Fuzzy Systems and Soft Computing
ISSN : 1819-4362

PERFORMANCE EVALUATION OF SUPERVISED MACHINE LEARNING METHODS ON RICE YIELD PREDICTION IN INDIA: A REVIEW

Pankaj Bhattacharjya Asstt. Professor, Dept. of Computer Science, Dimoria College, Khetri & Research Scholar, Dept. of Computer Science, University of Science & Technology, Meghalaya(USTM) Email: bhattacharjya.p@gmail.com
Dr. Kanak Chandra Bora Associate Professor, Dept. of Computer Science University of Science & Technology, Meghalaya(USTM) Email: boopborabora@yahoo.in

Abstract
Most of the country's economy is based on agriculture and agricultural products have been the foremost factor for our survival. Rice is one of the major crops of India and many other countries. Predicting Rice yield for the future is helpful for farmers and policymakers of a country for future planning. Machine Learning techniques can predict rice yield based on existing datasets like soil nutrients, weather data, and previous years' rice production. Supervised Machine Learning techniques like Random Forest (RF), Support Vector Machine(SVM), k-nearest Neighbors(k-NN), and XG Boost are widely used in predicting rice yield. The Performance of the models is generally performed by Coefficient of Determination (R2), Root Mean Square Error (RSME), and Mean Absolute Error (MAE). A survey is done on Supervised Machine Learning Methods carried forward by researchers for the prediction of crop yield across India along with the best performances of the models to help the researchers of this field for future research work.
Keywords: Rice Yield, RF, SVM, k-NN, XG Boos, RSME, MAE

Introduction

Agriculture has been the main livelihood for the maximum population across the countries and India in particular is also based on the agriculture economy. Indian agriculture is known for its diversity which mainly results in variation in resources and climate. India's economy is mainly dependent on agriculture. Rice is a staple food for a significant portion of the world's population, particularly in Asia. Globally, countries like China, India, Indonesia, Bangladesh, and Vietnam are among the top producers of rice. India is one of the top producers of rice with nearly 30 Percent of rice produced in India and plays a major role in global rice production.

India's economy is mainly dependent on agriculture. Ramisetty, S et al(2024) [1] observe that due to fluctuations in temperature, soil conditions, and agricultural practices, variations in rice production can affect both food security and economic stability. Making agriculture sustainable and resilient to the ongoing change in climate and social structure is a major challenge for scientists and researchers across the globe. The agricultural system demands transition and a multidisciplinary approach. According to Maya Gopal, P.S. (2020)[2] the field of machine learning (ML) has seen an expansion in the number of applications because of advances in computing. Due to its effectiveness in a variety of areas such as forecasting, defect detection, pattern recognition, and so on, machine learning is becoming more frequently employed throughout the world. Yield prediction is a major agricultural challenge. Farmers will be able to predict the yield of their crop using the results of this study before growing it in the cropland, allowing them to make informed decisions. Timely instructions to anticipate future crop output and analysis are essential to help farmers in maximizing agricultural yield.

In India, rice cultivation is a crucial agricultural activity, with a diverse range of rice varieties grown across different regions. Some of the key rice-producing states in India include West Bengal, Uttar Pradesh, Punjab, Andhra Pradesh, Bihar, and Telangana. The country has made significant advancements in rice production through the adoption of improved farming practices like sophisticated machinery, use of fertilizers, irrigation techniques, use of high-yielding rice varieties, alteration of paddy fields, etc.

Rice yields in India can vary based on factors such as weather conditions, water availability, soil quality, use of fertilizers and pesticides, and government policies. The government of India has been

implementing various agricultural schemes and programs to promote rice cultivation and increase productivity in the sector.

Overall, the global rice yield is influenced by various factors such as climate change, technological advancements, market demands, and government policies. Continuous efforts are being made to improve rice production and ensure food security for the growing population. Climate change has a significant impact on rice yield worldwide. Raising in temperature due to Climate can affect various stages of rice cultivation which in turn may affect the production of rice. Climate change has also caused unpredictable rainfall which causes either drought or flood which is affecting both quality and yield.

In coastal regions where rice is cultivated, sea level rise can lead to saltwater intrusion into rice fields, rendering the soil unsuitable for cultivation. The salinization of soil can reduce rice yield and productivity. Carbon dioxide in the atmosphere has been elevated which leads to long-term effects on rice yield.

Efforts to mitigate the impact of climate change on rice yield include developing climate-resilient rice varieties, improving water management practices, adopting sustainable agricultural techniques, and implementing policies to reduce greenhouse gas emissions. Adapting to changing climatic conditions is crucial to ensuring food security and sustainable rice production in the face of climate change.

Rice yield plays a crucial role in policy-making, estimation, and marketing in the agricultural sector. Here are some benefits of rice yield for these purposes:

Utilizing the rice yield data, policymakers make informed decisions on promoting food security by ensuring an adequate supply of rice. Based on rice yield data, policymakers design agricultural policies related to subsidies, pricing, land use, water management, and technology adoption to enhance rice production. Rice yield data is essential for formulating trade policies, including import/export regulations, trade agreements, and market interventions to maintain a balance between domestic supply and demand.

Accurate estimation of rice yield enables stakeholders to forecast production levels, plan distribution, and manage resources efficiently. By analyzing historical yield data and trends, stakeholders can assess production risks, such as climate variability, pests, and diseases, and implement mitigation strategies.

Marketing strategies can be made based on rice yield information. Prices can be fixed by signaling the level of supply available. Thus, farmers, traders, and consumers can make decisions on buying, selling, and pricing. Understanding rice yield patterns helps in planning market strategies, identifying potential consumers, managing inventory, and ensuring market competitiveness. Monitoring rice yield aids in maintaining quality standards, grading criteria, and certification processes for marketing purposes.

By utilizing rice yield data effectively, policymakers, stakeholders, and market participants can enhance decision-making, promote sustainable agricultural practices, and optimize marketing strategies in the rice industry.

Machine learning techniques can indeed help predict rice yield. By analyzing historical data on various factors such as weather patterns, soil conditions, agricultural practices, fertilization, pest control, and crop management, machine learning models can be trained to predict rice yield accurately. Some ways in which machine learning can assist in predicting rice yield include

1. Regression Models: Machine learning regression algorithms can be employed to predict the expected rice yield based on input variables such as weather conditions, soil quality, and farming practices.
2. Time Series Analysis: By analyzing time-series data related to rice cultivation, machine learning models can forecast yield trends over time and predict future yields based on historical patterns.
3. Image Recognition: Machine learning algorithms can analyze satellite images or drone data to assess crop health, detect diseases or pest infestations, and predict yield based on visual cues and patterns.
4. Predictive Analytics: By leveraging machine learning algorithms, predictive models can be built to estimate rice yield under different scenarios, helping farmers and policymakers make informed decisions.

5. Optimization Algorithms: Machine learning optimization techniques can be used to maximize rice yield by optimizing planting densities, irrigation schedules, fertilization strategies, and other factors that influence crop productivity.

Overall, machine learning offers a powerful toolset to analyze complex data patterns, identify key variables affecting rice yield, and develop predictive models to enhance decision-making in agriculture. By harnessing the capabilities of machine learning, stakeholders can improve crop management practices, optimize resource allocation, and ultimately boost rice yield.

Various researchers in India and across the globe have been studying constantly based on historical data and relevant features. Some of such methods that give optimum solutions to rice yield predictions are:

1. Random Forest:

- Random Forest is a versatile ensemble learning method derived from Decision Trees that can handle a large number of input features and noisy data effectively.
- It is robust against overfitting and can capture complex relationships between input variables and target outputs, making it suitable for predicting rice yield accurately.

2. Support Vector Machines (SVM):

- SVMs are powerful classifiers based on the concept of 'hyperplane ' that can also be used for regression tasks like predicting crop yield.
- They work well in high-dimensional spaces and are effective when the number of features exceeds the number of samples.
- SVMs can capture non-linear relationships in the data and generalize well to unseen data.

3. Gradient Boosting:

- Gradient Boosting algorithms like XGBoost, LightGBM, or CatBoost are popular for regression tasks and can be applied to predict rice yield.
- These algorithms work by building multiple weak learners sequentially, each correcting the errors of its predecessor, leading to highly accurate predictions.

4. Neural Networks:

- Deep learning techniques, such as neural networks, can also be used to predict rice yield.
- Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks are particularly useful for analyzing time-series data, which can be beneficial in predicting crop yield trends.

5. K-Nearest Neighbors (KNN):

- KNN is a simple and intuitive algorithm that can be effective for predicting rice yield, especially in cases where the relationships between features and output are not linear.
- It makes predictions based on the similarity of input data points to the training samples.

These supervised machine-learning techniques can be applied to historical data on rice yield, environmental factors, and other relevant variables to build predictive models for estimating rice yield. The choice of the best technique may depend on the specific characteristics of the dataset, the complexity of relationships, and the interpretability of the model required for the prediction task.

A literature survey on the prediction of rice yield using supervised machine learning methods can provide insights into current research trends, methodologies, and findings. Below is an overview of key points from recent studies in this area:

Key Areas of Focus

1. Data Sources:

- Remote Sensing Data: Utilizes satellite imagery and UAV data for features like NDVI (Normalized Difference Vegetation Index), soil moisture, and temperature.
- Weather Data: Includes parameters such as rainfall, temperature, cloud cover, and humidity.
- Soil Data: Features like soil type, pH, and nutrient levels (Nitrogen, Potassium, Phosphorous)
- Historical Yield Data: Past yield records for training models.

2. Commonly Used Supervised Learning Algorithms:

- Regression Models: Linear Regression, Support Vector Regression (SVR)
- Tree-based Models: Decision Trees, Random Forests, Gradient Boosting Machines (GBM), and Extreme Gradient Boosting (XGBoost).
- Neural Networks: Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN).
- Ensemble Methods: Combining multiple models to improve prediction accuracy (MAML)

3. Performance Metrics:

- Accuracy: The proportion of correct predictions.
- Mean Absolute Error (MAE): The average of the absolute differences between predicted and actual values.
- Root Mean Squared Error (RMSE): The square root of the average of squared differences between predicted and actual values.
- R-squared (R^2): The proportion of variance in the dependent variable that is predictable from the independent variables.

4. Challenges and Limitations:

- Data Quality and Availability: Incomplete or noisy data can affect model performance.
- Model Overfitting: Especially with complex models like neural networks.
- Scalability: Applying models to larger geographical areas.

Methodology

Review protocol: Before conducting the systematic review, a review protocol is defined. Firstly, the research questions are defined. When research questions are ready, research papers are searched from the open source available on the internet. The machine learning methods used in the rice yield predictions are searched and then from them supervised machine learning methods used in rice yield prediction are selected for review. All the relevant data from the selected studies were extracted, and eventually, the extracted data were synthesized in response to the research questions. The approach can be split up into three parts: plan review, conduct review, and report review (Klompenburga, T.V.)[3] The first stage is planning the review. In this stage, research questions are identified, a protocol is developed, and eventually, the protocol is validated to see if the approach is feasible. The data was extracted, which means that their information regarding authors, title, year of publication, and information regarding the research questions were stored.

Research Questions

1. Which supervised machine learning algorithms have been used in the literature for rice yield prediction?
2. Which features have been used in literature for rice yield prediction using supervised machine learning?
3. Which performance evaluation parameters and evaluation approaches have been used in literature for rice yield prediction to choose the best model?

Importance of present work

Indian agriculture is highly differentiated in terms of its climate, soil, water, crops, horticultural crops, plantation crops, medical crops, livestock, etc. Agriculture is facing the problem of changes in the resources that are directly affecting the crop yield in India and particularly in Assam. Hence the agricultural productivities are unpredictable. For balanced and sustainable growth of agriculture, these resources need to be evaluated, monitored, and analyzed, so that proper methods can be constructed. Accurate and eligible information about crop yield prediction is important for making decisions for agricultural risk management. Crop yield prediction is important for the supply chain operation of companies engaged in industries that use agricultural produce as raw material. An accurate estimate

of crop yield helps the allied companies in planning supply chain decisions like production scheduling and it is useful for businesses such as seed, fertilizer, agrochemical, and agricultural machinery industries and marketing activities based on crop yield.

Machine Learning techniques till now used widely in business and corporate sectors may be used in agriculture for data characterization, discrimination, and predictive and forecasting purposes. Some use of data mining in soil characteristic evaluation has already been attempted.

Machine Learning in agriculture is a relatively novel research field. Agriculture data are highly diversified in terms of nature, interdependency, and use of resources for farming. The major problem of using data mining in agriculture is to solve issues based on the available data and its meaningful outcomes. In data mining, clustering, and classification techniques make ingenious information in research and knowledge acquisition from integrated farming. And that produces better solutions for the farmers about their cultivation yield.

Agriculture is the most important application area, particularly in developing countries like India. The State of Assam is no exception. The use of information technology in agriculture may change the scenario of decision making and farmers can yield in a better way. For decision-making on several issues relating to the agriculture field: data mining plays a vital role. So, the proposed study may help enormously the farmers, and agricultural planners for their planning and decision-making in agricultural production, marketing, etc. in the State of Assam.

Review of literature

Each and every sector in this digital world is undergoing a dramatic change due to the influence of the IT field. But, to date, not much work has been done in the agricultural sector. The use of various machine learning techniques in the agricultural sector will be a continuing area of research. The ultimate goal is to increase the yield of the agricultural sector.

Chandraprabha, M.&.Dhanraj., R.K.(2023)[5] introduced a stacking-based ensemble learning technique termed Model Agnostic Meta Learning (MAML) approach was suggested for forecasting rice crop yield in a particular soil. Soil nutrients and crop production statistics were taken as the input for the proposed method. The input dataset has some missing values, and a preprocessing technique was used to remove the missing values. Normalization and mean of the attribute methods were used to preprocess the dataset. The reprocessed data were given as input for the three different classifiers SVM, CNN, and DBN. The output of these classifiers was given to one Meta-learning technique termed Model Agnostic Meta-Learning (MAML) approach. This MAML classifies the input data and predicts the output as six different classes. The amount of rice crop yield, in particular, is the final prediction. The accuracy rate obtained by the proposed method was 89.5%. A comparison graphical representation proves that the proposed method results are significantly better than those of the existing methods. Thus, the suggested method's ensemble classifier was the best alternative for forecasting rice crop output based on soil nutrient levels.

Vargas, V.W. et. al (October 2022)[6] carried forward a systematic mapping methodology for conducting an evidence-based literature review of research publications addressing preprocessing techniques for imbalanced data in ML. Through the 8-step filtering process domain areas of ML applications such as health, Finance, engineering, software, and biology were studied.-standard or modified clustering-based sampling techniques for balancing data were used. Over-sampling was found to be the most common and best-performing sampling method. They also suggested a hybrid sampling technique that can suppress oversampling in the future. It is also found in the study that classical ML models like SVM, KNN, and LR are the most frequently used but ensemble learning models can give better results. Future work may explore the usage of simulation-based oversampling for balancing data in ML Applications.

The study by Shah, A., Dubey, A. et al.(2018)[7] seeks to develop a strategy for automating agricultural activities and predicting crop production by identifying relationships between yield factors such as temperature and rainfall. Because crop productivity is affected by a variety of environmental factors and may vary under harsh situations, the author ruled out Linear Regression. To anticipate crop yield, they used Multivariate Polynomial Regression, which provided an R-squared value of 0.89 and a Root Mean Square Error (RMSE) value of 9.4. The authors used Random Forest Regression and Support

Vector Machine Regression to improve accuracy. With an RMSE of 5.48 and an R squared value of 0.968, SVM Regression surpassed all trained models and was used in the final approach to predict crop yield.

Panigrahi, B. et al (2023)[8] developed the most effective Machine Learning (ML) model possible for estimating agricultural yields for three different types of crops: Bengal gram, Groundnut, and Maize. It was beneficial for the farmers to take the required measures to boost crop production, which may also be useful in the process of decision-making. To make accurate predictions of the crop yields, the data was trained on six different kinds of alternative regression models, including Linear Regression, Decision Tree Regression, Gradient Boosting Regression, Random Forest Regression, XGboost Regression, and Voting Regression. The values for all 3 metrics: Mean Absolute Error (MAE), R2 score, and Cross-Validation score for all the 6 models mentioned were performed. These criteria are critical in evaluating models and selecting the most suitable model for the final predictions. Random Forest Regression and XGBoost Regression gave better results than the other machine learning algorithms. Based on observation, the XGBoost Regression model was overfitting with a high R2 score and a lower Cross Validation score. Furthermore, with an MAE score of 468.16, MSE score of 825.29, R2 score of 0.7952, and a Cross-Validation score of 0.6087, the Random Forest Regression model outperformed the XGBoost Regression model after parameter tuning.

Devdatta, A..& Bondre, (2019)[13] made soil classification, crop yield prediction, and fertilizer recommendations. For that, they applied two machine learning techniques namely Support vector Machine and Random Forest methods. Soil classification was done using soil nutrient data. The two machine learning algorithms are used for soil classification. The two algorithms classify and display confusion matrix, precision, recall, and average value and at the end accuracy in percentage as output. The accuracy of Random forest in soil classification is found to be 86.35% whereas the accuracy of SVM is 73.75%. So, the random forest method is good for soil classification. Crop yield prediction was done using crop yield data, nutrients, and location data. These inputs are passed into the SVM algorithms. These algorithms predict crops based on present input. The accuracy of the SVM algorithm is 99.47% for yield prediction and RF accuracy is 97.48%. So, for crop yield prediction SVM algorithm is good. Suitable fertilizers for crops and a requird amount of fertilizers were also recommended.

A study by Venugopal, A., Aparna, S., et al.(2021) [9]uses Machine Learning Classification Models to forecast crop output based on temperature, rainfall, and area. Classifier models include Logistic Regression, Naïve Bayes, and Random Forest. Precision farming should emphasize quality above environmental factors. It lists 14 districts in Kerala's crop names, regions, yields, temperatures, rainfall, humidity, and wind speed. The user registers through a mobile app and inputs location and region into the database. Current weather from an open weather map is used to predict yield. Regression model accuracy was 87.8% and Naive-Bayes was 91.58%. Random Forest Classifier's 92.81% accuracy is the highest among classification methods since it uses Bagging. The accuracy gained after using the Random Forest model to evaluate the test data was 91.34%. This study is done so that crops can be grown more efficiently and effectively.

Jiya, E.A. et al (2023)[10] examined the performance of five distinct models for predicting rice yield based on climatic variables. The models were evaluated using both class-based and error-based metrics to assess their accuracy in predicting class instances and numerical values, respectively. The class-based metrics revealed that out of five machine learning algorithms viz. Logistic Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), Random Trees (RT), and Naïve Bayes (NB); Random Forest (RF) and Random Trees (RT) surpassed the other models by achieving a True Positive rate of 1, indicating their precise predictions of rice yield classes. This paper developed a rice yield prediction model for Katsina state in Nigeria, using climatic data and rice yield data. The result of the data analysis reveals that precipitations have no significant relationship with rice output but rather temperature has a closer relationship. Also, of all the 5 models, Random Trees has the highest accuracy in predicting rice yield.

Joshua, V. et al (2021)[11] Predicted crop yield out using statistical and machine learning algorithms. Specifically, the statistical study of likely MLR techniques and machine learning algorithms such as SVM, GRNN, RBFNN, and BPNN were considered for evaluation to attain crop yield prediction of higher accuracy. Model performance metrics are adapted to scrutinize the accuracy level of the

different algorithms. With the observed outcomes, the following conclusions were made: Machine learning algorithms attained exceptionally greater yield prediction accuracy than statistical methodology based on the results of evaluation metrics. Among the four machine learning algorithms such as SVM, RBFNN, GRNN, and BPNN, GRNN predicted the yield more precisely.

Kamath, P. et al.(2021)[12] made region-specific crop yield analysis and it is processed by implementing by random forest algorithm. In this project, they have chosen a dataset in .csv format. For training purposes, 80% of data is used and the remaining 20% of data is used for testing. After the successful training and testing, the next step was finding the accuracy of the model. They have achieved a good accuracy which means this model is good for predicting yield. They have designed the Website which consists of three Functional Modules 1)Crop Module: This module provides the list of available crops. On selection of each one of them gives a detailed description of the crop. 2) Soil Module: This module provides the list of available soils. On selection of each one of them gives a detailed description of the soil. 3) Weather Module: In this module by entering the city name the user can get the live weather forecast.

Taher et al., (2021)[14] used a classification algorithm to create a method that correctly classifies data using the training data set. The soil is the most important aspect of agriculture. The classification of soils based on the nutrients found in the soil, such as potassium, nitrogen, sulfur, phosphorus, iron, zinc, manganese, boron, and copper, as well as its physical properties, such as pH, organic carbon, and electric conductivity, is extremely useful for increasing agricultural production. In the paper, they compared four algorithms such as NB, K-NN, DT, and RF. In comparison to the other four, the K-NN classification algorithm produces a better result for this dataset, correctly classifying the full number of instances. To forecast soil features, they suggested the K-NN algorithm.

Archana, K. (2020)[20]categorized the soil samples according to the soil type, land type, and macronutrients Nitrogen (N), Phosphorus (P), and Potassium (K) present in the soil the suitable crop along with its appropriate fertilizer was suggested to the agricultural stakeholder. The month in which the yield would be high was also suggested to the user. The yield calculation is also provided for the crop selected by the farmer. The proposed crop recommendation system provides 82% accuracy.

Venugopal, A., S, et al(2021)[9] try to reinforce crop production with the aid of machine learning techniques. The technique which results in high accuracy predicts the right crop with its yield. The machine learning algorithms are implemented on Python 3.8.5(Jupyter Notebook) having input libraries such as Scikit Learn, Numpy, Keras, and Pandas. Developed Android application queried the results of machine learning analysis. Flutter-based Android app portrayed crop names and their corresponding yield.

Findings from Literature Review and Discussion:

Table 1: Some Findings from the Literature Reviews:

Name of the Author	Paper Name	Parameters/ features Used	Methods used for Rice Prediction	Methods used for performance evaluation	Performance Evaluation
Wijayanti, E.,B. et.al(2024)[15]	Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting	Area, rice yield for previous years, avg. temperature, avg. rainfall	eXtreme Gradient Boosting	Coefficient of Determination (R2), Root Mean Error(RME),Mean Absolute Error(MAE),	when compared with other methods it gave highest accuracy
Ramisetty,S. & Kumar,M.(2024)[1]	Zone II & III: Machine Learning Based Rice Yield Prediction in Andhra Pradesh	Soil Information (pH Values), humidity (rainfall), Solar Information (Temperature), nutrients(nitrogen, Phosphorus, Potassium), field Management (Urea)	Support Vector Machines, Random Forest, and Gradient Boosting	Mean Squared Error (MSE), Cross Validation using MSE	Random Forest method gives result with highest accracy
Joshua, V. Et. Al(2021)[11]	Exploration of Machine Learning Approaches for Paddy Yield Prediction in Eastern Part of Tamil Nadu	Soil Information (pH Values), humidity (rainfall), Solar Information (Temperature), nutrients(nitrogen, Phosphorus, Potassium), filed Management (Urea)	Support Vector Regression(SVR), General Regression Nural Networks(GRNN),Radial Basis Functional Neural network(RBFNN), Back Propagation Neural Network(BPNN)	Coefficent of Determination (R2), Root Mean Square Error(RSME),Mean Absolute Error(MAE), Mean Absolute Percentage Error(MAPE), Coefficient of Varience(CV), Normalized Mean Squared Error(MMSE)	SVR gives result with highest accuracy
Wilson, A.(2023)[16]	Machine Learning Model for Rice Yield Prediction using KNN Regression	Soil Nutrients, Temperature, Humidity, Precipitation, Crop Yield	Decision Tree Regression, Random Forest Regression, Linear Regression,K Nearest Neighbour Regression,XGboost Regression and Support Vector Regression	No mention of method for accuracy determination	KNN with highest accuracy (98.77%)
Jiya, E.A. et al (2023)[10]	Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms	Climate Data like elevation, max temperature, min temperature, wind, relative humidity, and Crop Data (Yield/metric tons)	Logistic Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), Random Trees (RT), and Naïve Bayes (NB)	MAE, RMSE, RAE, and RRSE	Random Forest and Random Tree(RT) highest accuracy in prediction
Dahikar ,S.S. &.Rode.,V.(2014)[17]	Agricultural Crop Yield Prediction Using Artificial Neural Network Approach	Soil type, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, and humidity	ANN	No mention of method for accuracy determination	ANN

Venugopal, A. et al.(2021)[9]	Crop Yield Prediction using Machine Learning Algorithms	crop names, regions, yields, temperatures, rainfall, humidity, and wind speed	Logistic Regression model, Naive-Bayes, Random Forest Classifier	No mention of method for accuracy determination	Random Forest Classifier with 92.81% accuracy
Panigrahi, B. et.al(2023)[8]	A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models	water, seeds, fertilizers, pesticides, and insecticides, previous years' yield	Linear Regression, Decision Tree Regression, Gradient Boosting Regression, Random Forest Regression, Xgboost Regression, And Voting Regression	Mean absolute error,Mean squared error, R2 score, Cross-validation score	XG Boost Model and Random Forest Reggression gives the best result but Random Forest Regressor is best even on unkown data
Chandraprabha, M. et al (2020)[5]	Ensemble Deep Learning Algorithm for Forecasting of Rice Crop Yield based on Soil Nutrition Levels	Soil nutrients and crop production statistics	Deep Neural Network (DNN), Deep Belief Network (DBN) and Support Vector Machine (SVM), k-Nearest Neighbur (k-NN) and inputs are fed into a stacking based ensemble deep learning strategy termed Model Agnostic Meta-Learning (MAML) for classification	False Positive Rate (FPR), accuracy, specificity, recall, F-1 score, False-negative Rate (FNR), precision, False Predictive Value (NPV), Error and Mathew Correlation Coefficient (MCC)	MAML perorms better than the other ML methods with 89.5% accuracy
Banerjee, S. et. Al (2023)[18]	A Region wise Weather Data-Based Crop Recommendation System Using Different Machine Learning Algorithms	weather data like temperature, rainfall, humidity, Solar hour and previous years' crop data	Decision trees, Naïve Bayes, SVM, Random Forest, Logistic Regression	No mention of method for accuracy determination	Random Forest predicts with highest accuracy
Suganya, M. et al(2020)[19]	Crop Yield Prediction Using Supervised Machine Learning Techniques	weather data like temperature, rainfall, humidity, Soil Nutrient data like N,P,K	Logistic Regression,Linear Discriminant Analysis, K-Nearest Neighbors , Classification and Regression Trees, Support Vector Machines	Precession, Recall, F1-Score	Logistic Regression predicts with highest accuracy

From the table.1 it is evident that –

From the table.1 it is evident that –

1. Most of the researchers have taken data from 3 sources for the prediction of crop yield. They are :

a. Soil data i.e. pH value, Organic Carbon (OC) soil macronutrients like Nitrogen(N), Potassium(K), and Phosphorus(p). Some researchers also added calcium, magnesium, sulfur, manganese, copper, iron, etc. as features for their study.

b. Weather data which include temperature, humidity, rainfall, etc for the past few years.

c. Crop yield data for the past few years.

2. Mostly used Supervised Machine Learning Methods for Crop yield predictions are Random Forest (RF), Support Vector Machine (SVM), k-NN, ANN, XGBoost, ensemble ML method like Model Agnostic Meta-Learning (MAML)

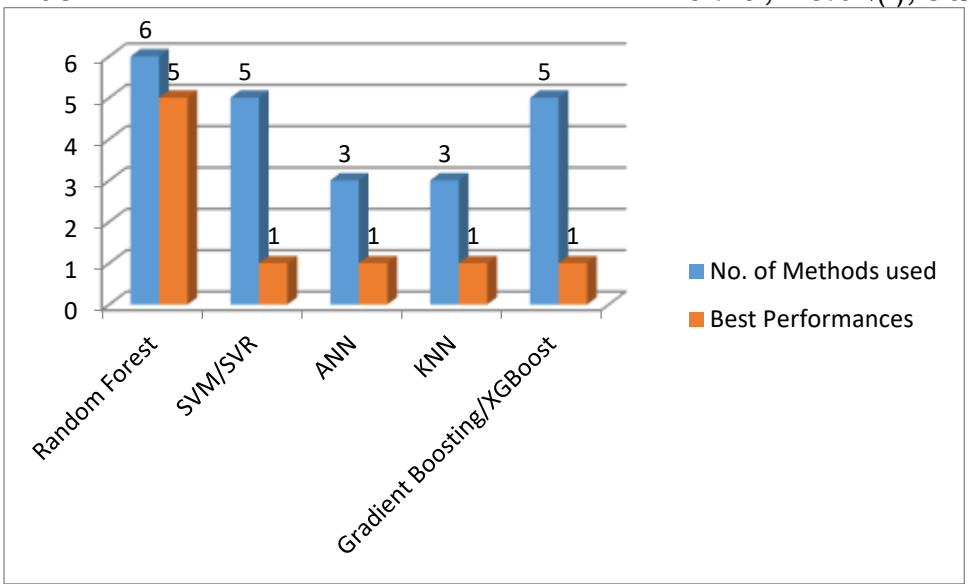
3. Performance evaluation is carried out by Coefficient of Determination (R2), Root Mean Square Error(RSME), etc.

4. Pre-processing of data for cleaning the noise is carried out by all researchers

5. Feature extraction is carried out by the researchers for dimensionality reduction and to fit the model with the preprocessed data.

6. Random Forrest (RF) and Random Forest Classifier/ Regressor outperform all other methods because they perform on a large number of decision trees but cannot handle complex data.

Figure:1-The Performance evaluation is depicted below in 3-D Column



In Figure-1 above 9 research papers have been taken for analysis. In most of the papers multiple ML methods are taken for rice yield prediction and the best performer method is depicted. It is clear from the figure-1 that Random Forest Method is the topmost choice of the researchers and its performance is also the best among all the ML Methods.

Conclusion

From the above study, it can be seen that Supervised Machine Learning Methods can be used effectively in predicting rice yield. Some methods can not handle complex types of data which causes overfitting, but proper choice of method can reduce the problem and can accurately predict the rice yield for the future. Deep learning Methods are the advanced methods that can eliminate the shortcomings of the traditional ML methods. If the researchers in this field apply Deep Learning Methods along with the ML Methods more accurate results can be achieved.

Bibliography/References:

1. Ramisetty,S., Bansode, D., Atmakur, V.K., Ghantasala,G.P., Usharsee,D.& Kumar,M.(2024).Zone II & III: Machine Learning Based Rice Yield Prediction in Andhra Pradesh.MATEC Web of Conference 392,01086 https://doi.org/10.1051/matecconf/202439201086
2. Maya Gopal,P.S. (2020, December). Efficient Paddy Crop Yield Prediction Using Machine Learning Technique. Ph.D thesis, Vellore Institute of Technology
3. Klompenburga.,T.V., Kassahuna.,A., & Catalb.,C.((177)2020) Crop yield prediction using machine learning: A systematic literature review. Computer and Electronics in Agriculture(Contents lists available at ScienceDirect)
4. Bhalerao, O. B., & J, R. M. (2015). Agricultural recommender using machine learning techniques by Lobo. 5(4).
5. Chandraprabha,M.& Dr.Dhanraj.,R.K.(2023). Ensemble Deep Learning Algorithm for Forecasting of Rice Crop Yield based on Soil Nutrition Levels. EAI Endorsed Transactions Research Article on Scalable Information Systems. doi: https://doi.org/10.4108/eetsis.v10i3.2610
6. Vargas, V.W., Aranda, J.E.S., Costa, R.D.S., Perira, P.R.D.S. & Barbusa, J.L.V.(2022), Imbalanced data preprocessing techniques for machine learning: A systematic mapping study.Knowledge and Information System, 65:31-57(2023). https://doi.org/10.1007/s10115-022-01772-8.
7. Shah, A., Dubey, A., Hemnani, V., Gala, D.,& Kalbande, D.R. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. Proceedings of International Conference on Wireless Communication . Lecture Notes on Data Engineering and Communications Technologies, vol 19. Springer, Singapore.
8. Panigrahi, B., Kathala,K.C.R, & Sujatha,M.(2023) Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models. International Conference on Machine Learning and Data Engineering, ScienceDirect. Procedia Computer Science

218 (2023) 2684–2693
9. Venugopal, A., Aparna, S., Mani,j., Mathew, R., & Williams, V.(2021). Crop Yield Prediction using Machine Learning Algorithms. International Journal of Engineering Research & Technology (IJERT) NCREIS – 2021.9(13)
10. Jiya,E.A., Illiyasu,U. & Mudashiru,A.(2023) Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms. Journal of Information Systems and Informatics 5(2) June 2023 e-ISSN: 2656-4882 p.ISSN: 2656-5935. DOI: 10.51519/journalisi.v5i2.506
11. Joshua, V., Priyadharson,S.,M. & Kannadasan,R. Exploration of Machine Learning Approaches for Paddy Yield Prediction in Eastern Part of Tamilnadu. Agronomy 2021, 11, 2068. https://doi.org/10.3390/ agronomy11102068
12. Kamath, P., Patil , P., Srilatha, S., Susma, & Somya, S. (2021). India: Crop yield Forecasating using Data Mining. Global transitions Proceedings.2, pp. 402-407. Dept. of Computer Science & engineering, Sri Madhwa Vadiraja institute of Echnology and Management.
13. Devdatta, A. & Bondre, M. S. (2019). Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms. International Journal of Engineering Applied Sciences and Technology, 4(5), 371-376. doi:http://www.ijeast.com
14. Taher, I. K., Abdulazeez, A. M., & Zebari, A. D. (2021). Data Mining Classification Algorithms for Analyzing Soil Data. Asian Journal of Research in Computer Science, 8(2), pp.17-28.
15. Wijayanti, E.,B., De Rosal Ignatius Moses Setiadi,J.,M.,S., & Setyoko,B.,H., (2024). Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting. Journal of Computing Theories and Application, ISSN:3024-9104, DOI : 10.62411/jcta.10057
16. Wilson.,A., Sukumar.,R., & Hemalatha, N. Machine Learning Model for Rice Yield Prediction Using KNN Regression. https://www.researchgate.net/publication/353944092.Preprint in agriRxiv · August 2021
17. Dohikar., S.S., & Rode.,S.(2014) Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2(1), ISSN (Print) 2321 – 5526
18. Banerje.,S & Mondal., A.C. (2023) A Region Wise Weather Data-Based Crop Recommendation System Using Different Machine Learning Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 11(3),pp.283-297, ISSN:2147-6799
19. Suganya.,M., Dayana.,R.& Rebathi.,R.,(2020).Crop Yield Prediction Using Supervised Learning Techniques.International Journal of Computer Engineering and Technology, 11(2), 2020, pp. 9-20
20. Archana, K. (2020, Jan-Feb). International Journal of Computer Science and Engineering (IJCSE), 9(1).