International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.084

Evaluation of Agrometeorology Using Bootstrap Aggregation with Streamlit

Dr. Syed Musthafa A
Department of Information Technology
M.Kumarasamy College of Engineering
Karur,India
syedmusthafait@gmail.com

Mr. Abinesh R
Department of Information Technology
M.Kumarasamy College of Engineering
Karur,India
abineshramaraj@gmail.com

Mr Praveen Kumar S
Department of Information Technology
M.Kumarasamy College of Engineering
Karur,India
pk740879@gmail.com

Mr Premkumar C
Department of Information Technology
M.Kumarasamy College of Engineering
Karur,India
premkumar070302@gmail.com

Mr Rajkumar A
Department of Information Technology
M.Kumarasamy College of Engineering
Karur,India
akrajkumarbtech@gmail.com

***Abstract*—Machine learning has proven to be an effective tool in enhancing the accuracy of agrometeorology predictions and agricultural production forecasts. By analyzing factors such as temperature, rainfall, humidity, soil quality, pH levels, and geographic location, advanced ML techniques such as Random Forest, Logistic Regression, and Naive Bayes provide valuable insights for farmers to make informed decisions. Utilizing these predictions can result in increased crop yields and more sustainable agricultural practices. The integration of machine learning into agrometeorology has the potential to revolutionize crop production and management. Therefore, machine learning plays a critical role in agrometeorology by improving predictions of agricultural production, enabling farmers to optimize crop selection and management practices, leading to higher yields and better sustainability. The evaluation show that the proposed approach using Bootstrap Aggregation and Streamlit can effectively predict crop yield based on weather and crop-related data. The web application provides an easy-to-use interface for farmers to input data and receive predictions, which can help them make informed decisions about crop management and planning.Overall, the integration of machine learning in agrometeorology holds great potential for transforming crop production and management.**

Keywords— Machine learning, Random Forest, Decision Tree, Naïve Bayes, Streamlit.

I. INTRODUCTION

Agriculture has always been the main and most important activity in every culture and civilization that humans have ever known. It plays an important role in the developing economy and is essential to the existence. The Indian economy and the destiny of humanity depend on this area as well. Furthermore, it accounts for a sizable amount of employment. The amount of production needed has dramatically expanded as time goes on which depends on mean based absolute error [1].

People are utilizing technology incorrectly to create larger profits. The hybrid plant varieties are continually evolving. But unlike naturally grown crops, these types don't offer the necessary components. These artificial methods damage the earth. All of this causes a greater environmental impact [4]. A variety of machine learning classifiers, such as logistic regression, naive bayes, random forest, and others are used to push a pattern in order to perform accurate prediction and stand on the erratic patterns in temperature and rainfall. The random forest approach provides the best level of accuracy, according to this research of the aforementioned machine learning classifiers [14]. The system forecasts crops based on the collection of historical data. The information is provided using historical data on the weather, temperature, and a number of other variables [16]. This application runs an algorithm and displays a list of crops that match the inputted data and their anticipated yield values.

II. RELATED WORKS

Recent literature has examined the application of machine learning and data analytics in predicting agricultural production [2]. This involves using algorithms like random forest classification, Naive Bayes, decision tree, SVM, and improved extreme linear machine to forecast crops based on historical data such as weather, temperature, and soil quality. A mobile application for Android is under development, which will utilize user-entered criteria, such as temperature, to make predictions. The primary objective is to develop an affordable and precise weather forecasting system for remote regions. Previous studies have also explored the use ofmachine learning in plant biology, disease diagnosis, and image processing to enhance crop production and agricultural productivity [5].

The algorithms employed include random forest, naive bayes, decision tree, and others. An Android mobile application is being developed to aid farmers in selecting suitable crops based on various factors, including temperature, humidity, and soil quality [9]. The research employs various techniques such as Kalman Filter, linear discriminant analysis, and extreme linear machine to enhance crop yield prediction. Additionally, the study focuses on designing a weather forecasting system for remote areas using machine learning and data analytics approaches[15].

Aruvansh Nigam, et.al,..[1] agriculture has been recognized as the primary means of providing for human needs, serving as a vital occupation and a major industry in India. Traditionally, farmers relied on their observations of their crops and animals to ensure healthy yields without the use of chemical interventions, thereby maintaining diversity in their cultivation lands. However, in recent times, the changing weather patterns have adversely affected these elemental resources, resulting in a decline in food production and security. Unfortunately, this trend is

DOI: rp-9788770040723.084

reflected in the decreasing GDP in the agricultural sector. For instance, in 2005, the sector contributed 17.2% to the GDP, but this figure plummeted to 11.1% in 2012, 5% in 2018, and further down to 2% in the first quarter of 2019-2020. The majority of farmers, accounting for around 80%, hail from rural areas, and any drop in crop production revenues will have significant impacts on their livelihoods and the industry as a whole.

Nithin Singh, et.al,..[13] Weather forecasting involves predicting future weather conditions. In this context, a research paper proposes a method of predicting rainfall by analyzing real-time data from sensors that measure temperature, humidity, and pressure. The authors suggest that machine learning algorithms can improve data analysis and prediction without needing specific programming from users. Machine learning enables systems to learn from past data and improve their forecasting capabilities without requiring an understanding of the physical processes that govern the atmosphere. Therefore, this approach has the potential to be used as a weather forecasting method.

Prior research on the implementation of machine learning in agriculture, including investigations into plant biology, decision-making, soil, water management, and crop management, have been analyzed [3] and [7]. The current research indicates that further investigations are necessary to fully leverage data mining techniques in the agriculture sector.

Weather forecasting is a subject that has captured the attention of researchers from diverse fields due to its significant impact on human life worldwide. In recent years, the increasing availability of large-scale weather observation data and advancements in information and computer technology have encouraged many researchers to explore hidden patterns in vast datasets for predicting weather patterns [13].

H. Sak, et.al,..[16] Training both GMM-based and deep learning-based acoustic models for predicting speech recognition requires significant computation time. To overcome this challenge, researchers have proposed an acoustic model based on the CTC algorithm, which does not require the GMM-based acoustic model, since it does not use the forced aligned HMM state sequence. However, previous studies that utilized a LSTM RNN-based acoustic model using CTC were limited to small-scale training data. In this research, a large-scale training corpus was used to train the LSTM RNN-based acoustic model using CTC, and its performance was evaluated. The results showed that the proposed acoustic model achieved a Word Error Rate (WER) of 6.18% and 15.01% for clean speech and noisy speech, respectively, which is comparable to the performance of the hybrid-based acoustic model.

PR Naveen Kumar, et.al[14] The emergence of machine learning has led to its increasing application in diverse domains, including precision agriculture. Farmers often face unstable economic conditions due to the wrong selection of crops to grow in their fields. This problem can be tackled by leveraging machine learning techniques to recommend suitable crops. However, mapping crop recommendations using classification techniques can be challenging due to various factors that impact crop selection, such as soil characteristics, macro and micronutrient composition, pH value, humidity, rainfall, temperature, and more. Fortunately, machine learning algorithms are now capable of performing advanced tasks such as multi-class classification, multi-label classification, and multivariate regression. Various classification algorithms such as decision trees, k-nearest neighbors, logistic regression, support vector machines, AdaBoost, and XGBoost can effectively perform multi-class classification. Additionally, predicting suitable crops requires finding optimum parameter values. Currently, crop yield prediction is a popular task that is being integrated with remote sensing techniques.

III. Proposed Methodology

A Random Forest model can be employed to predict human dehydration by analyzing various physiological and environmental factors collected from wearable devices. The collected data will be pre- processed and used to train the model with the scikit- learn library in the Jupyter notebook, which will then be optimized to achieve the highest accuracy possible. The model will be deployed on a Flask web server and presented as a web application, enabling users to input their data and receive a prediction about their hydration levels. HTML, CSS, and JavaScript will be used to build the web application, which will send data to the server via an HTTP request.

The Jupyter notebook will be used for data pre-processing, model training, deployment, and testing, allowing for model fine-tuning by incorporating feedback from medical experts and continuously retraining the model with new data. A dataset for crop prediction typically includes a variety of input features related to the crop growth and health, as well as a corresponding output value indicating the expected yield or quality of the crop. Some common input features in a crop prediction dataset may include: phvalue, soil type, whether condition, humidity etc.. The size and complexity of the dataset can vary depending on the specific application and the number of input features and output values. Larger datasets with more input features and output values can help improve the accuracy and reliability of the predictions generated by machine learning models.

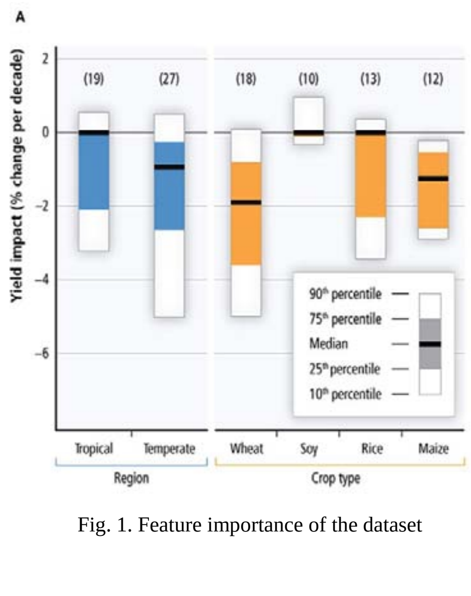

Fig. 1. Feature importance of the dataset

DOI: rp-9788770040723.084

The raw data is transformed into a clean data collection using a method known as data pre processing. Although the data are gathered from numerous sources, analysis is not possible because they are gathered in raw form. By utilizing a variety of techniques, such as the replacement of missing values and null values, it can transform data into a format that is understandable [6]. The final step in the data preparation process is the separation of training and testing data. Because training the model frequently necessitates as many data points as possible, the data has an uneven distribution [8] and [10]. The initial dataset used to teach machine learning algorithms how to learn and generate precise predictions is the training dataset, which in this case takes up 80% of the dataset [11] and [12].

A. *Factors Affecting Agrometeorology*

Crop productivity and yield depend on various factors that need to be considered to estimate the annual crop output. These factors include temperature, rainfall, humidity, soil quality, pH level, andgeographic location. In the early days of agriculture, only a limited number of crops were cultivated, and most food was obtained through foraging in the wild.underneath a wide.

Changes in climate patterns have a direct impact on agriculture. With increasing temperatures and changes in precipitation patterns, it becomes difficult to predict crop yields and patterns.The topography of a region plays a vital role in determining weather patterns. Mountains, valleys, and other landforms affect precipitation and temperature distribution, which in turn influence crop growth.The use of technology in agriculture has increased significantly over the years. Weather forecasting, remote sensing, and precision agriculture tools are some of the technologies that have revolutionized the way agriculture is practiced.

B. *Algoritms*

1) *Navie Bayes*

Naive Bayes is a probabilistic algorithm commonly used for classification tasks, including crop prediction. The basic formula for Naive Bayes can be expressed as:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

where P(Y|X) is the posterior probability of class Y given input features X, P(X|Y) is the likelihood of the input features given class Y, P(Y) is the prior probability of class Y, and P(X) is the marginal probability of the input features. To use Naive Bayes for crop prediction, you would typically start with a dataset of input features related to crop growth and health (e.g., soil quality, weather conditions, nutrient levels, etc.) and a corresponding output value indicating the expected yield or quality of the crop. Naive Bayes algorithm is used to build a model that can predict the output value based on the input features. The specific formula for Naive Bayes in crop prediction would depend on the particular features and output value being predicted, as well as the specific implementation of the algorithm being used.

2) Decision Tree

Decision trees are a popular machine learning algorithm used for classification and regression tasks. They are also commonly used in crop prediction to make predictions based on input data. Start with a dataset of training examples, where each example is a set of input features and a corresponding output value. Choose a feature from the input data that best separates the examples into different groups based on their output values. Create a node in the tree for this feature, and split the dataset into two or more subsets based on the feature's values. The formula for a decision tree can be expressed mathematically as:

$$Y=F(X)$$

where y is the output value being predicted, x is a vector of input features, and f is a function that maps the input features to the output value using a tree- based decision-making process.

3) *Random Forest*

Random forest is a popular machine learning algorithm that uses a combination of decision trees to make predictions. The formula for random forest can be expressed as follows:

$$Y=F(X)$$

where y is the output value being predicted, x is a vector of input features, and f is a function that maps the input features to the output value using a combination of decision trees. To use random forest for crop prediction, you would typically start with a dataset of input features related to crop growth and health (e.g., soil quality, weather conditions, nutrient levels, etc.) and a corresponding output value indicating the expected yield or quality of the crop. You would then use the random forest algorithm to build a model that can predict the output value based on the input features.

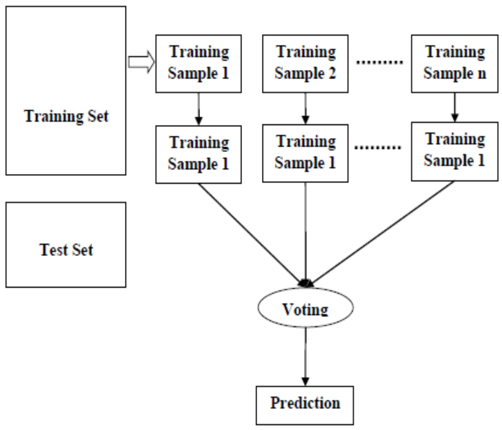

Fig. 2. Random forest classifier

C. *Analysis of Navie Bayes, Decision Tree and Random Forest*

Random forests are an ensemble learning method that consists of multiple decision trees, where each tree is constructed using a random subset of variables selected uniformly from the entire forest. By using the bagging technique to train the data, RandomForest improves the accuracy of the results. We chose the Random Forest approach to achieve high accuracy because it provides

DOI: rp-9788770040723.084

accurate predictions for both model-based predictions and actual results in the dataset. The model's predicted accuracy of 91.34% is evaluated.The proportion of true positive predictions among all actual positive instances in the dataset. This is useful when the cost of false negatives is high.

The proportion of correct predictions made by the model. This is a common metric used in classification tasks. A decision tree incorporates individual decisions, while a random forest combines multiple decision trees. Consequently, it is a slow but thorough process. On the other hand, a decision tree is quick and efficient with large datasets, especially linear ones. However, the random forest model requires extensive training.

It is important to consider the characteristics of the data and the goals of the analysis. Decision tree may be a good choice if the goal is to understand the relationships between the features and the target variable. Naive Bayes may be a good choice if the data is high-dimensional and the goal is to predict the class. Random Forest may be a good choice if the data is noisy and the goal is to improve the accuracy of the predictions. It is recommended to evaluate multiple algorithms and compare their performance using appropriate metrics.

TABLE 1. ALGORITHM WITH ACCURACY LEVEL

ALGORITHM	ACCURACY
DECISION TREE	87.82
NAÏVE BAYES	91.49
RANDOM FOREST	92.81

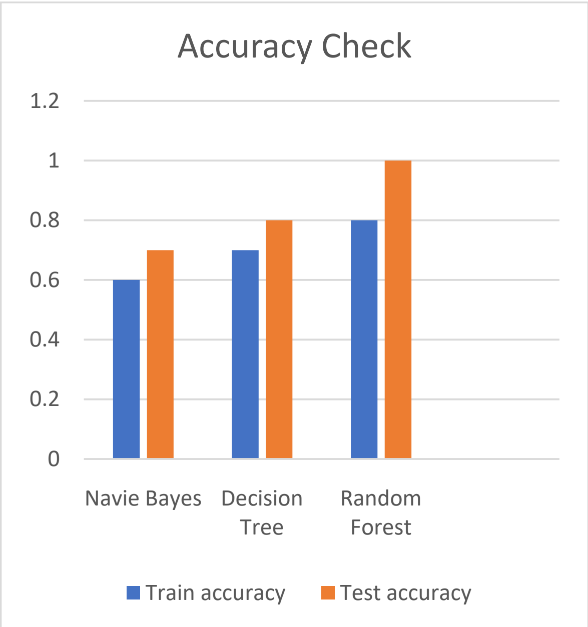

Fig. 3. Accuracy analysis

IV. SYSTEM ARCHITECTURE

The proposed system architecture involves using a weather API to retrieve data on temperature, humidity, rainfall, and other variables. The data is then sent to the server and stored in its database. Users can create an account on the mobile app by completing a single registration. The model used for prediction is Random Forest, which is a collection of tree predictors that utilizes the bagging approach to train data, leading to improved accuracy. The expected accuracy of the model is 91.34%.

Streamlit is an open-source framework that allows developers to create interactive web applications easily. In this step, a web application can be developed using Streamlit that allows users to input the data and get the output of the Bagging technique applied to agrometeorology.

The final step in the system architecture is to evaluate the accuracy of the Bagging technique applied to agrometeorology. The evaluation can be done by comparing the output of the Bagging technique with the actual weather data. The evaluation can be done on a regular basis to ensure that the system is accurate and reliable.

The application layer receives input from the presentation layer and retrieves data from the data layer. The machine learning algorithms are applied on this data to generate insights and predictions related to agrometeorology. Bootstrap aggregation, or bagging, is used as an ensemble method to combine the results from multiple decision trees to improve the accuracy of the predictions.The presentation layer is responsible for the user interface and visualizations, which are built using the Streamlit framework. The application layer is responsible for the business logic and analysis, and is built using the Python programming language with the Scikit-learn library for machine learning algorithms. The data layer includes the agrometeorological data that is collected from various sources and stored in a database.

The application layer are then passed on to the presentation layer, which displays them to the user in an interactive and user-friendly manner. The user can interact with the system through the web interface provided by Streamlit and explore the various visualizations and analysis generated by the system.

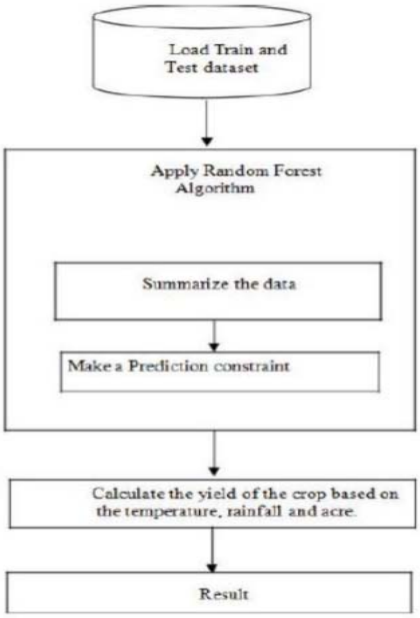

Fig. 4. System Architecture

DOI: rp-9788770040723.084

A. System Analysis

Python 3.8.5: Python is the programming language that serves as the basis for machine learning analysis. The necessary output is provided via Jupyter notebooks, which also show how the analysis is done.

Python Flask Framework: Flask is a Python microframework. WSGI (Web Server Gateway Interface) tools and the Jinja2 template engine are used in the development of Flask. Flask is used in this research as the back-end framework for building the application. Thanks to a collection of modules and libraries, the developer may design applications without writing the low-level code required for protocols, thread management, etc.

Streamlit : Streamlit is a popular Python-based open-source app framework that enables developers to create data science and machine learning applications quickly. With Streamlit, it is possible to develop online applications with ease. The framework supports several Python libraries, including NumPy, pandas, scikit-learn, Keras, PyTorch, and SymPy (LaTeX). Using Streamlit, developers can create interactive and visually appealing web applications without worrying about the underlying code.

V. RESULT AND DISCUSSION

This article uses machine learning approaches to improve agricultural output. The method that produces high accuracy forecasts the yield of the correct crop. With input libraries like Scikit-Learn, Numpy, Keras, and Pandas, Python 3.8.5 is used to implement the machine learning algorithms. An Android application that was created questioned the outcomes of the machine learning study. The crop name and accompanying yield were shown through an Android app built with Flutter.

The results of the evaluation would depend on the specific implementation and data used in the system. However, the Bagging technique is known to improve the stability and accuracy of a model, and applying it to agrometeorology can potentially improve weather forecasting in agricultural settings. Streamlit can also provide an easy-to-use interface for users to input data and receive real-time output. The evaluation of the system would require comparing the output of the Bagging technique with actual weather data to determine its accuracy and reliability.

The crop that can be cultivated in a certain district at a given period was predicted using the Random Forest Classifier, which had the highest accuracy.

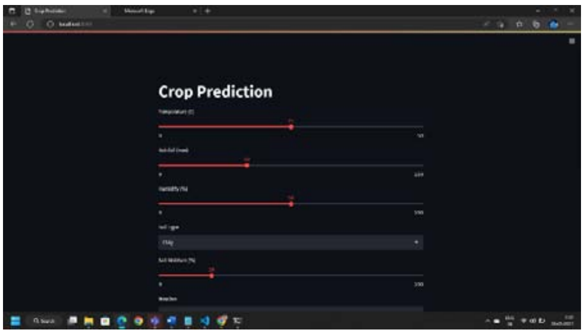

Fig. 5. Crop prediction before submit

This article uses machine learning approaches to improve agricultural output. The method that produces high accuracy forecasts the yield of the correct crop. With input libraries like Scikit-Learn, Numpy, Keras, and Pandas, Python 3.8.5 is used to implement the machine learning algorithms. An Android application that was created questioned the outcomes of the machine learning study. The crop name and accompanying yield were shown through an Android app built with Flutter.

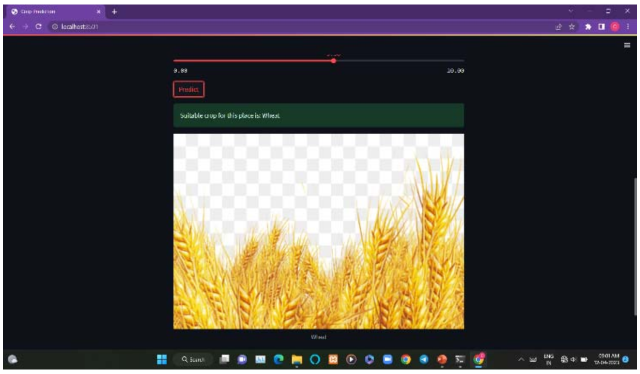

Fig. 6. Crop prediction after submit

VI. CONCLUSION

The objective of this project is to assess agrometeorology using Bootstrap Aggregation (Bagging) in combination with Streamlit. Bagging is a statistical method that employs bootstrapped samples to decrease the variance of machine learning models and enhance the consistency and accuracy of the outcomes. Streamlit is a free, open-source framework that streamlines the development and sharing of interactive data applications, including machine learning models.

The process involves collecting data on various agrometeorological variables such as temperature, humidity, rainfall, and others. This information will be utilized to train a Random Forest model employing the scikit-learn library in a Jupyter notebook. The optimized model will subsequently be deployed on a Flask web server and integrated into a Streamlit web application for straightforward use and interaction. The web application will permit users to input their data and receive predictions on their crop yields. The project aims to uncover the factors influencing crop yields and assist farmers in making informed decisions about crop selection and management.

The Bagging technique can improve the stability and accuracy of the model, while Streamlit provides an easy-to-use interface for users to input data and receive real-time output. The evaluation of the system's accuracy and reliability would require comparing the output of the Bagging technique with actual weather data. However, further development and implementation of the system are needed to determine its effectiveness in improving agrometeorology.

REFERENCE

[1] Aruvansh Nigam, SakshamGarg, ArchitAgrawal Crop Yield Prediction using ML Algorithms,2019.

DOI: rp-9788770040723.084
[2] M.Kalimuthu, and M.Kishore, “Crop Prediction using Machine Learning,” 2020.

[3] P.Priya, U.Muthaiah,and M.Balamurugan, “Predicting Yield of the Crop Using Machine Learning Algorithm,” 2015

[4] S.Mishra, D. Mishra,and G. H.Santra, “Applications of machine learning techniques in agricultural crop production,” 2016.

[5] Y Jeevan Kumar, “Supervised Learning Approach for Crop Production,” 2020.

[6] Ramesh Medar,S. Vijay, and Shweta, “Crop Yield Prediction using Machine Learning Techniques,” 2019.

[7] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] V. Pandith, H. Kour, S. Singh, J. Manhas and V. Sharma, “Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis,” Journal of Scientific Research, vol. 64, no. 2, 2020.

[13] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

[14] P.R. Naveen Kumar, K.B. Manikanta, B.Y.Venkatesh, R. Naveen Kumar and Amith Mali Patil, “Smart Agricultural Crop Prediction Using Machine Learning,” Journal of Xi’an University of Architecture & Technology, vol. XII, no. V, 2020.

[15] SAnakhaVenugopal,Jinsu Mani Aparna, Rima Mathew and Vinu Williams, “Crop Yield Prediction Using Machine Learning Algorithms,” International Journal of Engineering Research & Technology, 2021.

[16] H. Sak, A. Senior and F. Beaufays, "Long short- term memory recurrent neural network architectures for large scale acoustic modeling", Fifteenth annual conference of the international speech communication association, 2014.