UMYU Scientifica, Vol. 2 NO. 4, December 2023, Pp 031 – 038
https://doi.org/10.56919/usci.2324.004
ISSN: 2955 – 1145 (print); 2955 – 1153 (online)
A periodical of the Faculty of Natural and Applied Sciences, UMYU, Katsina

ORIGINAL RESEARCH ARTICLE

Crop Yield Prediction in Nigeria Using Machine Learning Techniques: (A Case Study of Southern Part of Nigeria).

Ahmed Abdulbasit[1], Sunday Eric Adewumi[2], Yemi-Peters Victoria[2].
[1]Division of Agricultural Colleges, Ahmadu Bello University, Zaria, Kaduna State, Nigeria.
[2]Department of Computer Science, Federal University Lokoja, Kogi State, Nigeria.

ARTICLE HISTORY
Received July 19, 2023.
Accepted November 02, 2023.
Published December 16, 2023.

KEYWORDS
Decision tree classifier, Machine Learning, random forest, root mean square error, and support vector machine.

ABSTRACT
A key tool for digitalizing the agriculture sector and other industries is using big data and machine learning to predict farm produce. The inability of farmers to accurately predict yield is a great problem using previous farming experience. This study adopted three (3) machine learning approaches, including a decision tree classifier, random forest, and support vector machine, to model data from different zones and make predictions. The techniques adopted were tested using root mean square error to ensure the right prediction algorithm is adopted and the right values are obtained. Results show prediction from the South East is the best in terms of yields and accuracy when tested and evaluated, with 138.9 %.

INTRODUCTION

The need for agricultural products rises daily as the economy expands. To increase the quality and output of agricultural products, the agriculture industry must not only employ traditional methods but also develop and support ongoing agricultural output improvement. Big data analytics and machine learning techniques can be applied to improve crop yield by tracking timely crop production and enhancing product output. In order to assist farmers in getting the best crop production, timely guidance for estimating agricultural productivity must be given. Agriculture faces a lot of difficulty in predicting yield because farmers prefer to predict using previous years' farming proceeds, which is not always correct and efficient on several occasions. Using machine learning techniques to predict will go a long way in having the correct yields after following the right agricultural practices.

Crop yield has been affected by so many factors. Such factors are climatic changes, farming techniques and the land upon which farming will occur. Farming, like any other profession, has gone beyond what farmers can just lay his/her hands on and make predictions about farm yield.

Deepak et al. (2020) the farming culture happens based on environmental changes and cultivation methods. Based on historical data, data analysis from crop production and farms recommends farmers crops that need to be cultivated based on former data. The absence of this led to time and financial loss to the farmers.

Aakunuri and Narasimha (2016) prediction is important for taking various policies, but the methods are costly, subjective, and time-consuming, and it doesn't react to differences that occur for seasonal effects and cycles.

As a result of Mayank et al., (2020), numerous algorithms or methodologies and agricultural production can be predicted using these techniques. The use of algorithms effectively, the expansion of the applications, and the function of big data analytics approaches in agriculture. No methods or solutions exist to resolve the situation by analyzing issues and challenges like temperature, weather, and other factors.

Numerous elements affect crop output. Climate change, farming practices, and the land itself are a few examples of such variables. The practice of farming, like any other profession, has developed beyond the point where a farmer can put his or her hands together and forecast the produce of a farm. Agriculture is based on environmental changes and production practices. Farmers are advised to produce particular crops based on historical data analysis from agricultural production and farming. Due to this, farmers lost money and time (Deepak et al., 2020).

Emphasized using predictions for determining the best course of action, but the methods are expensive, arbitrary,

Correspondence: Ahmed Abdulbasit. Division of Agricultural Colleges, Ahmadu Bello University, Zaria, Kaduna State, Nigeria. ✉ abdulbasitahmedcs@gmail.com. Phone Number: +234 706 650 8600.
How to cite: Ahmed, A., Adewumi, S. E., & Yemi-peters, V. (2023). Crop Yield Prediction in Nigeria Using Machine Learning Techniques: (A Case Study of Southern Part of Nigeria). *UMYU Scientifica*, 2(4), 31 – 38. https://doi.org/10.56919/usci.2324.004

and time-consuming and don't consider variations caused by seasonal influences and cycles (Aakunuri and Narsimha, 2016).

The production of crops and their phenology are influenced by soil characteristics, climate, and weather, which has presented a number of obstacles to farming (Kalaiarasi and Anbarasi, 2021).

While attempting to automate farming and other agricultural processes, a number of problems were fabricated. Implementing machine learning techniques can fix some of these problems by collecting data from some farms, evaluating the methods' accuracy, and hosting the application.

Wu et al. (2018) discovered embedded intelligence (EI), a brand-new discipline of study. This field aims to shed light on individual social patterns, spatial setting behaviours, and urban dynamics by analyzing the digital footprints people leave behind when interacting with the Internet of Things (smart cards, smart automobiles, and cameras). Big data knowledge extraction and technology mining were combined to provide high-depth ontologies or metadata for proactive decision-making and reasoning utilizing open data. It also examines the general structure, historical developments, key applications, traits, and research difficulties of EI.

Saeed & Lizhi (2019) built and demonstrated a model with high prediction accuracy when validated with a root-mean-square-error (RMSE) of 12% of the average yield and 50% of the standard deviation for the validation dataset utilizing anticipated meteorological data. Computing results showed that the model outperformed competing approaches like Lasso, regression trees (RT), and shallow neural networks (SNN). The outcomes also showed that environmental variables had a bigger influence on crop productivity than genetic variables.

Manoj et al. (2020) machine learning techniques were employed, and a statistical model were created to provide precise and accurate decision-making that helps farmers choose the most suitable crops to plant based on area and season with a low risk of losses (Manoj et al.,2020)

Priya et al. (2018) proved that agricultural output depends on historical information, including weather, soil, and prior crop output. His work focused on using random forests to forecast the yield using the provided information. Before sowing seeds on agricultural land, the models were tested using several samples and built utilizing real data.

According to research and conclusions of Thomas et al. (2021), the growth of machine learning and sensing technologies leads to a more affordable solution to the issue facing the agricultural sector. A thorough literature review was also done to find and gather the traits used as predictors in studies. Search criteria discovered 567 related publications across six online sites. A selection of 50 publications was made, and we carefully studied each one, assessed the features and approaches used, and made suggestions for further research. These models' most often utilized features were soil type, rainfall, and temperature, whereas artificial neural networks were the most frequently employed method. In order to find studies utilizing deep learning, we examined various electronic resources after making this statement based on a study of 50 papers employing machine learning.

Shivam et al. (2020) created a model that uses data from previous years' temperatures and rainfall to predict crop yield in advance. Three techniques—auto-regressive moving average (ARMA), ARMA with exogenous variables (ARMAX), and seasonal auto-regressive integrated moving average (SARIMA)—were utilized to make the forecast. Based on a comparison of the models' performances, the best one was chosen to forecast temperature and rainfall, which were used to forecast crop production (Shivam et al., 2020)

Kodimalar and Chellammal (2019) discussed how to use Naive Bayes, logistics, and linear regression to select the optimal algorithm from a list of alternatives and the challenges the researchers would face. A model combining big data computing technologies and machine learning was developed after research into the effectiveness of various machine learning techniques and their application.

Mohsen and Guiping (2020) presented a survey summarizing different approaches for predicting corn yield using machine learning, summarizing the developed models, and providing a numerical solution by expressing typical data preparation tasks carried out in the literature. This survey reviewed and presented several studies of machine learning techniques and their variation to predict the yield of corn.

Ramesh et al. (2019) used to forecast soil, weather, and sugarcane using the support vector regression (SVR) algorithm, which used soil and weather factors as input. When modelling meteorological variables and predicting soil, the methods were validated using historical data, which showed that the system had an accuracy of 85.24%.

Sivanandhini and Prakash (2010) employed a neural network model to anticipate the crop yield based on the proper agricultural parameters such as wind speed, temperature, humidity, and pressure. Predictions were made using feed-forward and recurrent neural network models. Finally, root mean square error (RMSE) was utilized to assess the model performance.

Karthikeya et al. (2020) addressed issues caused by population growth and proposed a method to enhance and boost productivity in line with the populace's needs. Using the KNN technique, the issue was resolved. The accuracy of the prediction was obtained with the aid of the KNN algorithm, which considered factors including rainfall, soil type and humidity.

Surya and Laurence (2018) gathered data for crop yield prediction using regression methods. Regression analysis effectiveness in predicting or projecting agricultural productivity was looked at. For sugarcane, banana, maize, turmeric, cotton and coconut. A predictor model was employed to estimate agricultural products in tons.

Rohit et al. (2017) created an accurate and effective precision farming model with support vector machines and Bayesian models. Additionally, it provided a system for choosing crops that would increase yield rates and the kind of crop to choose based on the soil, water, weather, and crop type.

Shreya et al. (2016) used classic clustering methods to improve the performance of rough k-Means and means ++ that were used along with a designed initial cluster center and a random selection of the number of clusters. As a consequence of the initial cluster centric selection, the modified K-Means method was used to increase precision and produce high-quality clusters.

Rushika et al. (2018) produced a number of comparisons between conventional techniques and the decision tree, KNN, and Naive Bayes models that could analyze the soil and forecast crop yield. Furthermore, soil fertility and plant nutrition were examined. The exact nature of the finding allowed for the enhancement of soil characteristics. Additionally, it provides some instructions to help farmers evaluate the quality of their soil using built-and-developed analyses. The technique evaluates soil quality to predict the crop that will grow well in a given soil type and maximize crop yield by suggesting the appropriate fertilizer.

Guoyong and Jim's (2020) imitation of maize identifies the benefits and drawbacks of each approach utilizing classical models, process-based models, and machine learning algorithms.

Abhinav et al. (2021) provided analyses of machine learning applications for "smart farming" and how they help the research community understand the adoption and use of digital practices in agriculture. To evaluate and track agricultural yield and quality, it also systematically assessed computer vision applications. This included looking at how different sets of crops were classified. The method was applied to cattle production by merging data from collar sensors to diagnose eating disorders, reproductive patterns, and cow behaviour.

Mamunur et al. (2021) presented the use of machine learning to predict crop yield with a focus on palm oil using a critical evaluation of the most up-to-date machine learning prediction algorithm application of machine learning in comparative analysis and the palm oil industry as well as an exploration of remote sensing application, disease recognition, and plant growth.

Vaishali et al. (2020) predicted the production of mustard crops, and a model using five approaches was created. Five variables were examined to evaluate the approach's success: precision, recall, accuracy, specificity, and f-score. Several trials were carried out to find the most accurate way to anticipate the output of the mustard crop. The most accurate methods for estimating the yield of mustard crops were found to be KNN and ANN.

Aditya et al. (2017) developed a model to calculate various maize, wheat, and cotton crop yields using regression techniques and other mathematical functions, including pure quadratic, quadratic, polynomial, and interactions. Finally, regression models that offered precise yield predictions for cotton, wheat, and maize were proposed. The optimum regression model was determined using root mean squared error (RMSE), mean percentage prediction error (MPPE), and R2 values based on crop factors and soil weather.

Sangeeta and Shruthi (2020) crop yields were calculated by using datasets linked to crop production, such as moisture, rainfall, and temperature from prior years to forecast. Regression models were used during training and construction, pre-processed and fitted into the trained data to reduce the cost function by determining the optimum fit-line. The difference between the actual and expected values was reduced to the least amount possible.

Anitha (2017) adopted data mining tools for cultivation plans for a high crop. The data set used consisted of actual information provided by farmers who cultivate alongside the river basin. K-means and other decision tree classifiers were used to cluster agronomic data. The performance of several classifiers was compared and validated, and it was established through review and experimentation that the random forest performs better than other classification algorithms. Farmers can utilize the research's final rules to make informed choices before harvest.

Saranya et al. (2020) compared and developed different machine learning techniques; datasets from prior crop years were examined using decision trees, support vector machines, multiple linear regression, and artificial neural networks. Different algorithms' advantages and disadvantages were listed.

Manjula and Djodiltachoumy (2017) constructed a technique to predict agricultural yield using past data. This was done by applying association rule mining on the dataset. The experimental results showed that the proposed technique successfully forecasts crop yield.

Dilli et al. (2020) addressed the requirement for repeatable workflows that will aid in comprehending the significance of various data sources, characteristics, and predictors. Additionally, it developed elements that allow numerous nations to be served with little configuration change by integrating crop simulation results with reusable early or late-season meteorological, soil, and remote-sensing data from databases. The results were compared with a simple approach, and a linear yield trend was predicted. The improvement of the baseline gained by the inclusion of new data, which led to the use of machine learning in large-scale farming, served as the impetus for this (Dilli et al., 2020)

Kuldeep et al. (2020) presented a literature review of several empirical models, ANNs, statistical approaches, and machine learning regression tools employed with satellite data.

The significance of the prediction models are: (1) Better prediction accuracy (2) Fast in terms of processing (3) Ability to choose the best among the models. (4)The local farmers can use it to predict crop yield

The objectives of the study are (1) Collect data from three zones of the country **(2)** Evaluate the accuracy of each of the techniques (3) Choosing the best in terms of accuracy.

MATERIAL AND METHOD

Data Collection

The dataset for the prediction was obtained from three (3) different colleges of agriculture in the southern part of the country, as shown in the Table 1 below.

Table 1: Dataset collected

S/no	Name of College (Farm)	State	Zone
1	Federal College of Agriculture, Akure, Ondo State	Ondo State	South West
2	Federal College of Agriculture, Ishiagu Ebonyi State	Ebonyi State	South East
3	Edo State College of Agriculture, Iguoriakhi, Edo State	Edo State	South-South

Description of the Proposed Model

We developed an effective recommendation system for the system by looking at the preceding data. We used machine learning techniques to process enormous volumes of data in order to foresee and interpret the results. The implementation used decision tree models, random forests, and support vector machines to improve the classifier. The input dataset was compiled from various agricultural colleges around the country. The historical dataset provided the input necessary for data processing, which was analyzed using root mean square error. The best was picked for prediction to provide farmers with relevant guidance for selecting the best crop to be cultivated for maximum yields.

RESULTS

Values for rainy and dry season farming showing the efficiency and accuracy for each zone as shown in the Table 2 below

Using Root means square error shows decision tree classifier performs better during the rainy season with an accuracy of 183.7 for South West, as shown in Figure 1 below.

Table 2: Accuracy from the dataset gotten

S/no	Zone/Season	Support Vector Machine	Random Forest	Decision Tree Classifier
1	South West/Rainy	923.2	353.2	183.7
2	South West/Dry	923.2	353.2	496.2
3	South East/Rainy	699.3	187.2	138.9
4	South East/Dry	844.8	187.4	147.6
5	South-South/Rainy	2400.8	623.9	620.0
6	South-South/Dry	2086.4	704.1	644.2

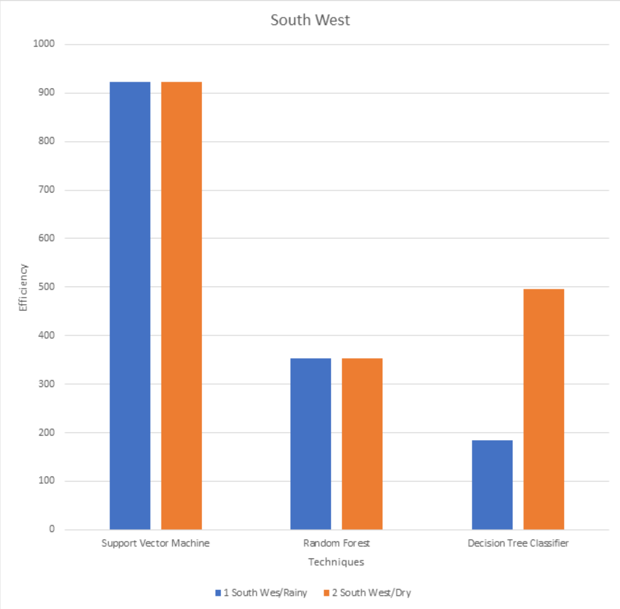

Figure 1: Showing accuracy for South West

Using Root means square error shows decision tree classifier performs better during the dry season with an accuracy of 138.9 for South East as shown in Figure 2 below.

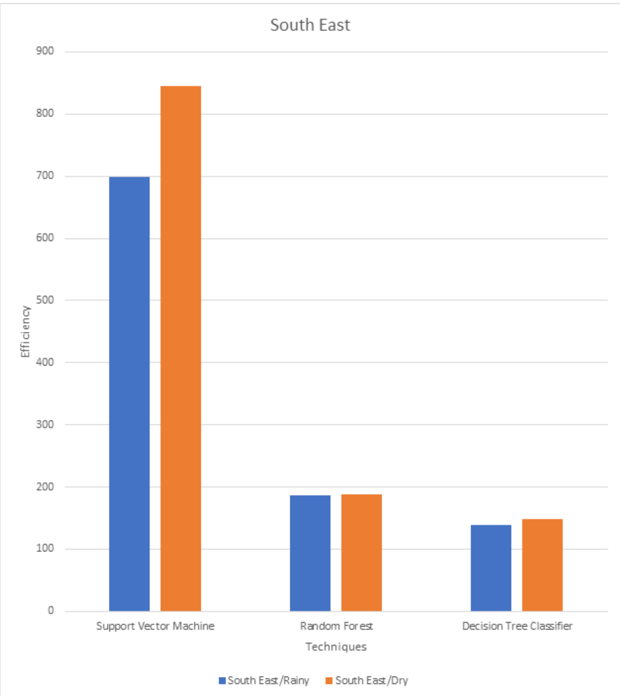

Figure 2: Showing accuracy for South East

Using Root means square error shows decision tree classifier performs better during the rainy season with an accuracy of 620 for South-South, as shown in Figure 3 below.

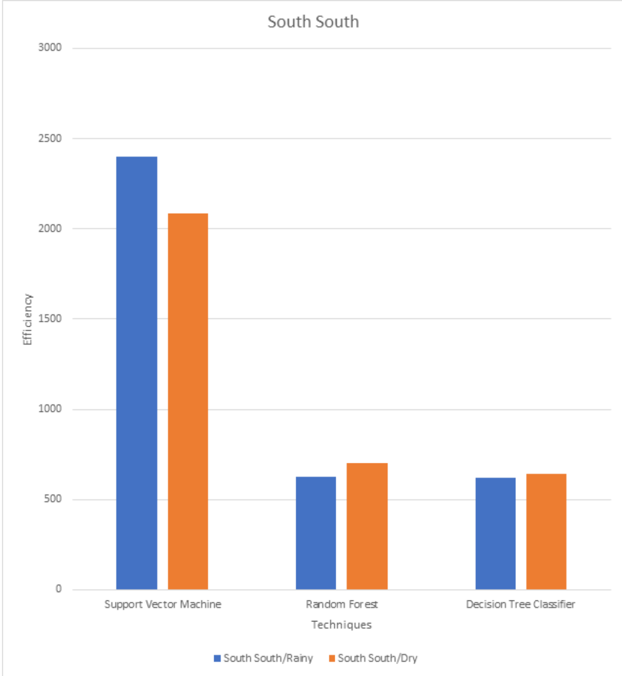

Figure 3: Showing accuracy for South South

Using Root means square error shows decision tree classifier performs better during the rainy season with an accuracy of 138.9, as shown in Figure 4 below.

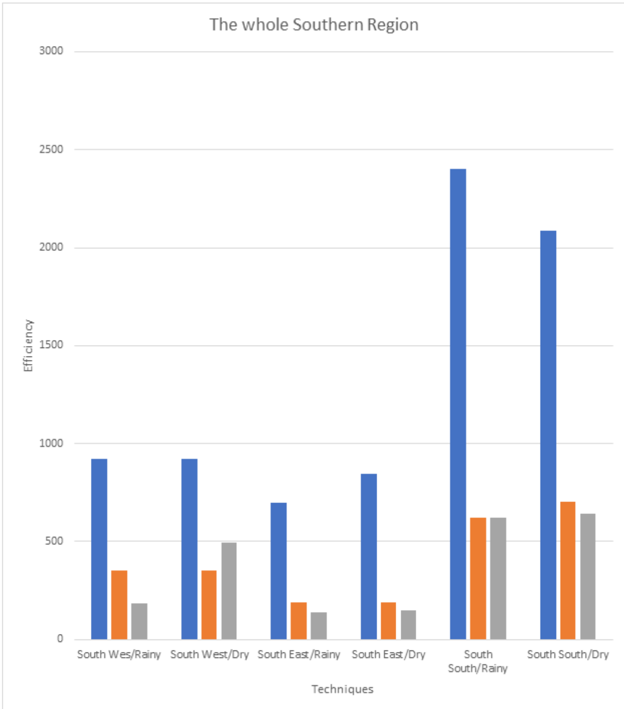

Figure 4: Showing accuracy for the whole southern region

The best model, the random forest, gives the best prediction, while the support vector machine gives the worst prediction, as shown in Figure 5.

```
In [278]: #predicting the crop yield for plant seeds of 30kg, 2.0hectares of land and 20bags
          the_best_model_prediction= model11_rfe.predict([[30, 2.0, 20,1,1,1]])
          print(the_best_model_prediction)

          the_worst_model_prediction= model5_svr.predict([[30, 2.0, 20,1,1,1]])
          print(the_worst_model_prediction)

          [38740.]
          [27503.34567107]
```

Fig. 5: Showing the best and worst prediction

It allows farmers to interact with the software through a web browser that allows farmers to input the number of seeds, size of the land, number of bags, soil type, type of seeds, and colour of seeds, as shown in Figure 6.

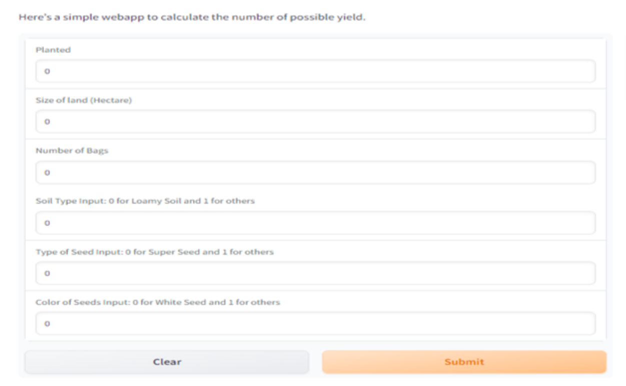

Fig. 6: Web interface of the prediction model

DISCUSSION

Comparison between rainy and dry seasons shows that prediction in the South East using a decision tree classifier has the best in terms of accuracy and output for rainy season farming, while southwest prediction using a decision tree classifier during rainy season has the best in terms of accuracy and output and south-south prediction using decision tree classifier for rainy season farming has the best in terms of accuracy and output. In Conclusion, rainy season farming in the Southeast using a decision tree classifier has the overall best accuracy and output among all the regions.

CONCLUSION

The benefits of the experience gained from working on this paper cannot be overstated. It has successfully bridged the gap between manual prediction techniques and automated techniques leveraging software development. It is open to new developments in the field of machine learning as well as modifications to existing techniques. The development and assessment of a prediction model. The new technology will go a long way toward addressing problems that the traditional method of predicting seasonal crop yield encounters because it indicates that the Southeast produces better during the dry season because irrigation is largely adopted, which lowers the amount of eroded farm products.

RECOMMENDATION

After outlining everything required for this paper to be implemented successfully, the following suggestions are made to improve and address some shortcomings:

1. Ensuring that weather and
2. Soil texture should considered.

REFERENCES

Aakunuri, M., & Narsimha, G. (2016). Crop Yield Prediction with Aid of Optimal Neural Network in Spatial Data Mining: New Approaches. *International Journal of Information & Computation Technology*, 28.

Abhinav, S., Arpit, J., Prateek, G., & Vinay, C. (2021). Machine Learning Applications for Precision. *International for Electrical and Electronics Engineer*, 4845-4847.

Aditya, S., Sanjay, H. A., & Bhanusree, E. (2017). Prediction of Crop Yield Using Regression Techniques. *International Journal for Soft Computing*, 96.

Anitha, A. (2017). A Predictive Modeling Approach for improving Paddy Crop Productivity using Data Mining Techniques. *Turkish Journal of Electrical Engineering & Computer Sciences*, 4777. **[Crossref]**

Deepak, G., Deepika, J., Dharshini, M., & Vanathi, B. (2020). Crop Yield Prediction Based on Ensemble Model Using Historical Data. *International Journal of Advance Research and Innovative Ideas in Education (IJARIIE)*, 670.

Dilli, P., Hendrik, B., Allard, D. W., Sander, J., Sjoukje, O., Christos, P., & Joannis, N. A. (2020, December 4). Machine Learning for Large-scale Crop Yield Forecasting. *Agricultural Systems*, p. 2.

Guoyong, L., & Jim, H. W. (2020). Predicting Spatial and Temporal Variability in Crop Yields: An

Inter-comparison of Machine Learning, Regression and Process-based Models. *Environmental Research* (p. 2). China: Institute of Geographic Sciences and Natural Resources Research.

Kalaiarasi, E., & Anbarasi, A. (2021). Crop yield prediction using multi-parametric deep neural networks. *Indian Journal of Science and Technology*, 131. **[Crossref]**

Karthikeya, H. K., Sudarshan, K., & Disha, S. S. (2020). Prediction of Agricultural Crops using KNN Algorithm. *International Journal of Innovative Science and Research Technology ISSN No:-2456-2165*, 1422.

Kodimalar, P., & Chellammal, S. (2019). An Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques. *International Journal of Computer Engineering and Technology (IJCET) Volume 10, Issue 03*, 110. **[Crossref]**

Kuldeep, S., Sunila, & Sanjeev, K. (2020). Crop Yield Prediction Techniques using Remote Sensing Data. *International Journal of Engineering and Advanced Technology (IJEAT)*, 3683. **[Crossref]**

Mamunur, R., Bifta, S. B., Yusri, Y., Mohamad, A. K., & Nuzhat, K. (2021). A Comprehensive Review of Crop Yield Prediction using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. *Institute of Electrical and Electronics Engineers*, 63409.

Manoj, G. S., Prajwal, G. S., Ashoka, U. R., Prashant, K., & Anitha, P. (2020). Prediction and Analysis of Crop Yield using Machine Learning Techniques. *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181*, 19.

Manjula, E., & Djodiltachoumy, S. (2017). A Model for Prediction of Crop Yield. *International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4*, 298-300.

Mayank, C., Darpan, C., Chaitanya, C., & Mansing, R. (2020). Crop Yield Prediction Using Machine Learning. *International Journal of Science and Research (IJSR)*, 645.

Mohsen, S., & Guiping, H. (2020). Machine Learning Models for Corn Yield Prediction: A Survey of Literature. *International Journal of Environmental Sciences & Natural Resources*, 80.

Priya, P., Muthaiah, U., & Balamurugan, M. (2018). Predicting yield of the Crop using Machine Learning Algorithm. *International Journal of engineering Sciences & Research Technology*, 1.

Ramesh, M. A., Vijay, R. S., & Anand, A. M. (2019). Sugarcane Crop Yield Forecasting Model Using Supervised Machine Learning. *International Journal for Intelligent Systems and Applications*, 13.

Rohit, K. R., Ankit, P., Mitalee, P., Pooja, S., Suresh, R., & Avinash, D. (2017). Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique. *International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 12*, 950.

Rushika, G., Juilee, K., Pooja, M., Sachee, N., & Priya , R. L. (2018). Prediction of Crop Yield using Machine Learning . *International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056*, 2237.

Saeed, K., & Lizhi, W. (2019). *Crop Yield Predictions using Deep Neural Network*. United States of America: Frontiers in Plant Science.

Sangeeta, & Shruthi, G. (2020). Design and Implementation of Crop Yield Prediction Model in Agriculture. *International Journal of Scientific & Technology Research* , 544.

Saranya, C. P., Guru, M. B., Karuppasamy, M., Sunmathi, M., & Shree, S. K. (2020). A Survey on Crop Yield Prediction using Machine Learning Algorithms. *International Journal of Research and Analytical Reviews (IJRAR)*, 494.

Shivam, B., Rajat, B., Ankit, S. C., Akshay, K. D., & Indu, C. (2020). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall Parameters Predicted through ARMA, SARIMA, and ARMAX models. *Research Gate Conference Paper* (p. 978). India: Department of CSE & IT.

Shreya, S. B., Kalyani, A. B., Aarti, G. D., & Bhagyashree, R. G. (2016). Crop and Yield Prediction Model. *International Journal of Advance Scientific Research and engineering trends*, 23.

Sivanandhini, P., & Prakash, J. (2020). Comparative Analysis of Machine Learning Techniques for Crop Yield Prediction. *International Journal of Advanced Research in Computer and Communication Engineering*, 289.

Sivanandhini, P., & Prakash, J. (2020). Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network. *International Journal of Innovative Science and Research Technology*, 1092.

Surya, P., & Laurence, A. I. (2018). Crop Yield Prediction in Agriculture using Data Mining Predictive Analytic Techniques. *International Journal of Research and Analytical Reviews*, 787.

Thomas, V. K., Ayalew, K., & Cagatay, C. (2021). Crop Yield Prediction using Machine Learning: A systematic literature review. *Computers and Electronics in Agriculture*, 2.

Vaishali, P., Haneet, K., Surjeet, S., Jatinder, M., & Vinod, S. (2020). Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis. *Journal of Scientific Research*, 394. **[Crossref]**

Wu, Y., Li, S., Li, L., Li, M., Li, M., Arvanitis, K. G., . . . Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *International Federation of Automatic Control Conference Paper* (p. 32). China: Elsevier.