SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I7P104

Volume 10 Issue 7, 33-43, July 2023

Original Article

Forecasting the Factors Responsible for Improving the Yield of Sugarcane Crop using Artificial Neural Network

S. Muthukumaran[1], N. Gnanasankaran[2], Sandhya Soman[3], G. Rakesh[4]

[1]*Department of Master of Computer Applications, Dhanalakshmi Srinivasan College of Engineering and Technology, Mamallapuram, Tamil Nadu, India.*
[2,4]*Department of Computer Science, Thiagarajar College, Madurai, Tamil Nadu, India.*
[3]*Department of Computer Applications, GITAM Deemed to be University,Bangalore, Karnataka, India.*

[1]*Corresponding Author : muthumphil11@gmail.com*

Received: 24 April 2023 Revised: 22 June 2023 Accepted: 13 July 2023 Published: 31 July 2023

Abstract *- Sugarcane is a cash crop cultivated in India's Tropical and Sub Tropical regions, contributing 5.7% of the Gross Domestic Product (GDP) to the Indian Economy. Sugarcane Farming gives employment opportunities to 60 million rural families. It is cultivated all over India to a latitude of 80◦ N to 300◦ N. Sugarcane contains sucrose in its stem juice and is the primary raw material for producing sugar. Sugarcane crop production depends on the season, biological, and economic cause. The growing demand for sugarcane worldwide incorporates the backbone of sugarcane agriculture. This paper proposed A Hybrid Machine Learning Model (HMLM) for forecasting the sugarcane yield, which helps both farmers and the sugar mills to assist in annual planning. The proposed model used Backward Feature Elimination to select the factors that dominate sugarcane production. K-Means Clustering (K-MC) is applied to the selected attributes, and the dataset is partitioned. An Artificial Neural Network (ANN) is created for each clustered dataset, and the parameters influencing sugarcane production are found. The performance of each Neural Network (NN) created was analysed by implementing performance measures, and the findings were compiled. The information gained from the suggested model aids farmers in making decisions to increase sugarcane crop productivity.*

Keywords *- Artificial Neural Network, Backward feature elimination, K-means clustering, Machine learning, Sugarcane production.*

1. Introduction

Sugarcane has been cultivated in India since the Vedic period, found in the Vedic writings around 1400 to 1000 B.C. Sugarcane is a cash crop, and it grows well in tropical regions. It is cultivated over 49.18 lakh hectares throughout India. After the textile industry, the significant contribution to India's GDP comes from the sugarcane industry, which is 6%. Sugarcane is a primary raw material, and the sugar and juice extracted from sugarcane are used for manufacturing white sugar, brown sugar and Jaggery. A man consumes an average of 24 kg of sugar per year. Sugar is extracted 80% from sugarcane and 20% from sugar beet [1]. Molasses and Bagasse are the main by-products of sugarcane, and the sugarcane Bagasse extracted after manufacturing sugarcane is used to make paper, plastic, etc. Ethyl alcohol, Butyl alcohol is manufactured using molasses in distilleries [2, 3].

Molasses is used as food for livestock, and Rum is made from molasses. The green tops obtained during sugarcane harvesting are used as food for cattle. Sugarcane grows to a height of 5 to 6 meters with multiple stems. Sugarcane has high efficiency in storing solar energy and converting this solar energy into sucrose [4, 5]. Sugarcane juice has nutrients such as Carbohydrates, Protein, Calcium, Iron, Potassium and Sodium. The sugarcane cultivation and sugar industry help improve rural areas' socio-economic development by providing job opportunities to 7.5 percent of the rural population. Nearly 60 million sugarcane farmers depend on sugarcane farming by doing labour work in harvesting and ancillary activities [6].

1.1. Motivation

Land cultivation is very low when compared with these cash crops and thus results in a low sugarcane yield. Sugar Manufacturing Factories work only during the production season of sugarcane; the remaining time, the workers remain idle. This causes a severe financial crisis for both factories and workers. Some of the essential challenges faced by sugarcane farmers were using alternative sweeteners for sugar, like Acesulfame potassium, Agave nectar and honey, to decrease the use of sugar. High labour costs, costly machinery for sugarcane harvesting, and bribes asked by sugar factory workers decrease sugarcane yield. The price fixed for sugarcane by the government is also one of the major causes of the decline of sugarcane production.

1.2. Objective

The use of sugarcane is very high, but the yield of sugarcane is meagre. To overcome this demand high yielding varieties having early maturing crops should be cultivated. Any researcher should take all the factors affecting the production of sugarcane cultivation and find the appropriate solution to those problems. The following research contributions fulfil the objective of this research.

- This paper proposed a Hybrid Machine Learning Model to predict the factors responsible for the growth of the sugarcane variety such as CoC(Sc)22, CoC90063, TNAUSCSi7, and TNAUSCSi8 were identified with the help of the proposed model, and the discovered knowledge helps the farmers to improve the sugarcane yield.
- The proposed machine learning model used Backward Elimination to reduce the dimensionality of the Sugarcane Dataset (SD).
- K-MC was utilized for partitioning the dataset and classifying the farmers having high similarity in cropping patterns.
- For the clustered dataset, an ANN was built, and the aspects that sugarcane growers should concentrate on to enhance productivity were found.

2. Literature Review

Shyamal S Virnodkar et al. presented a two-dimensional Convolutional Neural Network (CNN) for classifying sugarcane in Karnataka. The authors collected the Canset dataset from the ESA Copernicus website and preprocessed the RGB images using the SCP plugin. The Sentinel-2 images were Layer Stacked, cropped into 10×10 patches, and then saved in .jpg format. CNN was built using GoogleNet, ResNet50, and DenseNet201 and images were classified as sugarcane and non-sugarcane. The authors proved that DenseNet201 and ResNet50 performed well than the CNN constructed using the remaining structure [7, 8].

Ali Kaab et al. employed ANN to predict sugarcane's input and output energy consumption. The authors use energy indices such as electricity used, machinery used during cultivation, labour cost, pesticides and fertilizers applied during the life cycle of sugarcane production. The author built NN with the parameters and found that reduced tillage, usage of adequate water supply through proper irrigation methods, usage of pesticides and proper nutrition management to sugarcane significantly impact the energy consumption and production of sugarcane cultivation [9].

Ana Claudia, dos Santos, used Random Forest Algorithm to anticipate sugarcane yield by employing images of LandSat, Agronomy and Whether forecasting data. The author collected the sugarcane production dataset from 2012 to 2015 in Brazil. The author built five Random Forest trees with Randomly Selected samples from the dataset. The performance of the Random Forest trees was measured using R^2 and RMSE. Using their proposed methods, the authors identified how the weather, Agronomy, and Energy indices have impacted sugarcane growth [10].

Ate Poortinga et al. presented a deep learning method to automatically map the sugarcane production region of Thailand to reduce Air Pollution caused during sugarcane burning. The authors collected the dataset from the weather department of Thailand from June to November 2017 and created a base map. The input images were preprocessed and stored in Google Cloud and mapped with the images in the Google Earth Engine. CNN was built for the input images using the MobileNetV2 Architecture, and the sugarcane cultivation regions were identified [11].

Lidan Meng et al. studied the reason for irreversible aggregate formation that occurred during sugarcane juice extraction from sugarcane crops. The fresh sugarcane juice was stored at 18° C temperature for six months. The irreversible aggregates were formed due to the interaction of proteins with phenolic and polysaccharides in the sugarcane juice [12].

Pengwen Wang et al. proposed a Multilayer Perceptron for producing sugarcane yield using an IoT-based intelligent farming environment. The authors studied the process of sugarcane transformation into sugar with three ANN's help. The authors relate the proposed method's performance with the classifiers such as decision trees and SVM [13].

3. Methodology

3.1. Data Collection and Dataset Description

The data for this study was gathered in the Tamilnadu districts of Cuddalore and Villupuram. The Villupuram district has 21 agricultural blocks, among them eight blocks, namely Koliyanur, Kandamangalam, Vikravandi, Ulundurpet, Tirunavalur, Tirukoilur, Mugaiyur, Tiruvennainallur were selected.

The Cuddalore district has a total of 13 agricultural blocks, and among them, six blocks, namely Cuddalore, Kurinjipadi, Panruti, Annagraman, Virudhachalam, and Chidambaram, were selected for this research. The profile of the farmers, such as Village details, name, survey number, whether the cultivated land is owned or leased, subsidy from the government, and name of the government scheme under which the crop was cultivated, were collected.

The overall number of farmers who participated in this survey was 10,000, with a minimum of 500 and a maximum of 800 farmers per block. The database used in this research is secondary data collected from District Agricultural Headquarters, Farmers, and various Private Consultancies in Cuddalore and Villupuram districts.

The dataset consists of 26 attributes and 10,000 records. The hectare is the measurement of land sugarcane cultivated by each farmer. A farmer with land below or equal to 2 hectares is classified as a small-scale farmer and practises the conventional furrow irrigation method for irrigating their land. The farmer with land of more than two and a maximum of 6 hectares is classified as a marginal farmer, and they practised drip irrigation for irrigating their land. Hence, the cultivation method attribute has two values, conventional and drip.

The season attributes tell about the season the farmer starts to grow sugarcane crops, and it has three values (Early-December to January, Mid-February to March and Late-April to May). The variety attribute has four values: the varieties of sugarcane taken for this study: CoC (Sc)22, CoC90063, TNAUSCSi7, and TNAUSCSi8. These varieties are chosen because they have a high drought tolerance, and the Tamilnadu Agricultural University of Coimbatore highly recommends cultivating in Cuddalore and Villupuram Districts. This study used two types of soil, namely sandy and clay. The NPK, Super Phosphate, Zinc Sulphate and Ferrous Sulphate attributes have the values of fertilizers applied to the sugarcane crop in kilograms.

The seed rate is the number of sugarcane seeds planted in the field. The phases of Germination, Tillering, Grand Growth, Maturity, or Ripening attributes have the values of water irrigated to the land for each phase in regular intervals, represented in millimetres. The weedicide attribute has the amount of Thiobencarb @1.25kg active ingredient applied to the land per hectare to control weeds. To control pests in the crop, pesticides such as Chlorpyrifos, Chloarantrainilipole, and profenofos were used as maximum ingredients in the pesticide, and it is measured in litres. The cropping type attribute has four values about the cropping type adapted by the farmer: sole sugarcane, sugarcane with intercrop as sweet corn, sugarcane with intercrop as groundnut and sugarcane with intercrop as green gram.

The sugarcane seed rate has the values of the total number of sugarcane buds used for cultivation in the field. The intercrop seed rate has the number of intercrop seeds sowed in the field, measured in kilograms. The intercrop yield has the amount of intercrop yielded in kilograms. The Millable cane attributes have the number of sugarcanes selected by the sugar mills in good condition for manufacturing sugar. The green top attributes have the amount of sugarcane top leaves extracted after removing the sugarcane, represented in tones. The trash attribute has the amount of trash extracted in sugar mills at the end of the sugar manufacturing process and is represented in tones. The sugarcane yield attribute has the overall sugarcane yield in tones.

3.2. Backward Elimination

Backward Elimination is a feature selection model utilized in constructing a machine learning model. It helps remove the attributes from the dataset that do not significantly affect the input variable [14]. The backward elimination process uses a significance value (p) of normally 0.05 and fits all input variables to the regression model. The characteristics with the maximum p-value are disregarded from the attribute list, and the loop continues until all attributes meet the significance threshold. The p-value is used to determine the relation between the independent and dependent variables, and each dependent variable's predicted value (y) may be defined as

$$y = \frac{\hat{p} - p0}{\sqrt{\frac{p0(1-p0)}{n}}} \tag{1}$$

In which $\hat{p}$, $p0$, and n portray the sample proportion in the dataset, the assumed population proportion in the null hypothesis and the overall records in the dataset [15, 16, 17]. A statistical model for examining the association between one or more dependent and independent variables is called Ordinary Least Squares (OLS). The difference between the identified and anticipated values of the dependent variable's sum of squares is minimized, and the equation is represented as a straight line.

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \tag{2}$$

Here the dependent variable is represented as Y, and the model intercept is represented as β_0, X_j is the input variable, and ε is the random error value in the regression model.

3.3. K-Means Clustering

Clustering comes under Unsupervised Machine Learning because the technique does not consider the class label. Cluster Analysis is used to partition the given dataset into different subsets. Data points with a high degree of similarity are sorted into one cluster, while those with a lower degree of similarity are placed into a different cluster. The clustered dataset is further analysed with classification algorithms such as decision tree and NNs, and the patterns hidden in the cluster is effectively identified [18]. This study utilizes a K-MC method to partition the SD. The K-Means method initially chooses the number of clusters to be partitioned and computes a centroid value for each selected cluster.

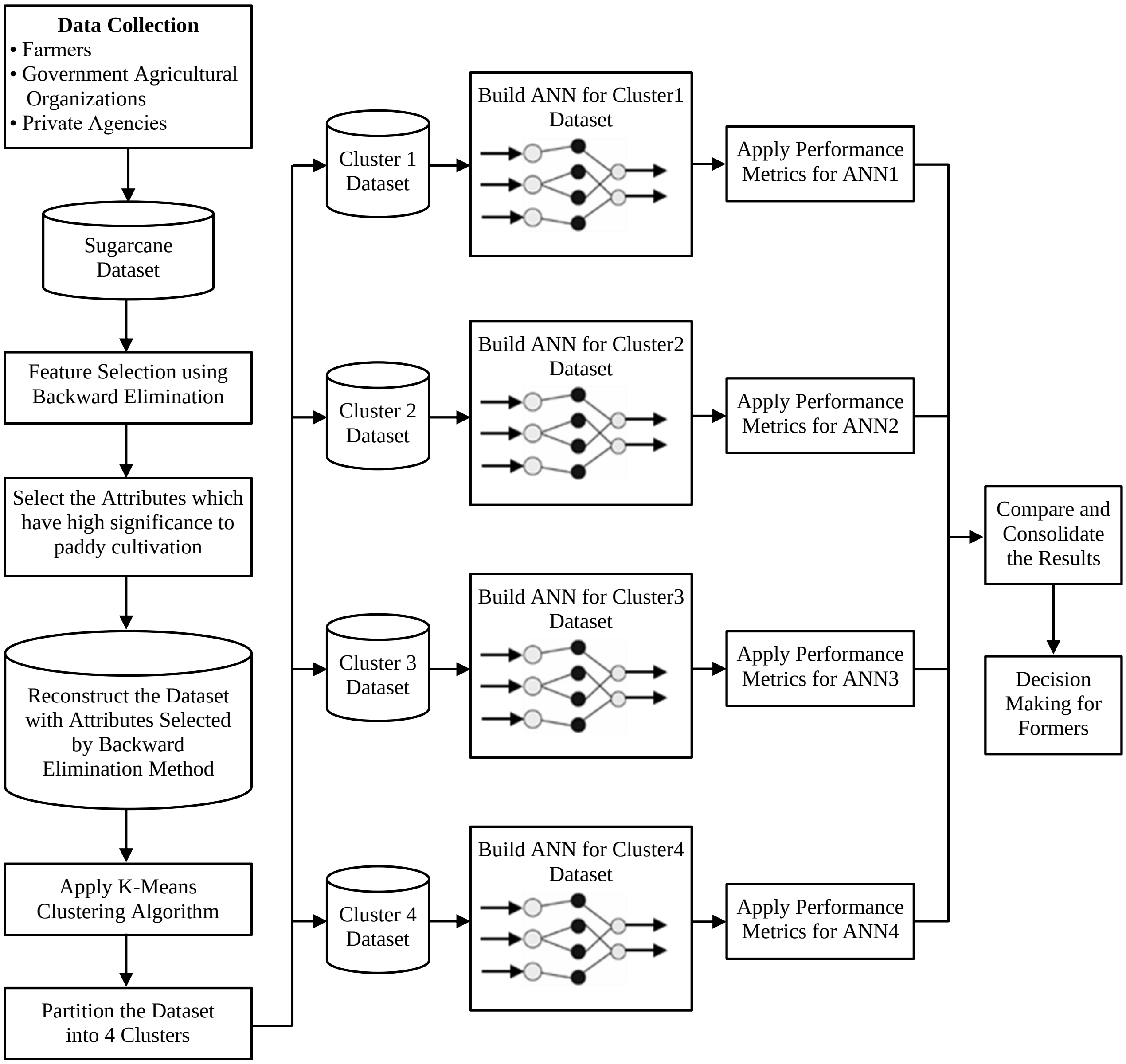

Fig. 1 Proposed Hybrid Machine Learning Model (HMLM) for sugarcane production

The algorithm then computes Euclidean distance for each data point and assigns the data points to a cluster closely related to the centroid distance of the cluster. This algorithm uses an iterative relocation technique to relocate the data points from one cluster to another and ensure the data points are mutually exclusive. The Euclidean distance between the data points in the SD was calculated by

$$d(i,j) = \sqrt{(x_{i1}-x_{j1})^2 + (x_{i2}-x_{j2})^2 + \cdots (x_{in}-x_{jn})^2} \quad (3)$$

The centroid of a cluster is calculated using

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} dist(p, c_i)^2 \quad (4)$$

3.4. Artificial Neural Network

ANN is a Backpropagation Algorithm that imitates the working of a human brain. ANN is a graph-like structure consisting of deeply interconnected nodes, like the nervous system of humans. During the training phase, the ANN learns about the dataset and tries to identify the pattern present [19, 20]. The ANN has an Input Layer (IL) that receives the input and assigns a weight to each input variable. The Transformation function is enforced on the input variables, and the values are fed to the Hidden Layer (HL). For the input values present in the dataset of sugarcane, the output values are computed using.

$$I_j = \sum_i w_{ij} O_j + \theta_j \quad (5)$$

$$O_j = \frac{1}{1+e^{-I_j}} \quad (6)$$

The HL values' weights are computed, and the computed value is then sent to the Output Layer (OL) to match the target class. If the intended output is not achieved, the data is returned to the IL, and the procedure is repeated until the anticipated target class is found [21]. The final output value is computed for each input sent to the OL by computing the error value.

$$Err_j = O_j(1-O_j)\sum_r O_j\left(1-O_j\right)(T_j - O_j) \quad (7)$$

The bias value used to update the values of the OL and feed forward the newly computed value back to the IL again was calculated using.

$$\theta_j = \theta_j + \Delta\theta_j \quad (8)$$

The weights for each input attribute are updated using

$$w_{ij} = w_{ij} + \Delta w_{ij} \quad (9)$$

4. Proposed Hybrid Machine Learning Model for Sugarcane Production

The proposed HMLM approach for sugarcane production is shown in Figure 1. The SD containing categorical attributes was transformed to numerical attributes, and preprocessing was done to ensure the dataset was clean. Backward Elimination was used to choose the most manipulating attributes, significantly improving the sugarcane yield. The SD was reconstructed using the attributes selected by Backward Elimination. K-MC was applied to the dimensionally reduced dataset, and the dataset was further partitioned into four clusters. The backpropagation Algorithm was applied to each clustered dataset, and NN was built. The factors that the farmers and sugar mill owners will concentrate on to improve the sugarcane yield were identified from the constructed NN.

4.1. Steps

The pseudo-code for the HMLM model is given below.

4.1.1. Pseudo Code for Backward Elimination Process

Input : D_s-Sugarcane Dataset, Xi-Input Variables associated to Sugarcane Cultivation, y-Variety of sugarcane (as CoC(Sc)22, CoC90063, TNAUSCSi7, TNAUSCSi8).

Output : A group of variables with a significant influence on sugarcane production

Method

- Set the significance level to 0.5
- For each X_i in DS do
- Using Equation (1), calculate the P value for the overall input variables concerning the target variable y.
- Compute the Ordinal Least Square value using Equation (2)
- Remove the attribute with the highest p-value
- End for
- Update the dataset D_s with the chosen attributes.

4.1.2. Pseudo Code for K-Means Clustering

Input : D_s-Sugarcane Dataset with Attributes selected by Backward Elimination

Output : Clusters of Partitioned dataset

Method

- Set the number of clusters k=4
- Repeat
- Compute the Euclidean distance for each data point using Equation (3)
- Compute the centroid for each cluster using Equation (4)
- Assign each data point to the centroid with the closest distance and form a cluster
- Until no change
- Partition the clusters into each dataset

4.1.3. Pseudo Code for Building the Artificial Neural Network

Input : Partitioned SDs D_{S1}, D_{S2}, D_{S3}, D_{S4}

Output : ANN1, ANN2, ANN3, ANN4-NN constructed for each dataset

Method

- Repeat
- In the SD, establish the weights and biases for the input parameters.
- For every input variable in the IL {
- Calculate input value by utilizing Equation (5)}
- For every input variable in the HL {
- Calculate the net input value concerning the previous value of the IL utilizing Equation (6)}
- For every input in the network built utilizing the SD {
- Calculate Error utilizing Equation (7)}
- For every input in the network built utilizing the SD {
- Calculate Bias utilizing Equation (8)}
- For every input in the network built utilizing the SD {
- Update Weights utilizing Equation (9)}
- While the target class is not anticipated

5. Results

5.1. Outcomes of the Backward Elimination Process

The proposed model is employed in Python 3.8, with Spyder as the Code Editor in the Anaconda Navigator Environment. The SD containing 26 attributes with 9999 records were loaded and preprocessed. The attributes such as

Cultivation Method, Season, Variety, Soil Types and Cropping Type contain categorical values, and it is converted to numerical values using Label Encoding and One Hot Encoding technique with the help of Label Encoder and One Hot Encoder in the sklearn preprocessing package. The dataset, after preprocessing, contains 36 attributes, and it is then applied with OLS Regression Model in the statsmodel package of Python. In the first iteration, the attribute Grand growth phase has the maximum p-value, which is disregarded from the list. In the second iteration, the attribute Gypsum has the maximum p-value, and it is disregarded from the list.

During the successive iterations, the attributes such as Soil type (Clay), Sugarcane Variety (CoC(Sc)22), Water Irrigation level (Germination Phase, Grand Growth Phase, Tillering Phase, Maturity Phase), profenofos, Season of planting, NPK, Super Phosphate, Gypsum, Zinc Sulphate, Ferrous Sulphate, Chloropyrifos, profenofos and Inter Crop Seed Rate were removed from the list. The dataset has undergone 13 iterations; at the 13th iteration, all the attributes have the p-value zero; hence, the iteration stops. The attributes that remained in the dataset were 21, shown in the following Figure 2.

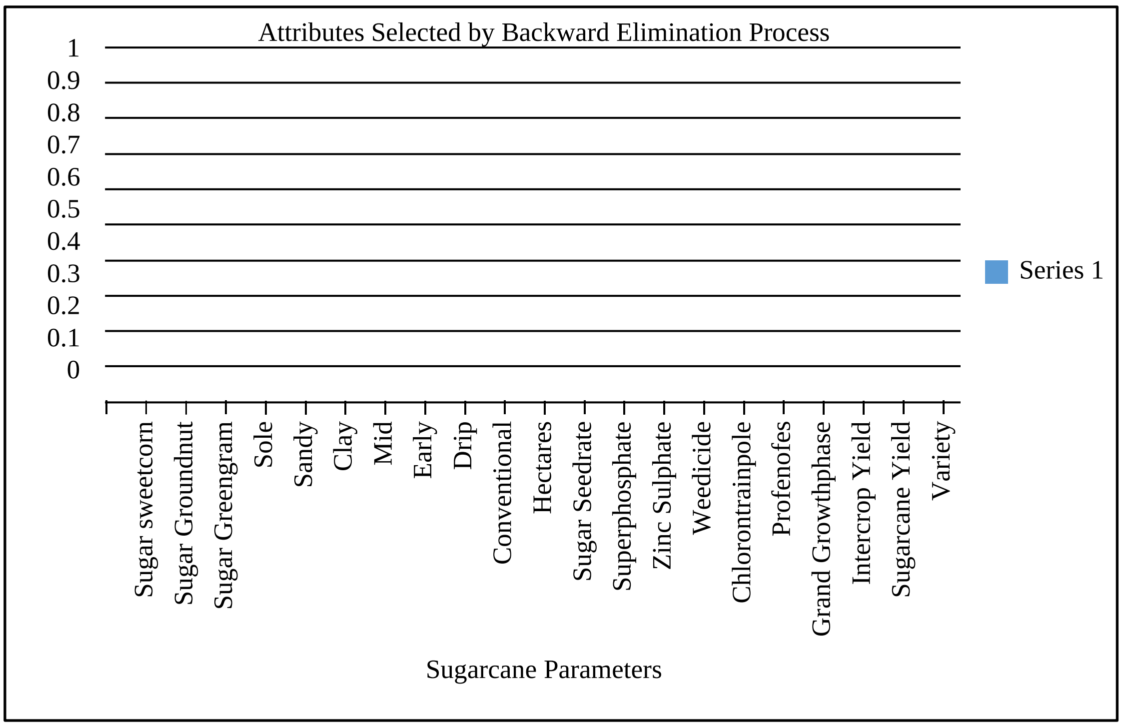

Fig. 2 Attributes elected by backward elimination process

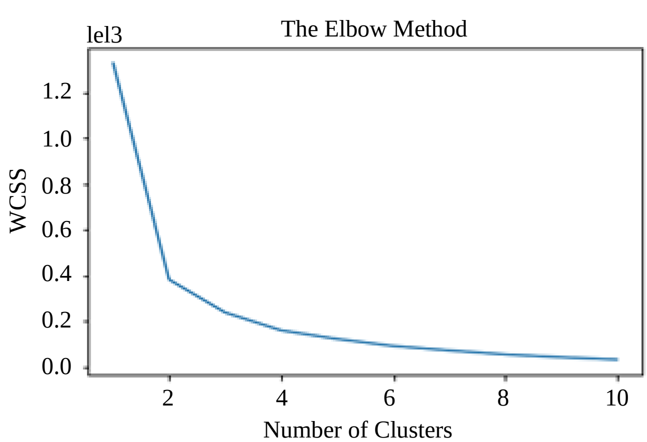

Fig. 3 Elbow method used for selecting the cluster numbers

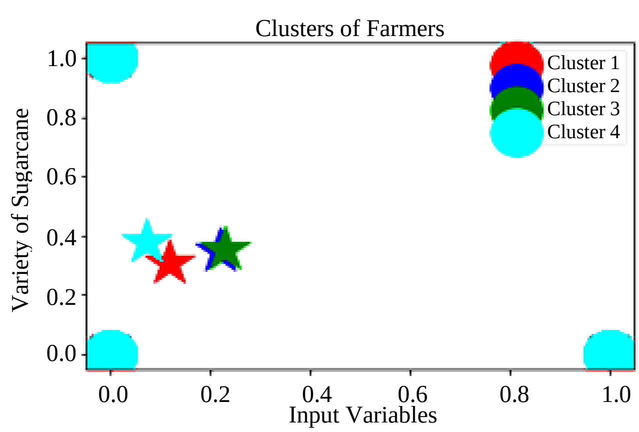

Fig. 4 Clustered SD for k=4

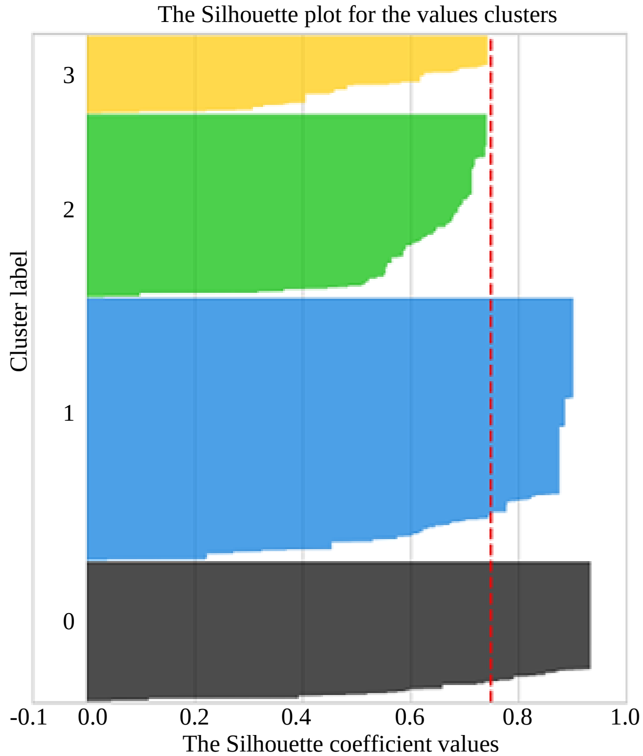

Fig. 5 Performance of K-MC using silhouette coefficients for k=4

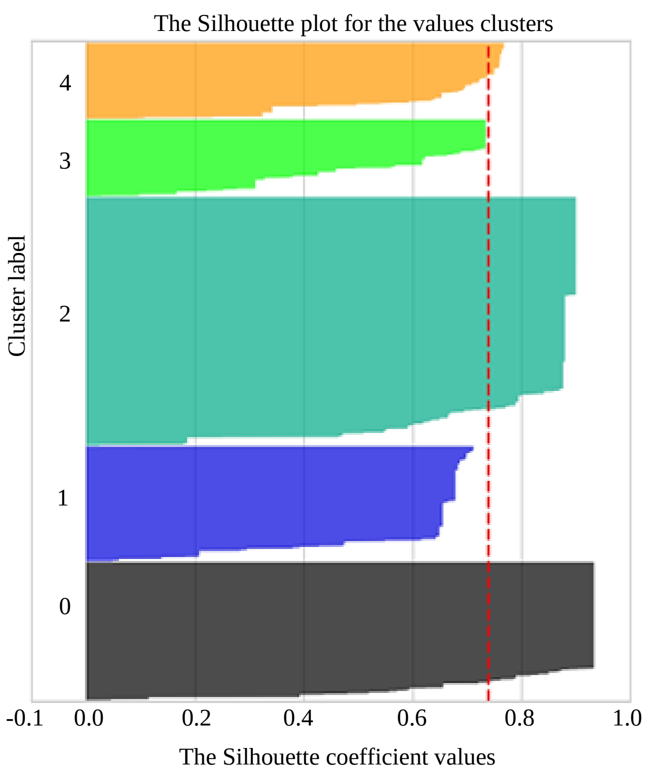

Fig. 6 Performance of K-MC using silhouette coefficients for k=5

5.2. Results for K-Means Clustering

The K-MC was implemented utilizing the K-Means technique present in the sklearn cluster. The number of clusters used to split the dataset was identified using the Elbow method and given in Figure 3. Figure 3 shows that for choosing the cluster numbers as four, the K-Means algorithm accurately clustered the dataset and a graphical representation of the cluster centroids was given in Figure 4. For measuring the performance of the K-Means approach, Silhouette Score is utilized, and the SD was partitioned into different K values such as 2, 3, 4, 5 and 6. The Silhouette Score for various clusters is given in Table 1 given below.

Table 1. Silhouette scores comparison for various clusters

Cluster Number	Average Silhouette Score
Cluster k=2	0.633
Cluster k=3	0.534
Cluster k=4	0.559
Cluster k=5	0.599
Cluster k=6	0.569

From Table 1, it is clear that the performance of the K-Means Algorithm is best for k value 2. Next, the algorithm performs best for clustering the dataset for k values 5 and 6. The Clustering of the dataset into 4 clusters comes into the fourth position, and the last one is the Clustering of the dataset for k value 3.

5.3. Results for Artificial Neural Network

The SD is partitioned into four individual datasets because the Elbow method suggests the dataset is accurately classified for k=4, as shown in Figure 3. The Cluster 1 (C-1) dataset has 4449 farmers, and all the farmers in this cluster cultivate CoC(Sc)22, TNAUSCSi7, and TNAUSCSi8 varieties of sugarcane. Cluster 2 (C-2) had 2299 farmers, Cluster 3 (C-3) had 2324 farmers, and Cluster 4 (C-4) had 927 farmers, and all are cultivating CoC(SC)22, CoC90063, TNAUSCSi7, and TNAUSCSi8 varieties of sugarcane.

The proposed NN is implemented in Python Language and is used in this study to construct the NN Tensorflow_2.3.0 version and Keras. All four datasets were partitioned into 80% and 20% training and testing set data. Relu is utilized as the activation function for the HL and IL. Softmax is used as the OL's activation function, and the built model is compiled using the Categorical_CrossEntropy method. The NN built for the C-2 dataset is given below in Figure 7.

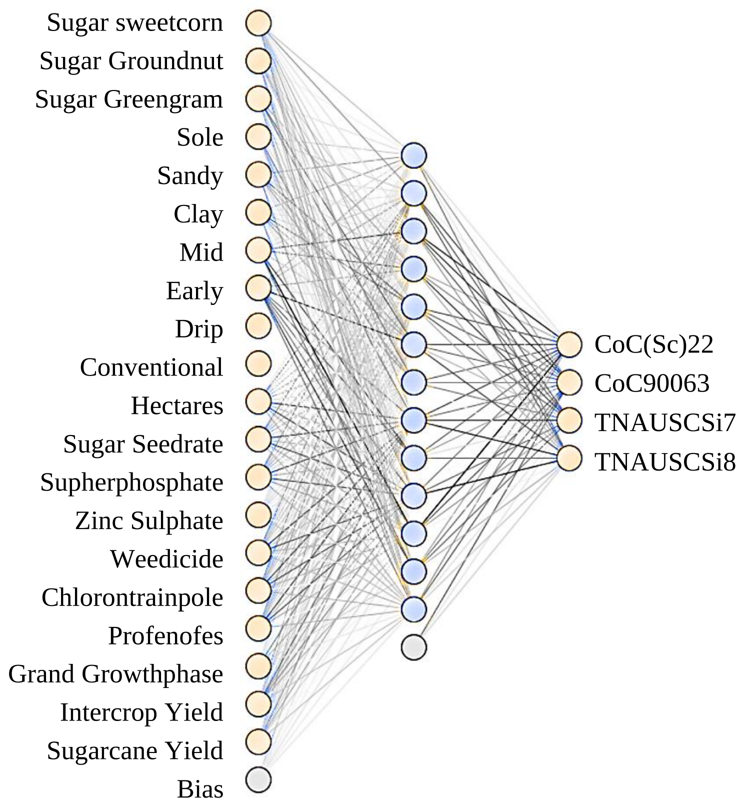

Fig. 7 NN constructed for C-2 dataset

From Figure 7, it is clear that the input variables, which are highlighted by dark lines, have a high impact in classifying the target class of the NN. The sugarcane varieties CoC(SC)22, CoC90063, TNAUSCSi7, and TNAUSCSi8 planted in the Mid and Early season gives a high yield.

The NN was constructed for each dataset. The Confusion Matrix obtained after building the NN for each cluster dataset is given below in the following figures.

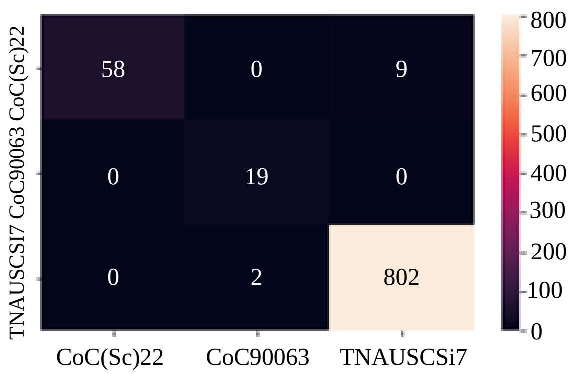

Fig. 8 Confusion matrix for C-1 dataset

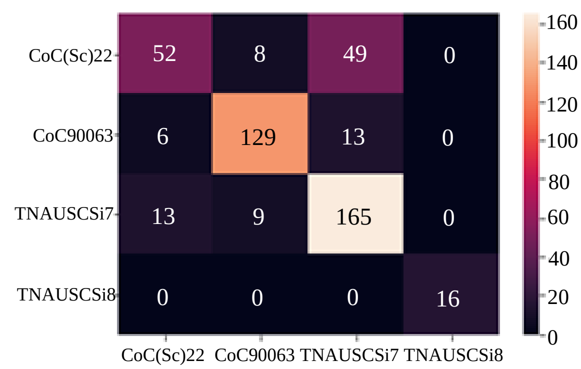

Fig. 9 Confusion matrix for C-2 dataset

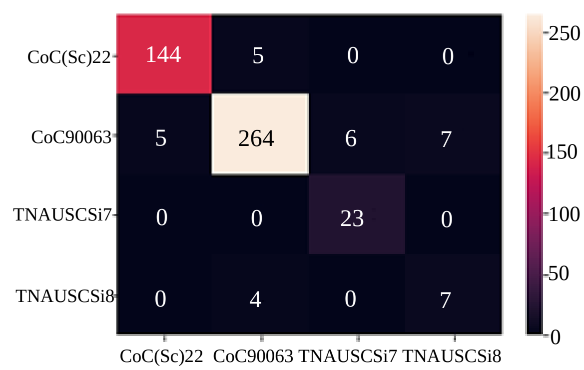

Fig. 10 Confusion matrix for C-3 dataset

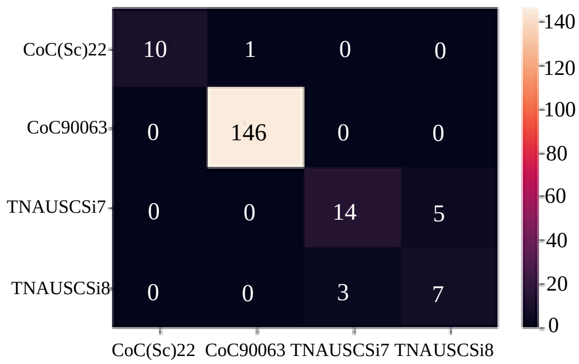

Fig. 11 Confusion matrix for C-4 dataset

The performance of the Back Propagation Algorithm was measured using the performance metrics calculated with the help of the confusion matrix, and it is listed below in the following Table 2.

Table 2 shows that the ANN classifier performed well on the C-1 dataset with a high accuracy 0.991. Next, the ANN classifier gives good results in classifying the C-4 dataset with an accuracy of 0.975. C-3 dataset comes third with an accuracy of 0.970, and the C-2 dataset comes fourth with an accuracy of 0.896.

Table 2. Performance comparison for ANN constructed for the four clustered SD

	ANN-Cluster1	ANN-Cluster 2	ANN-Cluster3	ANN-Cluster 4
Time	13.48 Secs	4.52 Secs	4.38 Secs	2.14 Secs
Recall	0.954	0.807	0.884	0.836
Specificity	0.964	0.915	0.976	0.982
Precision	0.964	0.835	0.806	0.849
Negative Predicted Value	0.987	0.924	0.970	0.986
False Positive Rate	0.033	0.083	0.023	0.017
False Negative Rate	0.045	0.191	0.114	0.163
False Discovery Rate	0.035	0.164	0.192	0.149
Accuracy	0.991	0.893	0.970	0.975
Error Rate	0.009	0.107	0.029	0.25
Number of Farmers	4449	2299	2324	927
Variety of Sugarcane Cultivated by Farmers of Each Cluster	CoC(Sc)22, TNAUSCSi7, TNAUSCSi8	CoC(Sc)22, CoC90063, TNAUSCSi7, TNAUSCSi8	CoC(Sc)22, CoC90063, TNAUSCSi7, TNAUSCSi8	CoC(Sc)22, CoC90063, TNAUSCSi7, TNAUSCSi8

6. Discussion

The consolidated results of this research show that pest management control plays a significant part in improving sugarcane production. Applying Chlorantraniliprole at the germination phase helps control shoot borer and Top Borer diseases caused in the sugarcane buds and helps to grow the sugarcane buds healthily. Selecting a suitable number of seed rates for each variety and planting them with proper seed treatment methods assists in enhancing sugarcane cultivation. Using more sugarcane seeds gives a high number of millable canes and gives high income to farmers. Practising Intercropping method in sugarcane cultivation helps control the weeds and effectively uses the 90cm space between two rows of sugarcane buds.

Intercropping gives additional income to farmers because these intercrops use sugarcane resources to grow, providing additional income and employment opportunities to farmers. The application of Gypsum increases soil fertility by improving the mineral contents such as Ca^{2+}, Mg^{2+}, SO_4^{2-}. This mineral content overcomes the impacts of water stress caused by rainless periods. It provides nutrition to the sugarcane buds' roots and helps grow the sugarcane crop healthily.

The selected sugarcane varieties CoC(Sc)22, CoC90063, TNAUSCSi7 and TNAUSCSi8 gives the best results in both the Conventional and Drip Irrigation Method. These varieties also support Intercropping as well as sole sugarcane production. The four selected varieties are best when cultivated during a particular season. The micronutrients such as Ferrous Sulphate and Zinc Sulphate applied to both Sand and Clay soil increased the iron content of the sugarcane leaf. This helps to improve the sucrose content and juice level of the sugarcane.

6.1. Limitations

Throughout India, sugarcane is grown in several ways. Only four sugarcane varieties have been studied in this research. The research's geographical focus was restricted to 10000 farmers for only two districts. The proposed method was tested on 10,000 records, but it has to be tested on much more significant amounts of data, and the algorithm's effectiveness needs to be assessed. Exactly 25 historical dataset-derived variables were used in this research. IOT sensors are required to capture parameter values in real-time. A real-time dataset was created with the sensor readings, and the proposed algorithms were applied.

7. Conclusion and Future Work

This study proposed a Hybrid Machine Learning Model (HMLM) that reduces the dimensionality of the dataset using Backward Feature Elimination. The proposed model used K-MC to classify the farmers with higher similarity and partition the dataset into four individual datasets. ANN is built individually for each dataset, and the factors responsible for the growth of the sugarcane varieties, such as CoC(Sc)22, CoC90063, TNAUSCSi7 and TNAUSCSi8, were identified. The classifier performance was enhanced by adjusting the classifier's parameters, and the results were compared using the performance metrics. A machine learning algorithm in sugarcane cultivation performed well in analyzing the dataset and correctly identifying the factors responsible for sugarcane growth.

In future, the designed machine learning model is recommended to use in various domains such as Medical, Educational and Finance Sectors. The researchers also planned to analyze the study of crops such as cashews and black gram using the proposed method in future.

References

[1] Mohammad Hajeb et al., "Simultaneous Retrieval of Sugarcane Variables from Sentinel-2 Data using Bayesian Regularized Neural Network," *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103168, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Athiraja Atheeswaran et al., "Expert System for Smart Farming for Diagnosing Sugarcane Diseases using Machine Learning," *Computers and Electrical Engineering*, vol. 109, p. 108739, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Anuj Rapaka, and A. Clara Kanmani, "An Optimized Hyper Parameter Tuned Convolution Neural Frame for Potato Leaves Disease Prediction," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 2, pp. 180-195, 2023. [CrossRef] [Publisher Link]

[4] S. K. Tyagi et al., "Technological Advancements in Jaggery-Making Processes and Emission Reduction Potential via Clean Combustion for Sustainable Jaggery Production: An Overview," *Journal of Environmental Management*, vol. 301, p. 113792, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mahima Begum, Dhiman Dev Singha, and Bijnan Chandra Bordoloi, "Trend of Sugarcane and Jaggery Production in Assam and Associated Problems and Prospects," *SSRG International Journal of Agriculture & Environmental Science*, vol. 3, no. 6, pp. 13-20, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[6] A. Narayanamoorthy, *The Irrigation Future of India: Overview and Synthesis*, Global Issues in Water Policy, vol. 29, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Shyamal S. Virnodkar et al., "CaneSat Dataset to Leverage Convolutional Neural Networks for Sugarcane Classification from Sentinel-2," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3343-3355, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Marife Kung Villareal, and Alejandro Fernandez Tongco, "Sugarcane Classification using Spectral Signature and Object-Based Image Analysis (OBIA) in LiDAR Data Sets," *SSRG International Journal of Agriculture & Environmental Science*, vol. 6, no. 4, pp. 9-16, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ali Kaab et al., "Combined Life Cycle Assessment and Artificial Intelligence for Prediction of Output Energy and Environmental Impacts of Sugarcane Production," *Science of the Total Environment*, vol. 664, pp. 1005-1019, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ana Cláudia dos Santos Luciano et al., "Empirical Model for Forecasting Sugarcane Yield on a Local Scale in Brazil using Landsat Imagery and Random Forest Algorithm," *Computers and Electronics in Agriculture*, vol. 184, p. 106063, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ate Poortinga et al., “Mapping Sugarcane in Thailand using Transfer Learning, a Lightweight Convolutional Neural Network, NICFI High Resolution Satellite Imagery and Google Earth Engine,” *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 1, p. 100003, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Lidan Meng et al., “Understanding the Pathways for Irreversible Aggregate Clusters Formation in Concentrated Sugarcane Juice Derived from the Membrane Clarification Process,” *LWT*, vol. 151, p.112204, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Pengwen Wang, Behzad Aalipur Hafshejani, and Daluyo Wang, “An Improved Multilayer Perceptron Approach for Detecting Sugarcane Yield Production in IoT-Based Smart Agriculture,” *Microprocessors and Microsystems*, vol. 82, p. 103822, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Anteneh Agezew Melash et al., “Indigenous Agricultural Knowledge: A Neglected Human Based Resource for Sustainable Crop Protection and Production,” *Heliyon*, vol. 9, no. 1, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] S. Divya Meena et al., “Crop Yield Improvement with Weeds, Pest and Disease Detection,” *Procedia Computer Science*, vol. 218, pp. 2369-2382, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] S. Thirumal, and R. Latha, “Teaching and Learning based Optimization with Deep Learning Model for Rice Crop Yield Prediction,” *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 4, pp. 105-114, 2023. [CrossRef] [Publisher Link]

[17] D. Sheema et al., “The Detection and Identification of Pest-FAW Infestation in Maize Crops using Iot-Based Deep-Learning Algorithm,” *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 12, pp. 180-188, 2022. [CrossRef] [Publisher Link]

[18] Yongqiang Wang et al., “Multiobjective Optimization of Regional Irrigation and Nitrogen Schedules using the CERES-Maize Model with Crop Parameters Determined from the Remotely Sensed Leaf Area Index,” *Agricultural Water Management*, vol. 286, p. 108386, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] K. Archana, and K. G. Saranya, “Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier,” *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 5, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Jasmin Praful Bharadiya, Nikolaos Tzenios Tzenios, and Manjunath Reddy, “Forecasting of Crop Yield using Remote Sensing Data, Agrarian Factors and Machine Learning Approaches,” *Journal of Engineering Research and Reports*, vol. 24, no. 12, pp. 29-44, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ryoya Tanabe, Tsutomu Matsui, and Takashi S T Tanaka, “Winter Wheat Yield Prediction using Convolutional Neural Networks and UAV-Based Multispectral Imagery,” *Field Crops Research*, vol. 291, p. 108786, 2023. [CrossRef] [Google Scholar] [Publisher Link]