Journal of Engineering Sciences

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS

Dr. Pathan Ahmed Khan, Assistant Professor, Department of CSE
pathan.ahmed0504@gmail.com

Meesum S Hussain, BE, Department of CSE,
meesumshussain@gmail.com

Mir Mustafa Ali, BE, Department of CSE,
mirmustafa672@gmail.com

MD Zakir Ali Khan, BE, Department of CSE,
zakirdanish28.zk@gmail.com

ABSTRACT: Agriculture is the pillar of the Indian economy and more than 50% of India's population are dependent on agriculture for their survival. Variations in weather, climate, and other such environmental conditions have become a major risk for the healthy existence of agriculture. Machine learning (ML) plays a significant role as it has decision support tool for Crop Yield Prediction (CYP) including supporting decisions on what crops to grow and what to do during the growing season of the crops. The present research deals with a systematic review that extracts and synthesize the features used for CYP and furthermore, there are a variety of methods that were developed to analyze crop yield prediction using artificial intelligence techniques. The major limitations of the Neural Network are reduction in the relative error and decreased prediction efficiency of Crop Yield. Similarly, supervised learning techniques were incapable to capture the nonlinear bond between input and output variables faced a problem during the selection of fruits grading or sorting. Many studies were recommended for agriculture development and the goal was to create an accurate and efficient model for crop classification such as crop yield estimation based on the weather, crop disease, classification of crops based on the growing phase etc., This paper explores various ML techniques utilized in the field of crop yield estimation and provided a detailed analysis in terms of accuracy using the techniques.

Keywords: *Crop yield prediction, ML.*

1. INTRODUCTION

The history of agriculture in India dates back to the Indus Valley Civilization Era. India ranks second in this sector. Agriculture and allied sectors like forestry and fisheries account for 15.4 percent of the GDP (gross domestic product) with about 31 percent of the workforce. India ranks first globally with the highest net cropped area followed by US and China. Agriculture is demographically the broadest economic sector and plays a significant role in the

overall socio-economic fabric of India. Due to the revolution in industrialization, the economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth.

The problem that the Indian Agriculture sector is facing is the integration of technology to bring the desired outputs. With the advent of new technologies and overuse of non-renewable energy resources patterns of rainfall and temperature are disturbed. The inconsistent trends developed from the side effects of global warming make it cumbersome for the farmers to clearly predict the temperature and rainfall patterns thus affecting their crop yield productivity. In order to perform accurate prediction and handle inconsistent trends in temperature and rainfall various machine learning algorithms like RNN, LSTM, etc can be applied to get a pattern. It will complement the agricultural growth in India and all together augment the ease of living for farmers. In past, many researchers have applied machine learning techniques to enhance agricultural growth of the country.

This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc Objectives to be followed in the future are given below:

1. Depending on the dissimilar crop feature divisions, the modulating factor values of ML algorithms differ to attain perfect approximation.

2. When the quantity of input elements is reduced, ANN is utilized. The optimal feature was being empirically selected for appropriate crop yield estimation.

3. The advantage of ML method regression is to avoid difficulties of using a linear function in large output sample space and optimization of complex problems transformed into simple linear function optimization.

4. ML algorithm can be executed with an enormous soil dataset for crop yield estimation. 5. The ML techniques, through observation of the agricultural fields, provided the necessary support to the farmers in increasing crop production to a great extent.

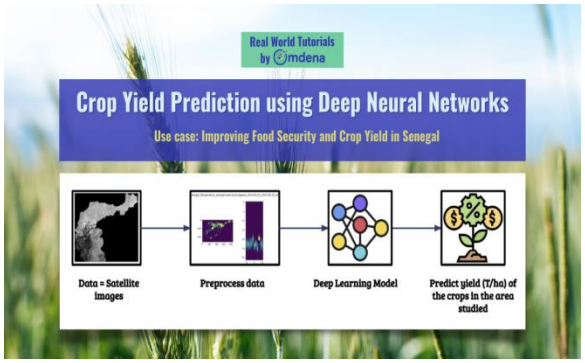

Figure:1. Example figure

2. LITERATURE SURVEY

2.1 PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM:

The agriculture plays a dominant role in the growth of the country's economy.Climate and other environmental changes has become a major threat in the agriculture field. Machine learning (ML) is an essential approach for achieving practical and effective solutions for this problem. Crop Yield Prediction involves predicting yield of the crop from

available historical available data like weather parameter,soil parameter and historic crop yield.This paper focus on predicting the yield of the crop based on the existing data by using Random Forest algorithm. Real data of Tami lnadu were used for building the models and the models were tested with samples.The prediction will helps to the farmer to predict the yield of the crop before cultivating onto the agriculture field. To predict the crop yield in future accurately Random Forest, a most powerful and popular supervised machine learning algorithm is used.

2.2 Applications of machine learning techniques in agricultural crop production: a review:

This paper has been prepared as an effort to reassess the research studies on the relevance of machine learning techniques in the domain of agricultural crop production. Methods/Statistical Analysis: This method is a new approach for production of agricultural crop management. Accurate and timely forecasts of crop production are necessary for important policy decisions like import-export, pricing marketing distribution etc. which are issued by the directorate of economics and statistics. However one has understand that these prior estimates are not the objective estimates as these estimate requires lots of descriptive assessment based on many different qualitative factors. Hence there is a requirement to develop statistically sound objective prediction of crop production. That development in computing and information storage has provided large amount of data. Findings: The problem has been to intricate knowledge from this raw data, this has lead to the development of new approach and techniques such as machine learning that can be used to unite the knowledge of the data with crop yield evaluation. This research has been intended to evaluate these innovative techniques such that significant relationship can be found by their applications to the various variables present in the data base. Application/Improvement: The few techniques like artificial neural networks, Information Fuzzy Network, Decision Tree, Regression Analysis, Bayesian belief network. Time series analysis, Markov chain model, k-means clustering, k nearest neighbor, and support vector machine are applied in the domain of agriculture were presented.

2.3 A Model for Prediction of Crop Yield:

Data Mining is emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Data mining techniques are the better choice for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This research proposes and implements a system to predict crop yield from previous data. This is achieved by applying association rule mining on agriculture data. This research focuses on creation of a prediction model which may be used to future prediction of crop yield. This paper presents a brief analysis of crop yield prediction using data mining technique based on association rules for the selected region i.e. district of Tamil Nadu in India. The experimental results shows that the proposed work efficiently predict the crop yield production.

2.4 Agricultural crop yield prediction using artificial neural network approach:

By considering various situations of climatologically phenomena affecting local weather conditions in various parts of the world. These weather conditions have a direct effect on crop yield. Various researches have been done exploring the connections between large-scale climatologically phenomena and crop yield. Artificial neural networks have been demonstrated to be powerful tools for modeling and prediction, to increase their effectiveness. Crop prediction methodology is used to predict the suitable crop by sensing various parameter of soil and also parameter related to atmosphere. Parameters like type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, humidity. For that purpose we are used artificial neural network (ANN).

2.5 Predictive ability of machine learning methods for massive crop yield prediction:

An important issue for agricultural planning purposes is the accurate yield estimation for the numerous crops involved in the planning. Machine learning (ML) is an essential approach for achieving practical and effective solutions for this problem. Many comparisons of ML methods for yield prediction have been made, seeking for the most accurate technique. Generally, the number of evaluated crops and techniques is too low and does not provide enough information for agricultural planning purposes. This paper compares the predictive accuracy of ML and linear regression techniques for crop yield prediction in ten crop datasets. Multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression and k-nearest neighbor methods were ranked. Four accuracy metrics were used to validate the models: the root mean square error (RMS), root relative square error (RRSE), normalized mean absolute error (MAE), and correlation factor (R). Real data of an irrigation zone of Mexico were used for building the models. Models were tested with samples of two consecutive years. The results show that M5- Prime and k-nearest neighbor techniques obtain the lowest average RMSE errors (5.14 and 4.91), the lowest RRSE errors (79.46% and 79.78%), the lowest average MAE errors (18.12% and 19.42%), and the highest average correlation factors (0.41 and 0.42). Since M5-Prime achieves the largest number of crop yield models with the lowest errors, it is a very suitable tool for massive crop yield prediction in agricultural planning.

3. PROPOSED SYSTEM

Most of the Exiting models utilized Neural networks, random forests, KNN regression techniques for CYP and a variety of ML techniques were also used for best prediction. The problems faced in existing research for crop yield prediction using machine learning are stated below:

1. Creation, repair and maintenance of ML algorithms required huge costs as they are very complex.

2. ML technique used for Crop yield prediction (mustard, wheat) combined input and output data but failed to obtain better results statistically

3. Due to the nature of linear connection in the parameters, the regression model was failed to

provide the exact prediction in a complex situation such as extreme value data and nonlinear data.

4. The existing K-NN models were used for classification for yield prediction but lowered the performance due to nonlinear and highly adaptable issues present in KNN.

They were operated in a locality model that incremented the dimensionality of the input vector made confusion for classification.

5. An appropriate decision was not taken during classification because a fewer quantity of data was available for estimation of crop yield.

Disadvantages:

The existing K-NN models were used for classification for yield prediction but lowered the performance due to nonlinear and highly adaptable issues present in KNN.

This paper focuses on the practical application of machine learning algorithms and its quantification. The work presented here also takes into account the inconsistent data from rainfall and temperature datasets to get a consistent trend. Crop yield prediction is determined by considering all the features in contrast with the usual trend of determining the prediction considering one feature at a time.

Advantages:

1. From the studies most of the common algorithms used were CNN, LSTM, DNN algorithms but still improvement was still required further in CYP.
2. The present research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction. • Ultimately, the experimental study showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction.

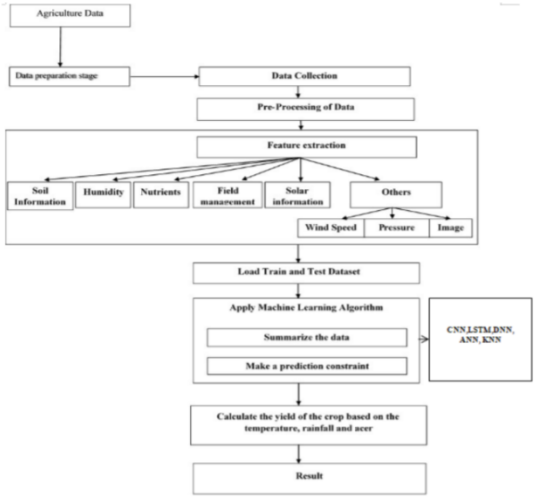

Fig.2: System architecture

MODULES:

Upload Crop Dataset: The crop production dataset that is used to predict the name and yield of the crop is fed into classification and regression algorithms.

Preprocess Dataset: Experiments were conducted on Indian government dataset and it has been established that Random Forest Regressor gives the highest yield prediction accuracy. Sequential model that is Simple Recurrent Neural Network performs better on rainfall prediction while LSTM is good for temperature prediction. By combining rainfall, temperature along with other parameters like season and area, yield prediction for a certain district can be made.

Train Machine Learning: This focuses on district wise yield prediction according to the crop sown in the district. Yield is being predicted for given crops district wise and crops with best yield.

Upload Test Data &Predict Yield: Results reveals that Random Forest is the best classifier when all parameters are combined. This will not only help farmers in choosing the right crop to grow in the next season but also bridge the gap between technology and the agriculture sector.

4. ALGORITHMS

Logistic Regression:- Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. When logistic regression algorithm applied on our dataset it provides an accuracy of 87.8%.

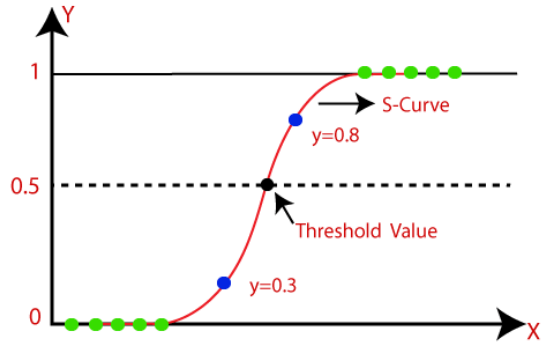

Fig.3: Logistic regression model

Naive Bayes:- Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. It provides an accuracy of 91.50%.

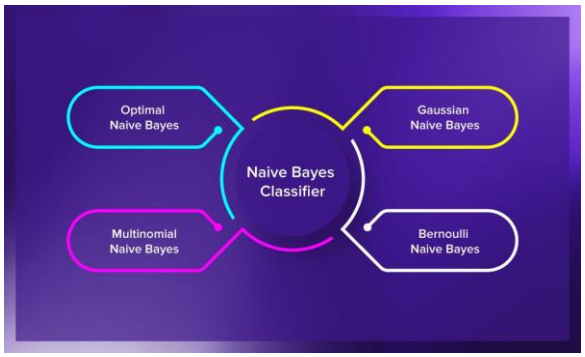

Fig.4: Naïve bayes model

Random Forest:- Random Forest has the ability to analyze crop growth related to the current climatic conditions and biophysical change. Random forest algorithm creates decision trees on different data samples and then predict the data from each subset and then by voting gives better solution for the system. Random Forest uses the bagging method to train the data which increases the accuracy of the result. For our data, RF provides an accuracy of 92.81%.

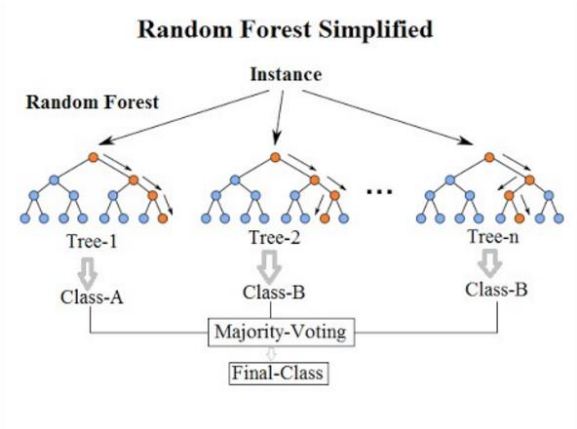

Fig.5: Random forest model

5. EXPERIMENTAL RESULTS

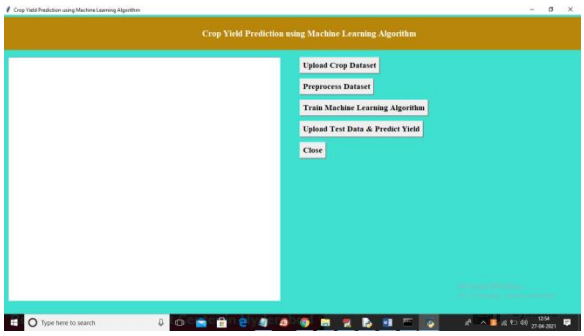

Fig.6: Home screen

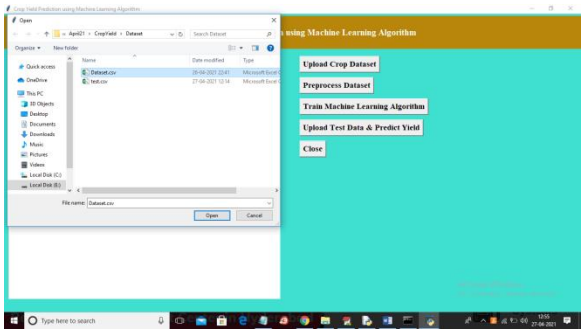

Fig.7: Upload heavy vehicle fuel dataset

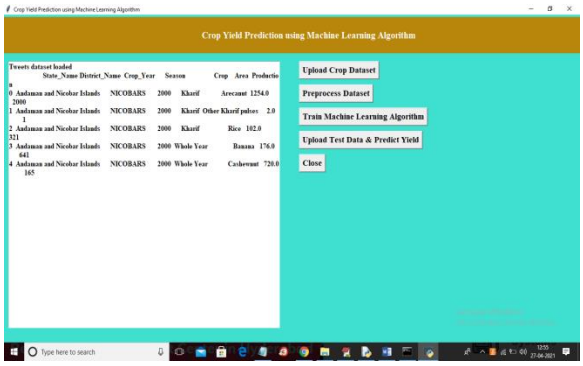

Fig.8: Dataset loaded

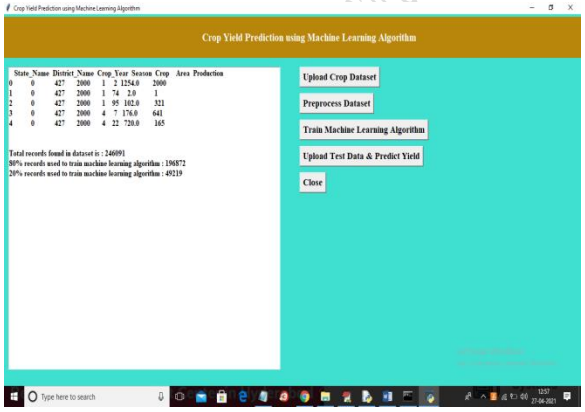

Fig.9: Preprocess dataset

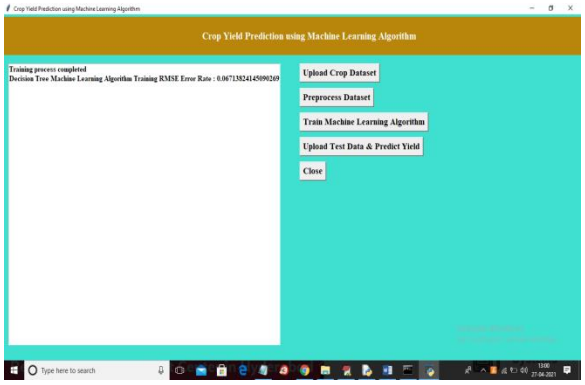

Fig.10: Train ML algorithm

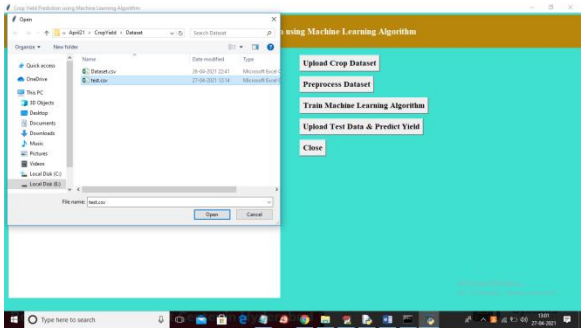

Fig.11: Test data upload screen

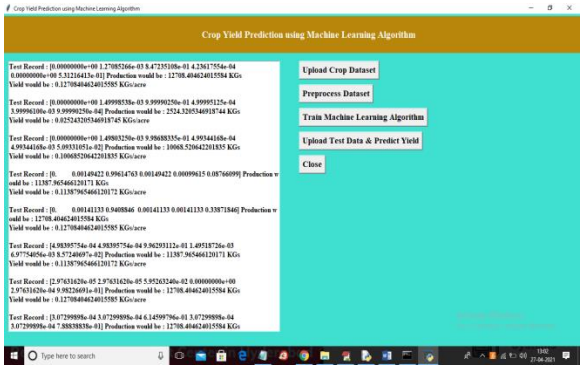

Fig.12: Upload test & predict yield

6. CONCLUSION

The present research work discussed about the variety of features that are mainly dependent on the data availability and each of the research will investigated CYP using ML algorithms that differed from the features. The features were chosen based upon the geological position, scale, and crop features and these choices were mainly dependent upon the data-set availability, but the more features usage was not always giving better results. Therefore, finding the fewer best performing features were tested that also have been utilized for the studies. Most of the exiting models utilized Neural networks, random forests, KNN regression techniques for CYP and a variety of ML techniques were also used for best prediction. From the studies most of the common algorithms used were CNN, LSTM, DNN algorithms but still improvement was still required further in CYP. The present research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction. Ultimately, the experimental study showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction. However, still more improvement in feature selection was required in terms of temperature variation aspects effects on agriculture. In the further studies, the key possibility that should be concentrated such as firstly the delay to border topographical areas required additional-explicit treatment. Next, a nonparametric portion of the model using machine learning algorithm and thirdly, using features from deterministic crop models to get perfect statistical CO2 fertilization. By following above-mentioned objectives, the crop yield estimation would be improved by further researchers. Additionally, in the crop yield estimation, fertilizer should also be considered for executing soil forecasts that agriculturalist to make a better judgment based on the situation of low crop yield estimation. Based on the outcomes obtained for the study further we need to build and develop a model based on DL for CYP.

REFERENCES

[1] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, “Prediction of crop yield using machine learning,” Int. Res. J. Eng. Technolgy, vol. 5, 2018.

[2] F. H. Tseng, H. H.Cho, and H. T. Wu, “Applying big data for intelligent agriculture-based crop selection analysis,” IEEE Access, vol. 7, pp. 116965-116974, 2019.

[3] A. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, “Crop Yield Prediction Using Linear Support Vector Machine,” European Journal of Molecular & Clinical Medicine, vol. 7, no. 6, pp. 2189- 2195, 2020.

[4] M. Alagurajan, and C. Vijayakumaran, “ML Methods for Crop Yield Prediction and Estimation: An Exploration,” International Journal of Engineering and Advanced Technology, vol. 9 no. 3, 2020

[5] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Predicition of Crop Yeild Using SVM Approch with the Facility of E-MART System" Easychair 2020.

[6] S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, pp.103090, 2020

[7] P. Tiwari, and P. Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[8] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.

[9] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," Ictact journal on soft computing, vol. 11, no. 01, 2020

[10] A. A. Alif, I. F. Shukanya, and T. N. Afee, "Crop prediction based on geographical and climatic data using machine learning and deep learning", Doctoral dissertation, BRAC University) 2018.

[11] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deeplearning-based detector for real-time tomato plant diseases and pests' recognition," Sensors, vol. 17, no. 9, pp. 2022, 2017.

[12] J. Sun, L, Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," Sensors, vol. 19, no. 20, pp. 4363, 2019.

[13] K. A. Shastry, and H. A. Sanjay, "Hybrid prediction strategy to predict agricultural information," Applied Soft Computing, vol. 98, pp. 106811, 2021.

[14] D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371-376, 2019.

[15] B. Devika, and B. Ananthi, "Analysis of crop yield prediction using data mining technique to predict annual yield of major crops," International Research Journal of Engineering and Technology, vol. 5, no.12, pp. 1460-1465, 2018.