International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 6, December 2023, pp. 6806~6816
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i6.pp6806-6816

Selection of crop varieties and yield prediction based on phenotype applying deep learning

Iniyan Shanmugam, Jebakumar Rethnaraj, Gayathri Mani
Department of Computing Technologies, SRM Institute of Science and Technology, Kaatankulathur, India

Article Info

Article history:

Received Sep 2, 2022
Revised Apr 13, 2023
Accepted Apr 15, 2023

Keywords:

Climate
Crop selection
Machine learning
Random forest
Regression models

ABSTRACT

In India, agriculture plays an important role in the nation's gross domestic product (GDP) and is also a part of civilization. Countries' economies are also influenced by the amount of crop production. All business trading involves farming as a major factor. In order to increase crop production, different technological advancements are developed to acquire the information required for crop production. The proposed work is mainly focused on suitable crop selection across districts in Tamil Nadu, considering phenotype factors such as soil type, climatic factors, cropping season, and crop region. The key objective is to predict the suitable crop for the farmers based on their locations, soil types, and environmental factors. This results in less financial loss and a shorter crop production timeframe. Combined feature selection (CFS)-based machine regression helps increase crop production rates. A brief comparative analysis was also made between various machine learning (ML) regression algorithms, which majorly contributed to the process of crop selection considering phenotype factors. Stacked long short-term memory (LSTM) classifiers outperformed other decision tree (DT), k-nearest neighbor (KNN), and logistic regression (LR) with a prediction accuracy of 93% with the lowest classification accuracy metrics. The proposed method can help us select the perfect crop for maximum yield.

Corresponding Author:

Iniyan Shanmugam
Department of Computing Technologies, SRM Institute of Science and Technology
Kaatankulathur, Chennai-603203, India
Email: iniyans@srmist.edu.in

1. INTRODUCTION

India is 70% dependent on agriculture as it is the main source of food and approximately 291.95 million metric tons of total food grain production are produced in India. The country has large productive lands near 15 agro-climatic zones, which are mentioned by Indian Council of Agricultural Research (ICAR) [1]. India's gross domestic product (GDP) mostly originates from farming, as agriculture is a dominant part of a country's economic growth, and therefore selecting the crops becomes very important in planning for agriculture. In general, environmental elements affect crops at every stage of growth, and most of the time it is avoidable, resulting in huge losses for farmers [2]. Agriculture's main goal is to produce as many crops as possible on a small amount of land. It will be much easier for an individual to cultivate a crop with optimum yield if the problem related to crop production is identified [3]. Farmers can benefit from choosing the right crop for a certain piece of land in order to boost crop yield and profit. Due to several complex features, the forecast of crop yield is difficult [4]. Fundamentally, crop yield is dependent on several components, including soil quality, landscapes, pest infestations, genotype, quality and water accessibility, climatic conditions, harvest planning [5]. When fewer resources, such as money, land, or fertilizers, are used to produce a maximum yield, the goal of agricultural planning is achieved, which can be accomplished with the help of various machine learning (ML) and artificial intelligence (AI) techniques [6]. ML approaches such as gradient boosting (GB) and random forest (RF) can be used to learn crop biophysical properties such as leaf area index and chlorophyll content.

Various climatic regions' remote sensing data sentinel-2 is analyzed to predict the crop yield. This work focused on developing agricultural systems that would assist in crop monitoring [7]. The estimation of yield performance using a machine learning approach for corn hybrids. The data was provided in the 2020 Syngenta Crop Challenge [8]. Prediction algorithms such as support vector regression, k-nearest neighbor (KNN) regression, gradient boosting (GB) regression, and decision tree (DT) regression are used in crop yield forecasting for five crops. Feature selection is based on agronomic principles and specific thresholds [9]. The various feature selection methods like aggregated boosted trees and random forest recursive feature elimination were applied to two crops: betel palms and mango plantations. Machine learning and feature selection together resulted in the object-oriented classification of the two crops in tropical agricultural regions [10]. The method for predicting the yield of alfalfa biomass using machine learning algorithms based on weather data and historical data from Kentucky and Georgia ReliefF is the feature selection method adopted [11].

The background study of crop selection and yield estimation is mainly focused on boosting crop output rates during unfavorable conditions, so it is critical to properly select a certain type of crop. It can be accomplished using a variety of methodologies, both manual and AI-based, although the manual method has the disadvantage of higher labor costs and longer processing times. Backtracking is also impossible due to certain errors. Applying univariate and multivariate analysis of features such as rainfall, humidity, seed quality, land quality, terrain or grassland, and yield rate to provide a clear picture of how elements are interdependent and affect yield rate. Crop selection is mostly influenced by favorable and unfavorable weather conditions. Composition of components such as *Cu, K, P, N, Mn, Fe, Ca,* and *carbon's pH* value. All of these characteristics, as well as new ones such as the need of the hour and a farmer's inclination to promote a specific crop in a region, play a vital role. Farmers can use an efficient predictor to help them make crop-growing decisions, and various supervised ML algorithms are engaged in this research work for the effective finding of crop selection and yield estimation for the farmers to avoid financial loss. This research work focused on crop selection and yield estimation processes to assist farmers in making better decisions regarding the optimal time to cultivate crops and the sorts of crops to cultivate based on environmental parameters in order to achieve a higher crop yield. Combined feature selection (CFS) is used for selecting the top k best features from the dataset, which majorly affects the yield estimation and crop selection. Various ML algorithms were analyzed, and an ensemble approach involved in random forest classifiers (RFC) outperformed others and produced better accuracy in the crop selection process.

2. RELATED WORKS

Crop selection method (CSM) to maximize crop yield rate using ML [12]–[14] proposed a system to get rid of crop selection issues. They spent a long time and a short time planning wheat crops and sugarcane crops, as well as deciding on a yield rate. They used CSM to predict crop yield condition, which increased the accuracy of the CSM algorithm in producing better results. A feature-based neural network model for weather forecasting [15] proposed a system for weather forecasting, which is also important in agriculture. So, in this system, weather forecasts have been done with the help of a neural network. Weather [16] primarily includes temperature from minimum to maximum and humidity, which have been predicted using extracted features from time series. Hence, future weather forecasts are done with the help of trained artificial neural networks, and the results came out with higher accuracy [17]. Data mining [18], an effective method for yield estimation and crop yield prediction. Their main challenge was to extract some information from the given raw data, and this was possible with the data mining that extracts the information from the data and classifies. Smart farming [19] toward unmanned farms proposes a new system of unmanned farms that do not require humans to enter farms for analysis. A naive Bayes map reduces an agricultural model's precision [20], and it proposes a method for efficient crop yielding in which the naive Bayes algorithm is used to efficiently create models. This system is useful in numerous crop identification applications; in other words, it is scalable. There is a yield graph that shows the perfect timing to sow seeds, grow plants, and harvest them. Also, in this paper, the author has found the worst as well as the best complexity for crop selection. This system used small-to large-scale farms to predict the model. It can further be useful in pesticide recommendations along with fertilizer and irrigation systems, which can be helpful for crops.

CSM to maximize crop yield rate using ML techniques [21] presented a crop selection method to pick a set of crops that will be planted during the growing season lection method to pick a set of crops that will be planted during the growing season. This technique increases the overall yield rate of plants planted during the growing season. In study [22], the utilization of different grouping methods that can be utilized for crop yield forecasts is described. A couple of information mining techniques, like naive Bayes, J48, arbitrary trees, support vector machines (SVM), and fake neural networks are introduced. Kushwaha and Bhattachrya [23] describe the similarity of a certain harvest for a specific climatic condition and the probability of improving the yield quality by utilizing different climate and sickness datasets. Fathima and Geetha [24] use information mining procedures on consistent data that help with data exposure. They use the k-implies bunching calculation to bunch the ranchers subject to the reap type and water framework limits. After achieving 77% accuracy using the naive Bayes calculation and zeroing in on soil boundaries in [25], researchers are expected to propose and implement a standard-based framework to assess harvest yield creation from a collection of past data by applying affiliation rule mining to farming data from 2000 to 2012. In [26], the proposed crop yield prediction model for soybean and corn crops using an ensemble ML model named a "multilayer stacked ensemble" using mutual information feature selection techniques outperformed other ML techniques. Iniyan and Jebakumar [27] used various ML classifiers for disease detection and classification in tomato plant leaves with the feature extraction techniques termed "CIELAB," which gave good accuracy in the detection and classification of infected diseases [28].

3. PROPOSED WORK AND ITS PURPOSE

Many algorithms proposed in crop selection and agricultural ML algorithm planning were successful. But every technique of ML did not show a high accuracy rate. Many papers concentrated on machine learning techniques for determining accuracy and reducing issues in agricultural planning. As we all know, agricultural products are influenced by a variety of factors such as climate and geography, all of which are extremely sensitive and risky and can be mathematically resolved by algorithms such as naive, SVM, and others. This algorithm gave much accuracy for crop selection with numerous factors to determine the selection of crops to save them from being damaged and also to make the process simpler for one to use it properly.

a. The purpose of incorporating machine learning in the field of crop yield prediction is to minimize the cost and resources involved, including labor costs, transportation costs, pesticide and fertilizer costs, and others.
b. The yield can be maximized to a large extent by having a prediction of the best-suited crop in a particular region based on several parameters and previous data using predictive modeling.
c. Land utilization and quality of crops can be improved by self-learning algorithms that can predict what needs to be done under what circumstances.

3.1. Dataset

In the crop production dataset, there are some initial missing numbers. The average scores of a similar soil variable from other regions can be used to fill in about 7% of the missing values in the soil data for some places. The average value of a similar management variable for other regions in the same year can be used to fill in the missing values in management data, which make up about 6% of the total. When replacing missing numbers, the median and mode methods were explored, but the mean method generated better predictions. The weather data does not contain any missing values.

3.2. Combined feature selection (CFS)

Better crop selection and yield estimation using machine learning algorithms required a good feature selection approach to identify the necessary characteristics in order to deliver more accurate crop selection and yield estimation outcomes. Various feature selection strategies, including CFS, mutual information-based feature selection (MIFS), and lasing-based feature selection (LFS), were utilized to forecast crop yield using ML and an ensemble approach. Based on an analysis of the effectiveness of feature selection strategies, machine learning, and ensemble models, in conjunction with MIFS, the multilayer stacked ensemble regression (MSER) ensemble ML method improved soybean and corn yield prediction accuracy with low error metrics. Based on this conclusion, the proposed work focuses primarily on the ensemble method of feature selection, which combines the K best characteristics of multiple feature selection strategies used in crop selection and yield estimation using machine learning models. Incorporating ANNOVA F-statistic feature selection (AFFS), MIFS, and recursive feature selection (RFS) into the CFS procedure resulted in better outcomes when applied to machine learning models, as shown in Figure 1.

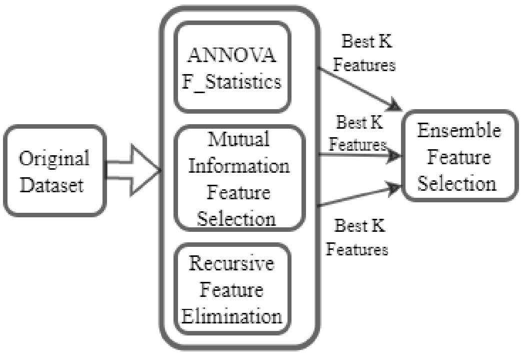

Figure 1. Combined feature selection

4. WORKFLOW AND DESIGN

The Figure 2 work flow diagram depicts how the system operates in crop variety selection. The first input is given, which implies we are given a set of datasets to classify. After that, the data is split into two parts: testing data and training data, with testing data accounting for 30% of the total and training data accounting for 70%. Training data will be labelled data that is predetermined and will be used to make predictions, as well as testing data to see how well it works. After training and testing, the data manipulation section is shown, which includes three processes: normalization, data cleaning, and transformation. The initial data will be studied in this step of data manipulation, which is a superior system. Second, during data cleaning, high cardinality factors are cleaned here. In transformation, the objective variable was erased from all gathered data and replaced with a straight-out factor in the model framework using one-hot encoding. The value 0 was assigned to the information's missing attributes. The persistent factors are scaled using min-max standardization, which converts esteems from 0 to 1 to square factors, causing the coefficients to shift dramatically. After the data modification step, the model selection training is completed. Pre-processed data was used in the creation and testing of the informative index. In supervised ML, data partitioning is a classification that is used to separate predefined classes of data. Both training and testing datasets can be obtained using data partitioning. If the training is successful, go to the trained model to acquire the desired result; if the training is unsuccessful, restart the training model selection process. In the assessment, we rated the classifier on inconspicuous test data and determined the R squared data for both the preparation and test information on the output, and the result will be displayed on the interface, which can be an Android application.

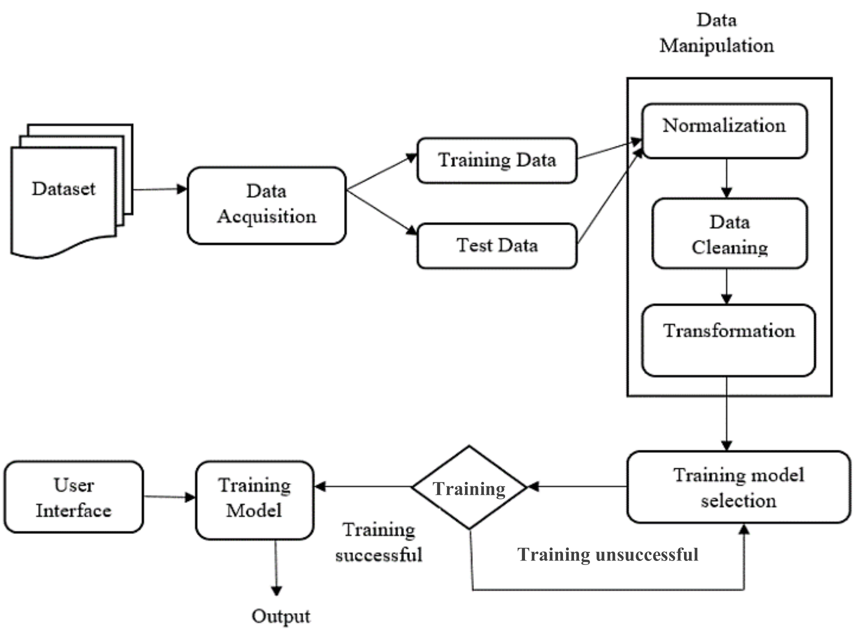

Figure 2. Work flow of crop selection and yield estimation

4.1. Models analyzed

In ML, regression algorithms are frequently employed to solve prediction issues. In order to forecast the crop production affected by numerous environmental parameters, including weather, soil, crop seasons, and crop regions, our proposed model also included regression techniques. The CFS-based model is evaluated with other regression techniques in terms of yield prediction accuracy levels using several accuracy measures like root mean square error (RMSE), mean absolute error (MAE), and mean square error (MSE). There are different machine learning approaches involved in this work to come up with the best crop selection in terms of environmental parameters.

4.1.1. Decision tree (DT)

A DT is a tree-model with left and right subtrees that predicts various outcomes such as likelihood, resource costs, and utility costs. It checks for impurity measurements to see if the stated requirements are met, and the node with the best result is chosen to divide. It is made up of flow control techniques that are conditional. Information gain and entropy are the two fundamentals.

4.1.2. Random forest (RF)

RF is made up of *n* DTs, each with its own set of parameters and training data. Assume a RF bag has 100 DTs. Because the settings and training data set are varied, the prediction obtained from each D-tree will differ significantly. Assume that each of these 100 trees was trained on a different subset of data. So, to make a final choice based on certain test data for a particular decision tree, use the voting method. The majority of trees' anticipated outcome is chosen as a feature to split. Because each feature has a varied value of entropy and information gain, the D-tree result changes from one tree to the next. Starting from the root node to the leaf node, RFs are used to find the most prevalent prediction and the feature that corresponds to it. When compared to a single D-tree, RF is more reliable and stable. It is the same as polling all ministers before making a decision instead of the Prime Minister making a single decision.

4.1.3. K-nearest neighbor (KNN)

It belongs to the supervised learning category and is beneficial for resolving classification and regression problems. For predictive classification, it is generally favored. It is also termed a "lazy learning algorithm" since it does not instantly learn from the training set. Instead, it saves the dataset and takes an action on it during the time of classification. It also has a non-parametric approach, so it makes no assumptions about the underlying data. It also suits our crop classification and yield estimation model. The results of KNN regression and classification also have great significance.

4.1.4. Multiple linear regression (MLR)

Numerous linear regressions are employed to estimate the outcome variable given multiple input factors. Multiple linear regression (MLR) is a statistical method used in ML to determine the linear relation between a dependent or desired output variable and an independent input variable. Since the target yield parameter has a linear relationship with the other independent input factors such as weather, soil, and other variables, MLR is more applicable to our problem. The crop yield parameter is considered an output variable, whereas phenotype factors like temperature, humidity, soil type, crop season, and crop region are treated as input parameters. The MLR applied to the dataset identifies the multiple significant parameters that affect crop yield through various training and testing processes to predict crop yield.

4.1.5. Logistic regression (LR)

It belongs to the supervised learning category and can be used to predict category-wise dependent variables from independent variables. It predicts the output of dependent numbers and returns a binary value of 0 or 1, true, false, and so on. But, for the most part, it provides value in a statistically significant range between 0 and 1. It can be observed that logistic regression and linear regression are similar, but the approaches used are different. Classification is aided by LR, while regression is aided by linear regression. Instead of using regression to fit a line, use the S-shaped logistic function, which produces two outputs ranging from 0 to 1. It is a notable ML algorithm because it can provide probability and do classification of fresh data using discrete data and classify diverse types of data.

5. RESULTS AND DISCUSSION

In general, environmental factors affect crops at every stage of growth, and most of the time it is avoidable, resulting in huge losses for farmers. There are many parameters that absolutely affect market rate, production rate, government policies, and so on. Crop cultivation is dependent on several parameters, including the geographical area, which can be a hilly area, a ground river, or a deep area, as well as the weather condition, which can be rainfall, temperature, humidity, or a cloudy day. The type of soil also plays an important role, like clay, saline, or sandy soil, and the composition of the soil, like its pH value, potassium, nitrogen, magnesium, copper, calcium, and iron. These different parameters are used for making predictions on crop selection. Earlier or traditional methods were based on the previous results of crop selection. Expert farmers used to analyze previous reports of their farming and proceed according to that, but that was a risky method because every time the same analysis could not produce greater results, and ultimately farmers had to face huge losses, which also directly affected economic growth.

5.1. Crop selection method

To improve crop yields, crop selection systems employ categorization algorithms. Depending on the level of accuracy we achieve, the algorithms may change. Naive Bayes, DT, RF classifiers, ensemble models, and advanced ML techniques such as artificial neural networks (ANN) and back propagation algorithms are among the algorithms used. Predictive modelling is divided into several levels. Extraction and collection of data are the first steps. Exploration and transformation of data is the second one. Predictive modelling is the third step. Finally, model implementation and deployment true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), recall and precision, the area under the curve of the receiver operating characteristic (AUC-ROC) metric, and log loss are some of the metrics used to measure accuracy. Keras, Scikit-Learn, Matplotlib, NumPy, TensorFlow, and Pandas are some of the Python libraries that will be used.

5.2. Data exploration

Using data exploration techniques, many classes are present in this data set, and some refer to null values. Feature-wise exploration is very much needed to analyze the importance of features to get more accuracy for further classification and estimation, as shown in Figure 3. As taken from the dataset from the past 10–12 years, and based on this dataset, the result will be evaluated. Examine the features provided in each column in the dataset. The graph depicts crop year and production. In this graph, we can see that as the year progresses, so does the production rate. Based on the existing data, analysis has been performed, as can be seen in the graph. This crop selection and yield estimation dataset consists of varieties of crop production data with respect to different environmental parameters such as soil type, temperature, humidity, and ph. Figure 4 clearly show that the feature exploration on rice crop production is primarily determined by district. The effective crop selection and yield estimation process mainly depends on the micro-level data or feature exploration strategies to be carried out.

5.3. Crop yield estimation

The proposed crop yield estimation model based on CFS is compared comprehensively with various machine learning algorithms. It was determined that the percentage of yield estimation accuracy, the regression precision parameter metrics, and the measurement between actual and estimated crop yield values were the most important parameters for comparative analysis. The initial crop production dataset was preprocessed by filling in the missing data; dimensionality reduction was also done in weather parameters. The proposed CFS used for selection of the k best features from different feature selection techniques was applied with various ML regression algorithms such as MLR, DT, KNN, and RF for better crop yield estimation. The CFS-based RF regression achieved higher accuracy when compared to the other regression algorithms. Figures 5 to 8 show the DT, KNN, MLR, and RF prediction results in terms of actual and predicted yields of rice crops based on phenotypic or environmental factors.

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production	Soil Type	temperature	humidity
count	13547	13547	13547.000000	13547	13547	13547.000000	1.326600e+04	13547	13547.000000	13547.000000
unique	1	31	NaN	3	87	NaN	NaN	24	NaN	NaN
top	Tamil Nadu	DINDIGUL	NaN	Whole Year	Groundnut	NaN	NaN	Laterite,Red Sandy	NaN	NaN
freq	13547	600	NaN	8117	546	NaN	NaN	1413	NaN	NaN
mean	NaN	NaN	2004.877537	NaN	NaN	7078.900642	9.103304e+05	NaN	27.103807	65.704816
std	NaN	NaN	4.714027	NaN	NaN	20874.776934	2.108750e+07	NaN	7.490034	23.828948
min	NaN	NaN	1997.000000	NaN	NaN	1.000000	0.000000e+00	NaN	8.825675	10.034048
25%	NaN	NaN	2002.000000	NaN	NaN	48.500000	5.700000e+01	NaN	22.920523	55.508826
50%	NaN	NaN	2004.000000	NaN	NaN	624.000000	8.410000e+02	NaN	26.030973	68.539971
75%	NaN	NaN	2009.000000	NaN	NaN	4472.500000	9.067250e+03	NaN	29.340573	83.789115
max	NaN	NaN	2013.000000	NaN	NaN	367554.000000	1.250800e+09	NaN	54.986760	99.981876

Figure 3. Feature wise exploration on crop selection dataset

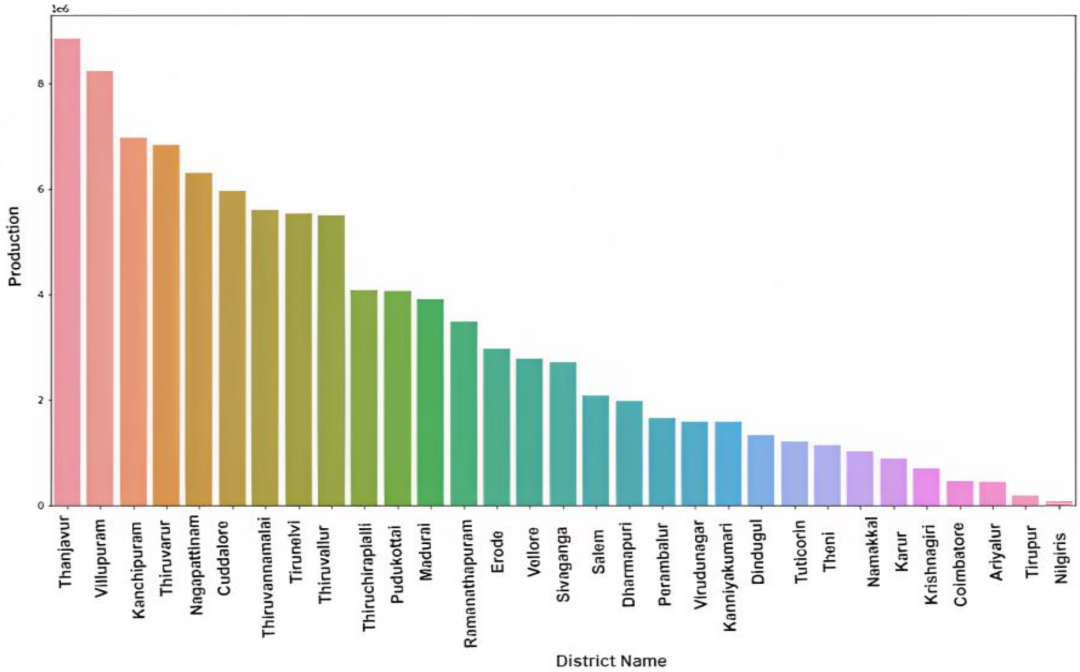

Figure 4. District wise rice crop production

Table 1 shows the comparative performance analysis of various ML regressions based on CFS for crop yield estimation or prediction. Table 2 shows the comparative performance analysis of various ML classifiers based on CFS for crop selection or classification. Figure 9 shows the performance analysis of crop modelling with machine and deep learning models. Figure 10 demonstrates the graphical interface of the crop selection model based on CFS identifying the suitable crop. CFS-based crop selection and yield estimation outperform the earlier MIFS-based crop yield prediction on corn and soybean crops [27]. Table 1 shows test data accuracy and loss percent. The accuracy with recurrent neural network (RNN) as well as long short-term memory (LSTM) neural networks is predicted in Table 2. Ten epochs are used to train and test the model. The accuracy has not much increased after 10 epochs. The model has been trained for a minimum of 20 epochs, increasing accuracy. RNN as well as LSTM networks are used to compare performance in each epoch in Table 3. Our research demonstrates that in this dataset, both stacked LSTM as well as RNN produce satisfactory results. However, when compared to the RNN, stacked LSTM exceeds the best outcome.

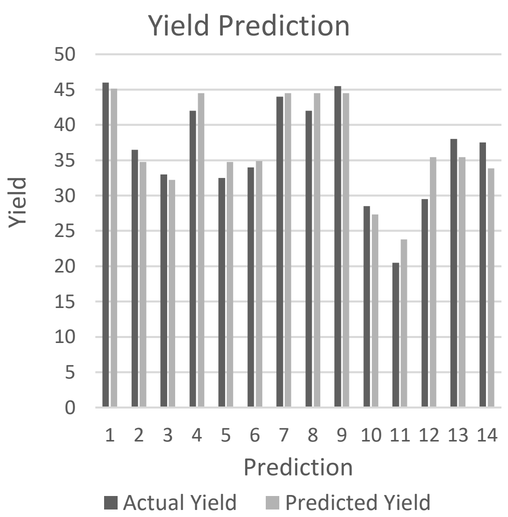

Figure 5. Actual versus predicted rice crop yield using DT

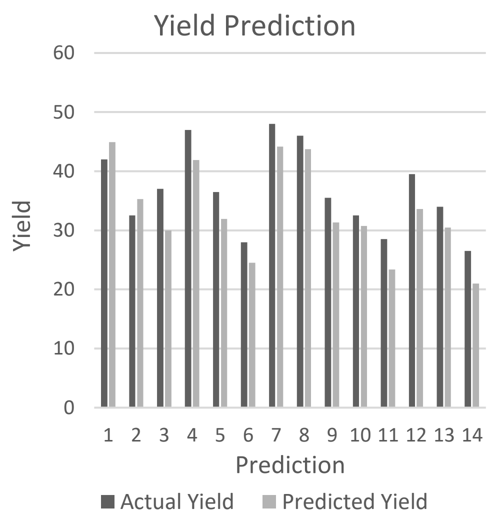

Figure 6. Actual versus predicted rice crop yield using KNN

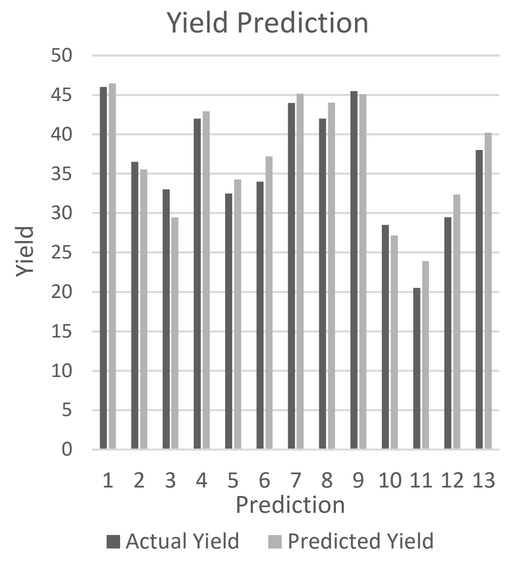

Figure 7. Actual versus predicted rice crop yield using MLR

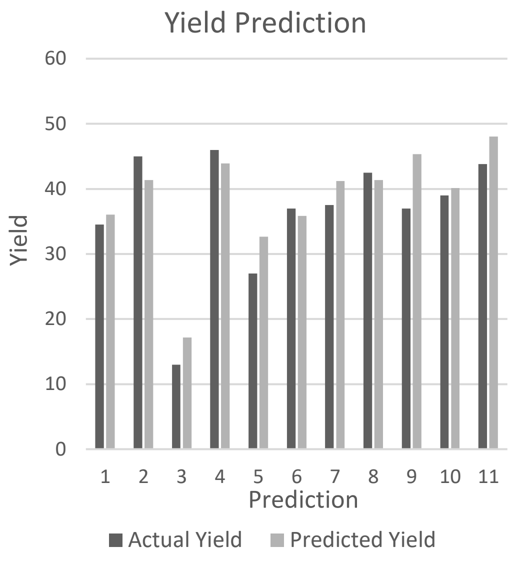

Figure 8. Actual versus predicted rice crop yield using RF

Table 1. Performance analysis of CFS based machine and deep learning classifiers

ML regressors	Estimated yield accuracy (%)	Precision	Recall	F1 score
MLR	87.31	0.87	0.89	0.87
DT regression	82.17	0.83	0.82	0.91
KNN regression	79.24	0.78	0.81	0.79
RF regression	88.21	0.87	0.93	0.88
RNN regression	89.32	0.89	0.92	0.91
Stacked LSTM regression	92.54	0.92	0.94	0.91

Table 2. Performance analysis of CFS based machine and deep learning classifiers

ML classifiers	Crop classification accuracy (%)	Precision	Recall	F1 score
DT classifier	86.67	0.87	0.86	0.87
LR classifier	76.47	0.76	0.78	0.81
KNN classifier	77.78	0.77	0.79	0.82
RF classifier	87.28	0.88	0.87	0.90
RNN classifier	89.51	0.89	0.90	0.91
Stacked LSTM classifier	93.38	0.93	0.94	0.93

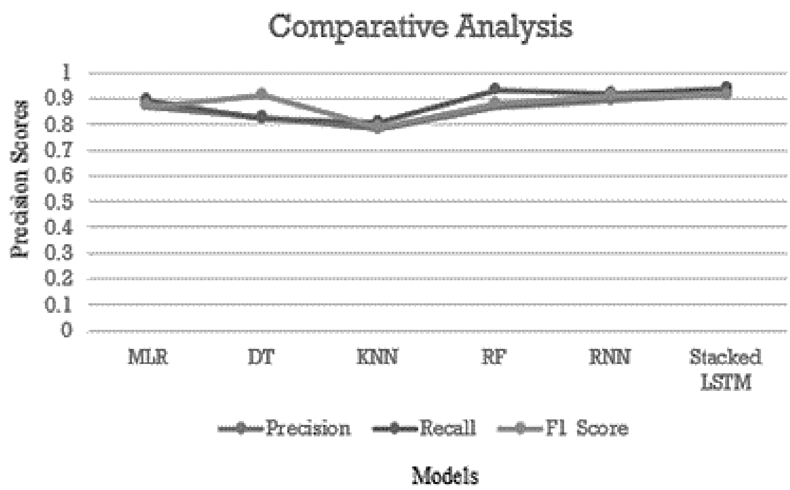

Figure 9. ROC curves show the performance of crop modelling

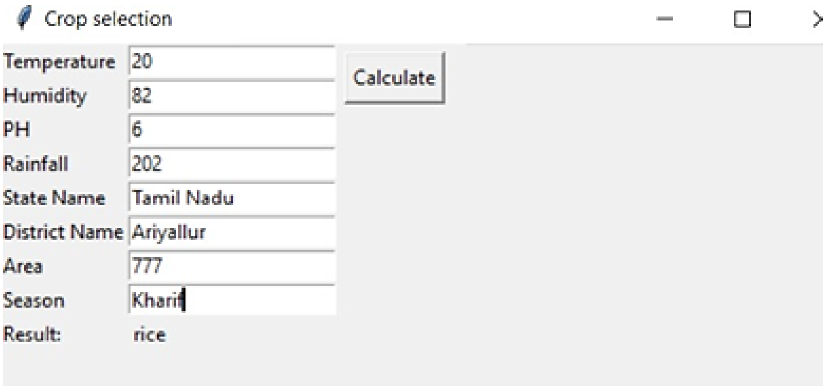

Figure 10. Graphical user interface of crop selection

Table 3. Accuracy in test data of each epoch

No of epochs	Stacked LSTM accuracy	Stacked LSTM Loss	RNN accuracy	RNN loss
1	0.91	7.01	0.80	19.12
2	0.84	14.74	0.75	26.31
3	0.87	12.32	0.77	24.14
4	0.90	9.75	0.81	18.13
5	0.91	8.25	0.82	20.00
6	0.92	8.65	0.86	13.21
7	0.92	8.21	0.86	13.21
8	0.87	13.75	0.88	11.14
9	0.88	11.95	0.88	11.14
10	0.93	7.14	0.89	11.13

6. CONCLUSION

The proposed CFS-based crop selection and yield estimation is useful for crop selection and yield estimation in a specific region and under specific climatic conditions. The weather type, region, state, district, soil type, and water density are all taken into account. as these are the most essential factors influencing crop productivity. As a result, the suggested model handles the problem of crop selection by taking into account all necessary criteria such as soil type, climatic conditions, state, district, and soil water content. This model can determine which crop is most suited for that particular field and thereby boost agricultural yield in that area. In view of this crop selection problem, numerous efforts have been made to save farmers from huge losses. The proposed CFS was used for the selection of the *k* best features from different feature selection techniques and was applied with various ML regression and classification algorithms. There are many algorithms available that can be used in prediction. Some of them are DT, KNN, LR, SVM, and RF.

A brief comparative analysis was also made between various ML regression algorithms, which majorly contributed to crop selection and yield estimation considering phenotypic factors. With the fewest classification and regression parameter metrics, the stacked LSTM classifier and regressor outperformed others with prediction accuracy of 93.388% and 92.541%, respectively. In future, the researchers may include both the genotype and phenotype factors for effective crop selection and yield prediction process. Ensemble deep learning approach may also provide better crop modelling.

REFERENCES

[1] B. B. Sinha and R. Dhanalakshmi, “Recent advancements and challenges of internet of things in smart agriculture: A survey,” *Future Generation Computer Systems*, vol. 126, pp. 169–184, Jan. 2022, doi: 10.1016/j.future.2021.08.006.

[2] T. A. Shaikh, T. Rasool, and F. R. Lone, “Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming,” *Computers and Electronics in Agriculture*, vol. 198, 2022, doi: 10.1016/j.compag.2022.107119.

[3] A. D. J. van Dijk, G. Kootstra, W. Kruijer, and D. de Ridder, “Machine learning in plant science and plant breeding,” *iScience*, vol. 24, no. 1, Jan. 2021, doi: 10.1016/j.isci.2020.101890.

[4] C. Bian *et al.*, “Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV data,” *Remote Sensing*, vol. 14, no. 6, Mar. 2022, doi: 10.3390/rs14061474.

[5] M. F. Danilevicz, P. E. Bayer, F. Boussaid, M. Bennamoun, and D. Edwards, “Maize yield prediction at an early developmental stage using multispectral images and genotype data for preliminary hybrid selection,” *Remote Sensing*, vol. 13, no. 19, Oct. 2021, doi: 10.3390/rs13193976.

[6] B. Aissa, T. Hamza, G. Yacine, and N. Mohamed, "Impact of sensorless neural direct torque control in a fuel cell traction system," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 2725–2732, Aug. 2021, doi: 10.11591/ijece.v11i4.pp2725-2732.
[7] M. Kganyago, P. Mhangara, and C. Adjorlolo, "Estimating crop biophysical parameters using machine learning algorithms and sentinel-2 imagery," *Remote Sensing*, vol. 13, no. 21, Oct. 2021, doi: 10.3390/rs13214314.
[8] F. B. Sarijaloo, M. Porta, B. Taslimi, and P. M. Pardalos, "Yield performance estimation of corn hybrids using machine learning algorithms," *Artificial Intelligence in Agriculture*, vol. 5, pp. 82–89, 2021, doi: 10.1016/j.aiia.2021.05.001.
[9] D. Paudel *et al.*, "Machine learning for large-scale crop yield forecasting," *Agricultural Systems*, vol. 187, Feb. 2021, doi: 10.1016/j.agsy.2020.103016.
[10] H. Luo *et al.*, "Combinations of feature selection and machine learning algorithms for object-oriented betel palms and mango plantations classification based on Gaofen-2 imagery," *Remote Sensing*, vol. 14, no. 7, Apr. 2022, doi: 10.3390/rs14071757.
[11] C. D. Whitmire, J. M. Vance, H. K. Rasheed, A. Missaoui, K. M. Rasheed, and F. W. Maier, "Using machine learning and feature selection for Alfalfa yield prediction," *AI*, vol. 2, no. 1, pp. 71–88, Feb. 2021, doi: 10.3390/ai2010006.
[12] R. Kumar, M. P. Singh, P. Kumar, and J. P. Singh, "Crop selection method to maximize crop yield rate using machine learning technique," in *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, May 2015, pp. 138–145, doi: 10.1109/ICSTM.2015.7225403.
[13] K. Kaur, "Machine learning: applications in Indian agriculture," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 4, pp. 342–344, 2016.
[14] S. M. Paras, A. Kumar, and M. Chandra, "A feature based neural network model for weather forecasting," *International Journal of Computational Intelligence*, vol. 4, no. 3, pp. 209–216, 2009.
[15] K. K. Kumar, K. R. Kumar, R. G. Ashrit, N. R. Deshpande, and J. W. Hansen, "Climate impacts on Indian agriculture," *International Journal of Climatology*, vol. 24, no. 11, pp. 1375–1393, Sep. 2004, doi: 10.1002/joc.1081.
[16] A. Raorane and R. Kulkarni, "Data Mining: An effective tool for yield estimation in the agricultural sector," *International Journal of Emerging Trends and Technology in Computer Science (IJETTCS)*, vol. 1, no. 2, pp. 1–4, 2012.
[17] H. Navarro-Hellín, J. Martínez-del-Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sánchez, "A decision support system for managing irrigation in agriculture," *Computers and Electronics in Agriculture*, vol. 124, pp. 121–131, Jun. 2016, doi: 10.1016/j.compag.2016.04.003.
[18] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environment, Development and Sustainability*, vol. 22, no. 6, pp. 5687–5708, Aug. 2020, doi: 10.1007/s10668-019-00445-x.
[19] R. Priya, D. Ramesh, and E. Khosla, "Crop prediction on the region belts of India: a naïve Bayes MapReduce precision agricultural model," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2018, pp. 99–104, doi: 10.1109/ICACCI.2018.8554948.
[20] A. Verma, A. Jatain, and S. Bajaj, "Crop yield prediction of wheat using fuzzy c means clustering and neural network," *International Journal of Applied Engineering Research*, vol. 13, no. 11, pp. 9816–9821, 2018.
[21] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques," in *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)*, 2016, pp. 1–4, doi: 10.1109/ICCTIDE.2016.7725357.
[22] S. Kujawa and G. Niedbała, "Artificial neural networks in agriculture," *Agriculture*, vol. 11, no. 6, May 2021, doi: 10.3390/agriculture11060497.
[23] A. K. Kushwaha and S. Bhattachrya, "Crop yield prediction using Agro algorithm in Hadoop," *International Journal of Computer Science and Information Technology and Security (IJCSITS)*, vol. 5, no. 2, pp. 271–274, 2015.
[24] G. N. Fathima and R. Geetha, "Agriculture crop pattern using data mining techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 5, pp. 781–786, 2014.
[25] N. Hemageetha, "A survey on application of data mining techniques to analyze the soil for agricultural purpose," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 3112–3117.
[26] S. Iniyan, R. Jebakumar, P. Mangalraj, M. Mohit, and A. Nanda, "Plant disease identification and detection using support vector machines and artificial neural networks," in *Advances in Intelligent Systems and Computing*, 2020, pp. 15–27.
[27] S. Iniyan and R. Jebakumar, "Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (MSER)," *Wireless Personal Communications*, vol. 126, no. 3, pp. 1935–1964, Oct. 2022, doi: 10.1007/s11277-021-08712-9.
[28] Y. Boujoudar, M. Azeroual, H. Elmoussaoui, and T. Lamhamdi, "Intelligent control of battery energy storage for microgrid energy management using ANN," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 2760–2767, Aug. 2021, doi: 10.11591/ijece.v11i4.pp2760-2767.

BIOGRAPHIES OF AUTHORS

Iniyan Shanmugam received a Bachelor's degree in Computer Science and Engineering from Bharathiyar College of Engineering and Technology, Pondichery University in 2010 and a Master's degree in Computer Science and Engineering from Sriram Engineering College of Anna University, Chennai in 2012. He is currently an Assistant Professor in the Department of Computing Technologies and pursuing his Ph.D. degree at SRM Institute of Science and Technology, Chennai, India. His research interest includes data analytics, internet of things, network security, wireless sensors network, machine learning and deep learning. He can be contacted at email: iniyans@srmist.edu.in.

Jebakumar Rethnaraj received a Master's degree in Computer Science and Engineering from the Sathyabama University, Chennai in 2005. He obtained his Ph.D. degree in the area of Information and Communication Engineering from Anna University Chennai in 2015. He is currently an Associate Professor in the Department of Computing Technologies at SRM Institute of Science and Technology, Chennai, India, Working here since 2006. His research includes wireless sensor networks, mobile ad hoc networks, cloud computing, big data, data mining and IoT. He can be contacted at email: jebakumr@srmist.edu.in.

Gayathri Mani is an Assistant Professor in Department of Computing Technologies, S.R.M Institute of Science and Technology, Kattankulathur, completed her Ph.D. in the area of biometrics. She has over eleven years of experience in teaching. Her research interest is security and privacy in biometrics, network security, internet of things, machine learning and deep learning. She can be contacted at email: gayathrm2@srmist.edu.in.