International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

ANALYSIS ON APPLICATIONS OF MACHINE LEARNING TECHNIQUES TO GENERATE CROP PREDICTIONS WITH BETTER PRECISION

Om Sai Madala[1], Attanti Brahmendra Charan[2]

[1,2]Student in Computer Science and Engineering at JNTUH COLLEGE OF ENGINEERING HYDERABAD

---***---

ABSTRACT: In most parts of India, agriculture has become a risky business and farmers suffer a lot due to unpredictable yield. The risk is mainly due to the availability of water resources for cultivation and getting profitable prices in the market. Prices alter between very high and very low, so crop planning has become very important for farmers to minimize losses. Machine learning techniques can help to understand the underlying patterns from mass data and these patterns can be used to help farmers with crop planning, also it would reduce the risk of crop failure and guarantee a maximum profit for farmers to sustain their livelihood. But human knowledge cultivation is not sufficient to cater to the demanding need due to the rapid growth in the world's human population. To address this problem, this paper has studied the use of machine learning tools. It experimented with more than 0,3 million data. This dataset identifies key parameters of cultivation collected from the Bangladesh Agriculture Department. This study compared the number of machine learning algorithms to neural networks.

Keywords: Machine learning, Crop prediction, Precision.

1. INTRODUCTION

The application of scientific methods and automated training in problem-solving was not a necessary trend in this technical era. In the areas of agriculture, medicine, education, etc., this made the first world countries more advanced. The use of technology leads to a booming economy,[1] that enriches it daily, whereas the excuse for turning the eye on it is too costly in developed countries such as Bangladesh. While agriculture is a key part of our economy, agriculture does not generally have the preferred output due to the fact that farmers do not use technology and scientific methods. The success of agricultural growth is supposed to be based on an economically viable and eco-friendly farming technology, which means that the available resources will be constantly adapted[1].

By 2050, the world's population is estimated to be over 9 billion by 2050, requiring substantial improvements in the output of large crops that contribute to global food security. The main objective of the majority of plant breeding programs for important plants like soya (Glycine max) is to boost yield and is a major food and feed source of protein or oil. However, the measurement of prime features such as yield, by a combination of qualitative and quantitative characteristics, requires time and labor in large reproductive populations of several thousand genotypes. The increase in yield is known as a highly complex and not linear process for genetic and environmental factors (Collins et al., 2008). As a result, secondary breeding approaches that have strong connections with the primary trait (e.g. yielding components and reflecting belts) allow plant breeders to recognize successful lines in early growing stages in an efficient and efficient way.

The aim of this research is therefore to explore opportunities that allow farmers to select crops efficiently and maximize their yields[10] through the use of a modern technology-driven prediction model. The search for alternatives to traditional farming systems is very important to address the ever-increasing food demand of excessive populations such as Bangladesh. In this document, the six main plant yield predictions, including three rice variations and 3 corn, jute, and potato, were considered for several machine learning techniques.

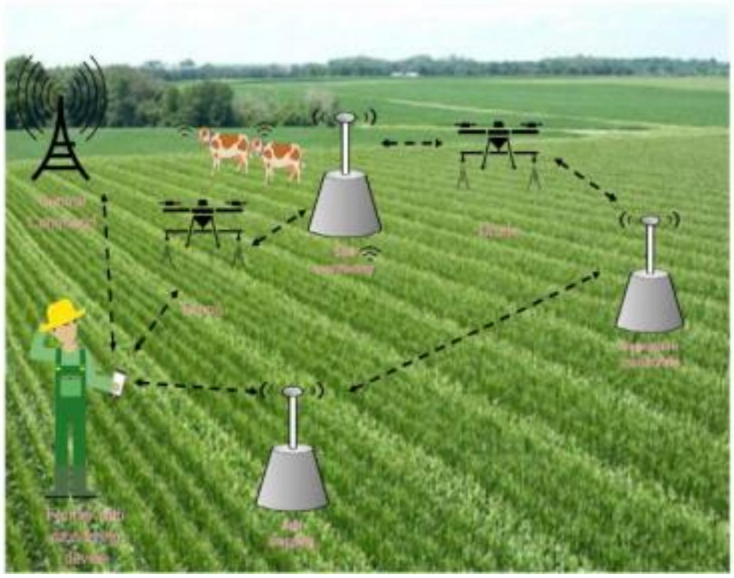


Figure 1. Precision Agriculture

MANY influential parameters have been accumulated, including many environmental parameters, diverse fertilizers, land type, soil structure, and essential land types, as a result of the successful crop selection and yield forecasting identified. To offer an effective solution to these parameters, 0.3 million data is also collected for machine educative models. [11-13]

The study investigated the prediction model using various learning techniques, namely the deep neural grid, the algorithm supporting vector machines, the random forest algorithm, and logistic regression. Environmental parameters, soil types, and composition are also found to have a notable effect on overall crop production. Many deep neural networks dedicated to hidden layers showed improved accuracy in comparison to other methods explored.

Precision agriculture is a technologically capable, sustainable farm management system, which is also abbreviated as digital agriculture. In essence, the modern support of agricultural decision support IT, software, and smart embedded devices are adopted[14] as demonstrated in Figure 1. Mechanized agriculture and the green revolution are two key elements of the first and second agricultural revolutions. The third revolution in agriculture[15] plays a major role in precision agriculture.

2. LITERATURE REVIEW

In[1], Ji et al. have created an agricultural administration model that wants to predict rice yields in planning with accurate and easy estimation techniques. The research aimed at (1) assessing the successful anticipation of the Fujian rice yield for the mountain district atmosphere (ANN) models, (2) assessing the performance of the ANN model relative to different parameters, and (3) comparing the effectiveness of multiple linear regression models with the ANN models.

Shastry and Sanjay submitted in [2] that on test datasets with the higher R2 stats and the lowest percentage prediction error, the Custom Artificial Neuronal Networks (C-ANN) model is better than MLRs and D-ANN models. Moreover, the prediction of crop yield is very important in the field of agriculture. In this study, we estimated the yield of wheat in terms of its various parameter and the improvement of wheat yield in the application of the C-ANN model.

The model for rice yields was introduced by Jabjone and Jiamrum [3] in Phimai District, Thailand. A multi-layer feeding neural networking model is used for model generation in conjunction with the backpropagation algorithm. Data are also used for model training between 2002 and 2007. The rice yield was predicted from 2008 to 2012 using this model. The input data from six weather factors were also used, namely rainfall, distribution of water, evaporation, perspiration, temperature, moisture, and wind speed. Using the Penman-Monteith equation, evapotranspiration (ET) was identified. Its consequence indicated that the lowest Root Mean Squared Error (RMSE) (10,57) and MAPE value were ANNs (8, 19, and 17). (2.3). There is a linear relation to rice output forecast for ANN (8, 19, and 17) and actual data (R2=0.99). Consequently, their model for predicting rice yield was precise and adequate.

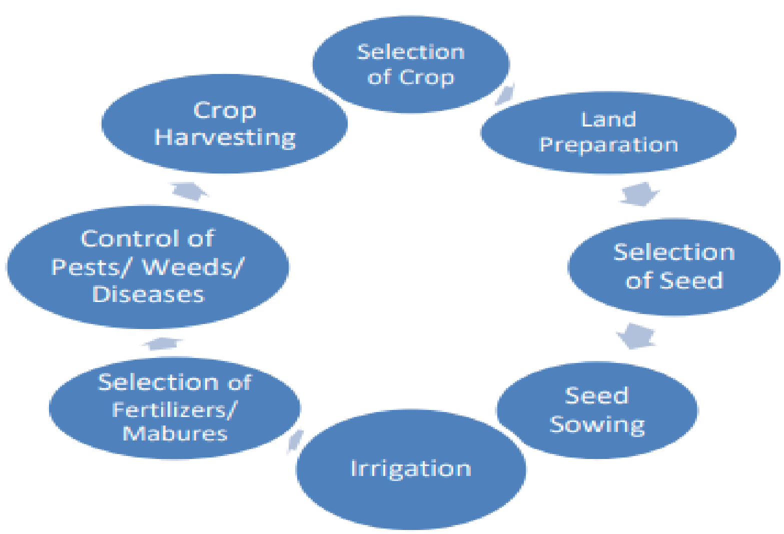
There are many research projects in the field of automated farming. However, Bangladesh is novel in this area of research, with data sets not well structured and the exact result difficult to predict. The data is based on Vinciya and A. Valarmathi's analyses of the organic, inorganic, and immovable data classification to prepare a prediction. In this research, data mining technology was used to extract helpful information and predict that multiple linear regression was employed in the selected region[4]. The Decision Tree algorithm was used to predict that the learning algorithm is controlled by a generalized prediction model, and for multiply linear regression [4]. The Decision Tree algorithm was used to predict the structured object rather than the discrete or real values [4]. Training data from classifications calculated the expected loss[4].

There has been an attempt to investigate the effect on soybean productivity by decision-tab induction techniques of climate parameters[5]. The Decision Tree results were framed in different rules for the end-users to be better understood[5]. Researchers, policymakers, and farmers can use the results of the study to predict and predict crop yields for the dynamics of the market[5].

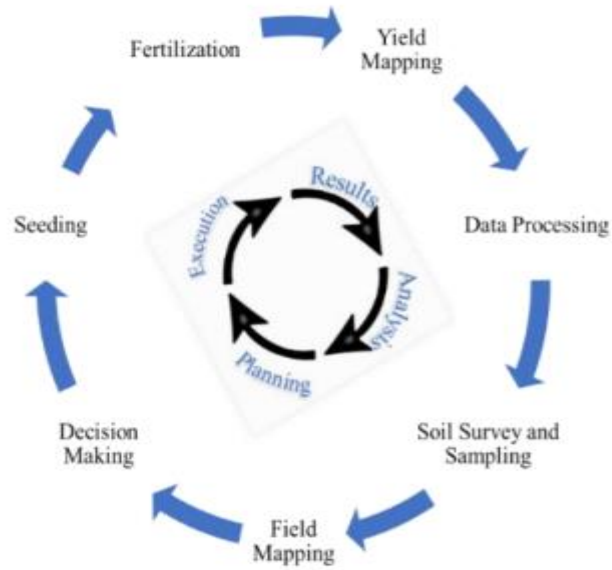
The Decision Tree is one of the most popular machine and data mining algorithms in today's world[6]. The classification of land capacity based on data in the 38 Soil Series (Maharashtra) of the Warda District was evaluated by the Iterating Dichotomizer 3(ID3) – The Classique decision box[6]. As attributes of the terrestrial capacity classification, the depth, pitch, drainage, texture, erosion, and permeability have been chosen[6].

3. MACHINE LEARNING APPLICATIONS IN PRECISION AGRICULTURE

Farmers in many countries rely on the reliability and experience of suggestions made by the elderly based on the traditional method of agriculture. This method makes it possible for farmers to experience random weather conditions caused by global warming and uneven precipitation. The manual sprinkling method has resulted in insufficient use of resources and environmental damage. AI- and IoT-enabled precision agriculture eliminates randomness and helps new-aged farmers to optimize their agriculture in every phase. Figures 2(a) and "Technology Facilitated" present an image of conventional agriculture and farm management systems (b).



(a)



(b)

Figure 2. (a) Traditional agriculture cycle (b) Precision agriculture cycle

Gaitán[7] conducted a systematic study on the effects and impacts on agricultural practices of extreme weather events like hail events, cold waves, heat waves. Inundations, droughts, freezing, hail, heatwaves, and plague outbreaks were noted.

In each agricultural activity as described in Figure 3, the AI systems are applicable and some are even further than those recognized in the convention. This section examines the state-of-the-art techniques that various researchers and practitioners worldwide propose/implement.

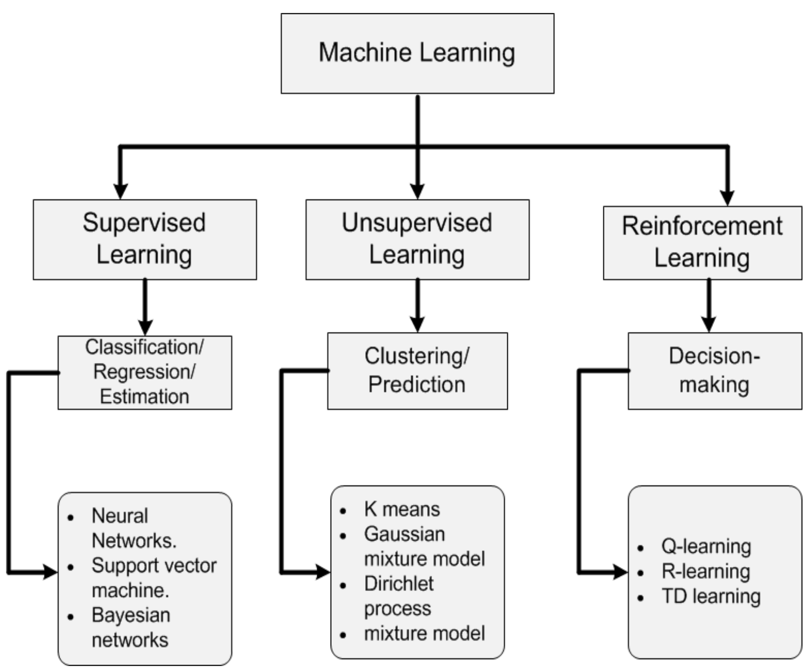


Figure 3. Categorization of Machine learning algorithms.

A. Soil Properties and Weather Prediction

The first and most crucial step in the selection of crops, soil preparation, selection of seeds and crop yield, and selection of fertilizers/manure is the prediction of land properties. The soil characteristics of the land are directly related to the geographical and climate conditions of the land being used and are therefore of great importance. The prediction of the soil properties mainly involves a prediction of soil nutrients, soil humidity, and climate conditions during the crop's lifecycle. Human activities have greatly affected soil properties and therefore our ability to grow crops[15]. In general, plants play a major role in 17 essential elements. Crop growth depends on the available nutrients in a given land. Electrical and electromagnetic sensors are the main control of soil nutrients. According to the nutrients, farmers decide which crop is ideal for the land. However, fertilizers, manure, etc. can add the nutrients but at an extra cost.

B. Crop Yield Prediction

The prediction of crop yield and how to yield may be increased are important pieces of information for any farmer. The value, soil type, and quality of the plant are parameters that play an important role in the forecasting of crop yield: temperature, rainfall, moisture, sunshine time, fertilizers, and harvest times. Manual agriculture can be seen scientifically as a feedback monitoring system in which corrective action is taken when a retrogression in a crop is observed. The yield of crops will depend greatly on the efficiency of optimal use of these resources. If an anomaly in the initial stage is not detected, the crop yield can be damaged uneventfully.

4. METHODOLOGY

Experimental Locations and Plant Materials

The research took place in 2018 and 2019 at Ridgetown Campus University of Guelph. There has been cultivation in two places: Ridgetown (42 to 2701 14.800N 81 to 520 48.000W, 200 m above sea level) and Palmyra (42 to 250 50.100N 81 to 450 06.900W, and 195 m above sea level) in Ontario (Canada), for two consecutive seasons of growth, 2018 and 2019, both under field conditions (Figure 4).

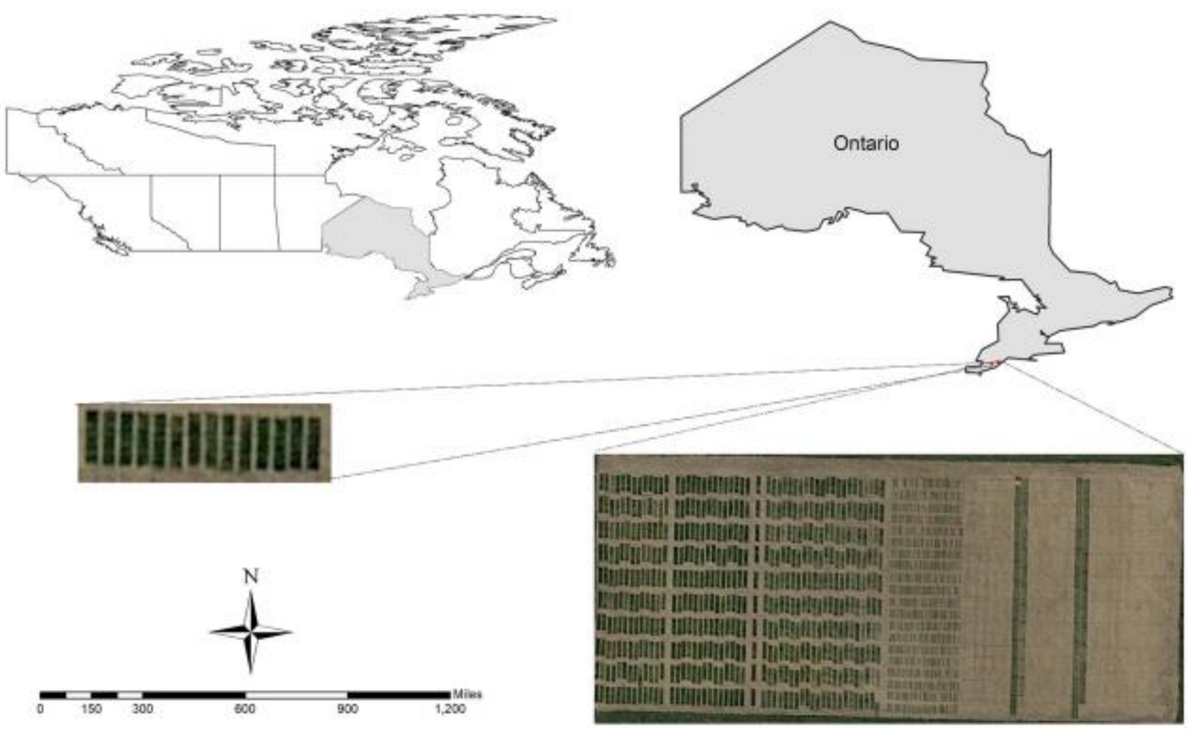


FIGURE 4: The location of the experiments in 2018 and 2019.

The main genotypes of genetic engineering and developmental cultivars used in the study were the soybean genotype derived from soy breeding programs in the University of Guelph, Ridgetown, collected in the past 35 years. The experiments took place with two replications in four environments using randomized complete block designs (RCBD) (two settings / two years). There were 500 environmental soybean plots and a total of 1,000 annual soybean plots.

Phenotypic Evaluations

Yield Collection

The soybean seed yield (Ton ha−1) was measured by three in five rows harvested, adjusted to 13 percent andhumidity. The best unbiased linear prediction (BLUP) as a mixed model was used for calculating the average seed yield in each soya genotype over different environments.

Variable Selection

The feature selection or variable is typically used in small training datasets before machine learning algorithms for data dimension reduction are developed. One of the common approaches for variable selection is a recursive removal approach (RFE). The most important output prediction variables can be defined and selected easily (Chen and Jeong, 2007). Therefore, the RFE was launched to show the original variable meaning and to remove the variables with the lowest significance range. The process was repeated repeatedly before the ranking was indicated for all reflection groups.

A. Supervised Machine Learning

By these algorithms in a set of predictors, the target variable (dependent variable) is predicted (independent variables). To make the function for mapping inputs to preferred outputs use this set of variables. The course continues until the model is accurate in terms of the training data[13]. The training process continues.

B. Decision Tree

Data mining techniques include pattern extraction and classification of large and uncertain data[11]. Algorithms for the decision tree are a common data mining method in which large volumes of data sets and similar pattern extracts are classified[11]. Decision tree is a classification model that breaks it down into smaller subsets for a dataset of a tree structure. The rooted node is selected in an entropy computing mechanism to gain information by the attribute information gained from a dataset.

C. KNN Regression:

● Implementation: In this study, the Euclidean differences are measured between average precipitation and the learned temperature values and test data. Then, it determines three data instances closest to the test case based on distances. From such k cases, the value of the output is the average of the rate of output of the instances selected.

● Pseudocode:

// calculating the Euclidean Distance

For j=0 to Yield.

length

D[j]=distance (rainfall, temp, rainfall[i], temp[i])

End For

//Finding the indices of 'k' no. mean Euclidean Distances

FindMinIndex (k, d, minIndex)

//calculating the predicted yield

Yield←← calculateAverage(y,minIndex)

//Calculate percentage if error

Error ←←percentageErrors (y1,yield)

D. KNNR

Euclidean distances are measured and after the KNN regression, the distance level is measured. The adjacent values can be found based on the optimum K number. He and his neighbor calculate the average reverse distance.

5. RESULTS AND DISCUSSION

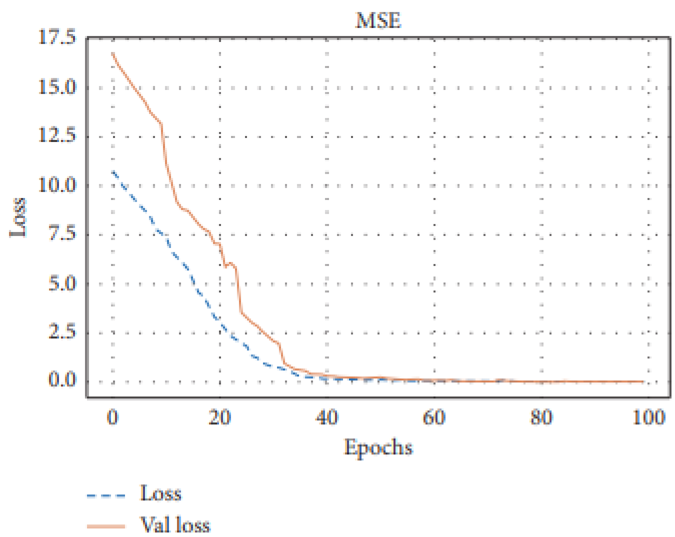


Figure 5: Validation of the model in terms of MSE.

Loss and loss of validation are measured in the course of training and testing to evaluate the model performance. In these two phases, therefore, Figure 5 represents the MSE of the model. The loss in the proposed model is shown to decrease gradually. This also enables the system to be predicted with increased precision. While the model's loss rate at the start was high, each iteration gradually increased the accuracy of the model, as illustrated in Figure 5. The loss will eventually decrease sharply, the system loss after 100 times is about 0.059, and the validation loss approximately 0.063. Further, the analysis of the results also shows that for the 2014 Dhaka District the actual results were 1.988 where the model proposed is near to the current model with a forecast of 2.01.

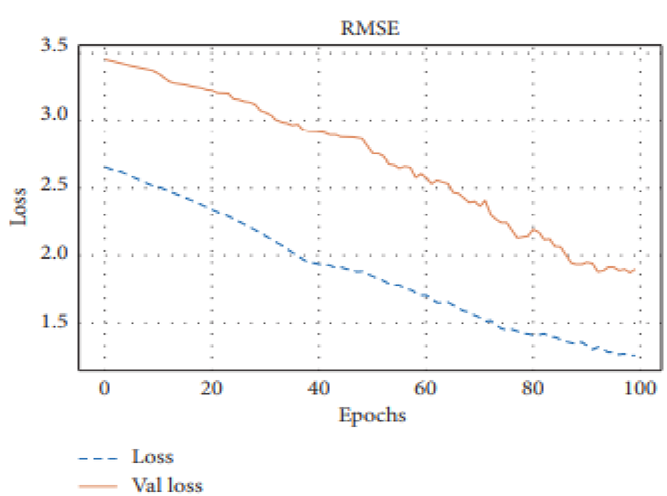


Figure 6: Validation of the model in terms of RMSE.

This is the second loss function for the model's performance assessment. The density of neurons is unchanged in the hidden levels in which the activation function is determined to be equal to the tank while maintaining a validation of 20%. Figure 6 shows, as the MSE, that the model RMSE also gradually decreases. We can keep our loss close to 0,267 and the loss of validation near to 0,42 after 100 epochs. The measured result, however, indicated slightly that the overfit was possible. The 2014 Dhaka District forecast results show that the model forecast a value of close to 1,47 with a true 1,988 value.

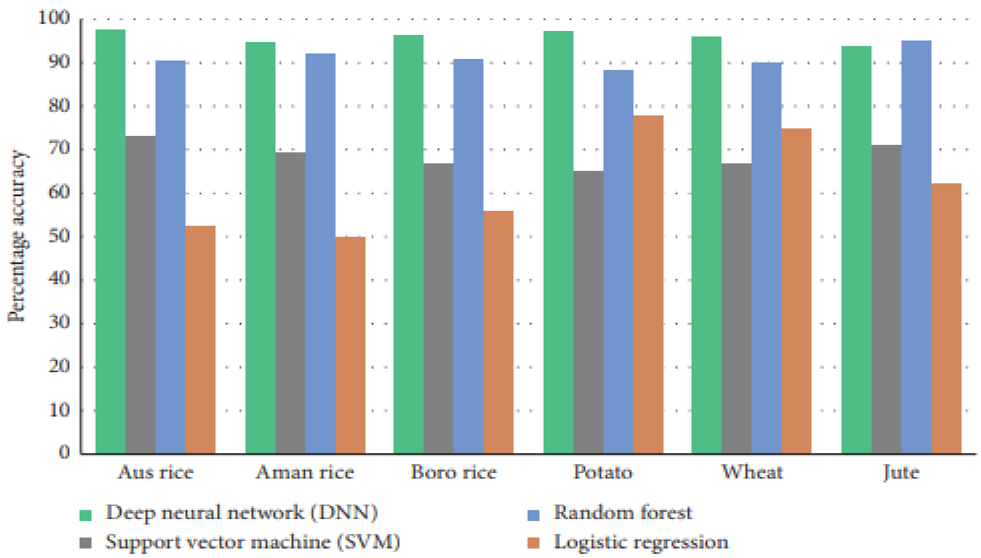


Figure 7: Accuracy percentage of various crops in different algorithms.

In addition, as illustrated in Figure 7, both potatoes and wheat are better precise in the deep neural network model. For potatoes, 97.3% of the neural deep-network model was precise. However, we found 65 percent precision with the same data set in the supporting vector machine, 88,3 percent random forest accuracy, and 78 percent regression-based accuracy in logistic models. The neural network model, therefore, exceeds the other three for potato selection and, given the accuracy results, the yield forecast.

CONCLUSION

An extensive study on the use of machine learning tools in agriculture was carried out in this paper. Agriculture, which employs most workers in countries such as Bangladesh, still employs traditional, human-based plant production methods. This research tries to integrate the prediction models into Bangladesh's old-fashioned farming system. This paper focuses on an additional attempt to collect large amounts of data from different agricultural government agencies in Bangladesh. The goal is to predict crop production using various machine learn algorithms based on the data collected. Neural networks have demonstrated that they have surpassed other machine learning algorithms and justify machine learning requirements in agriculture.

REFERENCES

[1] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," >e Journal of Agricultural Science, vol. 145, no. 3, p. 249, 2007.

[2] K. A. Shastry and H. A Sanjay, Cloud-based Agricultural Framework for Soil Classification and Crop Yield Prediction as a Service in Emerging Research in Computing, Information, Communication, and Applications, Springer, Singapore, Singapore, 2019.

[3] S. Jabjone and C. Jiamrum, "Artificial neural networks for predicting the rice yield in Phimai district of

*Thailand," International Journal of Electrical Energy, vol. 1, no. 3, pp. 177–181, 2013

[4] P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining," International Journal of Advanced Research in Computer Science and Software Engineering(ijarcsse), vol. 6, issue. 5, pp. 481-488, May.2016.

[5] S. Veenadhari, D. Bharat Mishra, and D. C. Singh, "Soybean Productivity Modelling Using Decision Tree Algorithms," International Journal of Computer Applications, vol. 27, no. 7, pp. 11–15, Aug. 2011.

[6] N. KUMAR, G. P. O. REDDY, S. CHATTERJI, and D. SARKAR, "An application of ID3 Decision Tree Algorithm in land capability classification," Nagpur 440033, India, 2012, pp. 35–42.

[7]. C. F. Gaitán, "Machine learning applications for agricultural impacts under extreme events," in Climate Extremes and Their Implications for Impact and Risk Assessment, Elsevier, 2020, pp. 119-138.

[8] H. Khushi, M. Moniruzzaman, and N. Tabassum, "Growth scenarios oF area, production and yield of rice in Bangladesh," Bangladesh Journal of Agricultural Economics, vol. 41, no. 2, 2020.

[9] X. Zhou, H. B. Zheng, X. Q. Xu et al., "Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 130, pp. 246–255, 2017.

[10] M. S. G. Adnan, A. Y. M. Abdullah, A. Dewan, and J. W. Hall, "*e effects of changing land use and flood hazard on poverty in coastal Bangladesh," Land Use Policy, vol. 99, Article ID 104868, 2020.

[11] M. Somvanshi and P. Chavan, "A review of machine learning techniques using decision tree and support vector machine," 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2016, pp. 1-7. doi: 10.1109/ICCUBEA.2016.7860040

[12] K-Nearest Neighbors Algorithm [Online]. Available at: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#kNN_regression Accessed: 27 Nov. 2016.

[13] P. Tan, M. Steinbach, and V. Kumar, Introduction to data mining, 1st ed. Dorling Kindersley: Pearson, 2015, p. 149.

[14] Yearbook of Agricultural Statistics - (2008-2015) [Online] http:bbs.gov.bd Accessed: 27 Nov. 2016.

[15] F. Baheri, F. Davardoost, and V. Ahmadzadeh, "Data mining with learning decision tree and Bayesian network for data replication in Data Grid", Proceedings of the Third International Conference on Contemporary Issues in Computer and Information Sciences, pp. 51- 51, 2012