of the algorithm. The requirement that the condition P` i=1 yiαi = 0 is enforced throughout the iterations implies that the smallest number of multipliers that can be optimized at each step. Whenever one multiplier is updated, at least one other multiplier needs to be adjusted in order to keep the condition true.

Fig 5: Analysis using SMO algorithm

E. RepTree

Figure 6 shows RepTree algorithm usich is using regression tree for analysis. This algorithm is the fast decision tree algorithm.

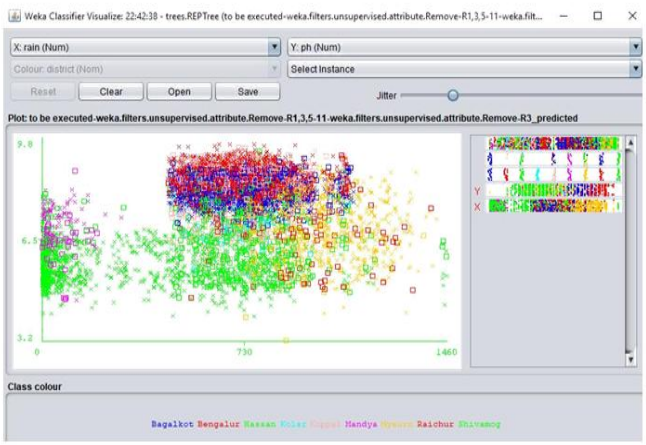

Fig 6: Analysis using REPTree algorithm

F. LMT

Figure 7 shows LMT algorithm which is supervised algorithm uses classification. Logistic model trees use a decision tree that has linear regression models at its leaves to provide a section wise linear regression.

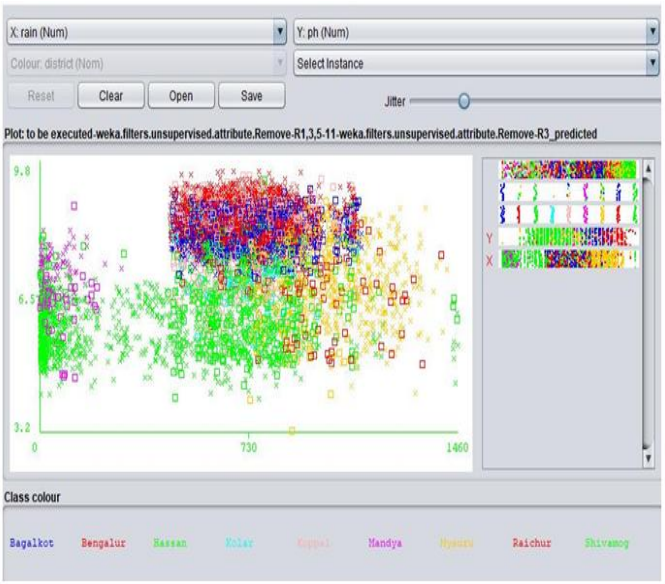

Fig 7: Analysis using LMT algorithm

V. RESULTS AND DISCUSSION

The cross validation method used to analyze the datasets. Various performance measures for all the datasets mentioned. . Comparative analysis of various decision tree classifications, simulation results as follows.

Table 2: Comparison of accuracy

Algorithm	Precision	F1-Measure	Recall	Time-Taken (sec)
LMT	0.559	0.59	0.633	11.62
Naïve Bayes	0.539	0.571	0.614	0.02
RepTree	0.576	0.596	0.635	0.1
SMO	0.548	0.566	0.616	0.65
ANN	0.591	0.574	0.615	9.74
J48	0.582	0.609	0.652	0.15

Precision (P): It is to determine of exactness. It is the ration of the predicted positive cases that were correct to the total number of predicted positive cases.

$$P = \frac{TruePositive}{TruePositive + FalsePositive}$$

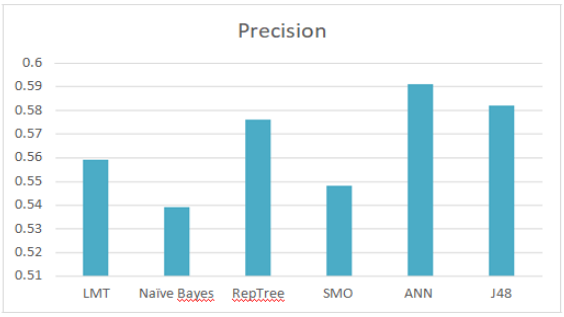

Fig 8: Accuracy of various algorithms using precision

Figure 8 shows that ANN is a better algorithmn to use , as it shows the highest and accurate precision. The second best to consider would be J48. Naïve bayes takes the least score , so it is not very reliable to use for this research.

Recall(R): Recall is determine of completeness. It is the proportion of positive cases that were correctly recognized to the total number of positive cases. It is also known as sensitivity or true positive rate (TPR).

$$R = \frac{TruePositive}{TruePositive + FalseNegative}$$

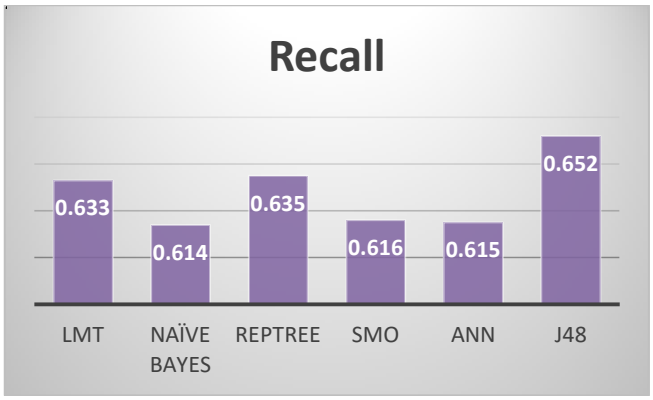

Fig 9: Comparison of accuracy with Recall

Figure 9 shows in terms of Recall J48 is again the better approach and next in line will be the LMT algorithm.

F-Measure: The harmonic mean of precision and recall. It is an important measure as it gives equal importance to precision and recall.

$$F - Measure = \frac{2 \times Recall \times Precision}{Precision + Recall}$$

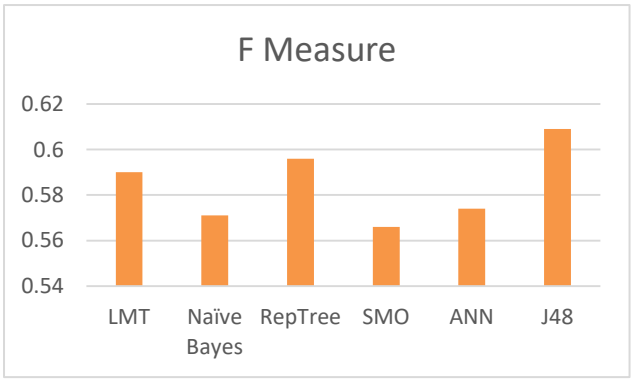

Fig 10: Comparison of accuracy with F-measue

Figure 10 shows the combination of Recall and Precision makes the F-Measure, J48 could be a better algorithm as it has the highest performance in both recall and precision.

Time-Taken-

The running time of an algorithm for a specific input depends on the number of operations executed. The greater the number of operations, the longer the running time of an algorithm. We usually want to know how many operations an algorithm will execute in proportion to the size of its input, which we will call N.

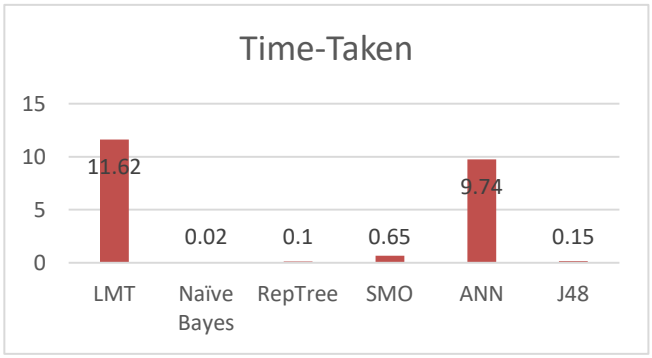

Figure 11: Time taken by algorithms

Time taken for the algorithm to process could be one of the criteria to consider as it will tell which algorithm is more efficient and faster also. In our case LMT is the slowest as it takes about 11.62 seconds to process and the quickest is Naïve bayes.

VI. CONCLUSION

This survey paper introduces cheap, practical and easy to develop and use model using neural network. This will help to analyze various papers with different parameters in dataset how this help to increase accuracy of dataset. Applying ML and DL on sensed data the crop yield, crop price and other various other prediction like rainfall, temperature etc., can be made. These techniques have made the predictions easy by providing high recommendations for correct decision making by farmers and government agencies to improve crop productivity.

REFERENCES

1. Zhao, Yi, Noemi Vergopolan, Kathy Baylis, Jordan Blekking, Kelly Caylor, Tom Evans, Stacey Giroux, Justin Sheffield, and Lyndon Estes. "Comparing empirical and survey-based yield forecasts in a dryland agro-ecosystem."*Agricultural and Forest Meteorology* 262 (2018): 147-156.
2. Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." In *INFORMS InternationalConference on Service Science*, pp. 139-147. Springer, Cham, 2019.
3. Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang and F. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," in *IEEETransactions on Computational Social Systems*,vol.6,no.3,pp.547-553,June2019.doi:10.1109/TCSS.2019.2914499
4. Verma, Akanksha, Aman Jatain, and Shalini Bajaj. "Crop Yield Prediction of Wheat Using Fuzzy C Means Clustering and Neural Network."*International Journal of Applied Engineering Research* 13, no. 11 (2018): 9816-9821.

5. N. Gandhi, O. Petkar and L. J. Armstrong, "Rice crop yield prediction using artificial neural networks,"*2016IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, 2016, pp. 105-110.
6. doi: 10.1109/TIAR.2016.7801222
7. Noureen, S. Atique, V. Roy and S. Bayne, "Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A case study of small scale agricultural load,"*2019 IEEE 62nd International MidwestSymposium on Circuits and Systems (MWSCAS)*, Dallas, TX, USA, 2019, pp. 521-524.doi:10.1109/MWSCAS.2019.8885349
8. Mohan, Pushpa, and Kiran Kumari Patil. "Deep Learning Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application."*International Journal of Intelligent Engineering and Systems* 11, no. 4 (2018): 167-176.
9.] Konstantinos P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis", Computersand Electronics in Agriculture 145:311-318 · February 2018
10. S. Ramesh , D. Vydeki, "Recognition and Classification of Paddy Leaf Diseases using Optimized Deep Neural Network with Jaya Algorithm", Information Processing in Agriculture,KeAi(Chinese roots Global Impact),2019
11. Yang Lu , Shujuan Yi , Nianyin Zeng , Yurong Liu , Yong Zhang, "Identification of Rice Diseases using Deep Convolutional Neural Networks", Neurocomputing, Volume 267, 6 December 2017, Pages 378-384
12. Arda Aydin, Anıl Çay, Burak Polat, "Image Analysis Applications in Plant Growth and Health Assessment", Journal of Agricultural Faculty of Mustafa Kemal University, 22(1):112-126 (2017)
13. T. N. Shaikh, Dr. S.M. Mukane, "Segmentation of Pomegranate Leaf for Detection of Disease Using Image Processing", International Journal of Innovative Research in Computer and Communication EngineeringVol.5, Special Issue 4, June 2017
14. Girish Athanikar, Priti Badar, "Potato Leaf Diseases Detection and Classification System", International Journal of Computer Science and Mobile Computing, Vol. 5, Issue. 2, February 2016, pg.76 – 88
15. Pranita P. Gulve, Sharayu S. Tambe, Madhu A.Pandey, S.S.Kanse, "Leaf Disease Detection of Cotton Plant using Image Processing Technique", Techno-Societal 2018pp55-65
Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data", Dept. of Computer Science NMIT, 2017.
16. D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data",JNTUH College of Engineering, Andhra Pradesh, India, Vol.2, Issue 9, September 2013.
17. S Pudumalar,E.Ramanujam,R. Harine Rajashree, C Kavya, T Kiruthik, J Nisha, "Crop Recommendation system for Precision Agriculture", Dept. of Information Technology, Thiagarajar College of Engineering , September-2016.
18. MotiurRahman M, Haq N, Rahman RM. Application of data mining tools for rice yield prediction on clustered regions of Bangladesh. IEEE. 2014; 2014:8–13.
19. Anchal Dass, Subhash Chandra et,all "Agronomic fortification of rice grains with secondary and micronutrients under differing crop management and soil moisture regimes in the north Indian Plains" Article In Paddy Water Environ (2017) 15:745–760 DOI 10.1007/s10333-017-0588-9,Springer.
20. Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia ,Sanya Bhardwaj (2015),"Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture", Innovations in Information, Embedded and Communication systems (ICIIECS).
21. Aakunuri Manjula, Dr.G .Narsimha (2015), "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction", Conference on Intelligent Systems and Control (ISCO).
22. Yash Sanghvi, Harsh Gupta, Harmish Doshi, DivyaKoli, AmoghAnshDivyaKoli, Umang Gupta (2015), "Comparison of Self Organizing Maps and Sammon's Mapping on agricultural datasets for precision agriculture",International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS).
23. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).
24. R. Suresh, R. Santhi, "Soil Test Crop Response Based Integrated Plant Nutrition System for Maize on Vertisol", *International Journal of Current Microbiology and Applied Sciences ISSN: 2319-7706* Volume 7 (8) 1631-1641(2018).

AUTHORS PROFILE

Mrs. Meeradevi obtained B.E in Information Science & Engg and M.Tech degree in Computer Science &Engg from VTU University, India. Currently, I am pursuing Ph.D from VTU University. I have published research papers in various International Journals and Conferences. My area of interest includes Machine Learning, Analytics, Wireless sensor networks, IoT and Computational Intelligence. I am IEEE member and Life member of ISTE.

Miss Sindhu has obtained BE in Computer Science and Engineering, and is currently pursuing M.Tech degree in Computer Network and Engineering from VTU, India. The areas of interest include Machine learning, Convolutional neural networks, Cyber security, IoT. She is presently working on Dockers and Kubernetes, application deployments.

Dr. Monica R. Mundada obtained her Ph.D. degree in Computer Science from Dr.M.G.R University. She completed her B.E degree in Computer Science &Engg from Gulbarga University and M.Tech degree in Computer Science &Engg from VTU University. Her areas of interest include Machine Learning, Big Data Analytics, and IOT. She has published research papers in various international Journals and Conferences. Presently she is working as Professor in the Department of Computer Science &Engg at MSRIT, Bangalore.