iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48709

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Machine Learning Use Case in Indian Agriculture: Predictive Analysis of Bihar Agriculture Data to Forecast Crop Yield

Vishakha Mistry[1], Abhishek Kumar Mishra[2], Nadiyah Ahmed[3]
[1]*Head of Department, IT Department,* [3]*Computer Science Research Associate, 360 Research Foundation, Nartakiganj, Bihar, India*
[2]*District Informatics Officer, National Informatics Center(NIC), Min. of Electronics and Information Technology, Gov of India, West Champaran, Bihar, India*

Abstract: ***As early as the Indus Valley Civilization Era, agriculture in India is recorded. The agriculture industry provides employment in developing countries such as India and is considered the backbone of the economy. The benefit of machine learning in farming is that it provides farmers with proper recommendations and judgments about crops. By applying machine learning to agriculture, farmers can increase efficiency, quality production, precision, and while consuming minimum human effort. This research work focused on applying various Machine Learning techniques for predicting the yield of the crop for the various districts of Bihar agricultural dataset. Here we used Random Forest, Decision Tree, SVR, XGBregressor, and Deep Neural Network for the prediction of crop yield, and their comparisons are made on the basis of MAE. This work will help farmers in predicting the yield of various crops based on past data. Therefore, farmers can select crops that suffer the fewest losses by using this tool.***

Keywords: Machine learning, Crop Yield prediction, Deep Neural Networks, Precision Agriculture, SVR, Random Forest, Decision Tree

I. INTRODUCTION

Agriculture is one of the largest sectors in India which provides employment to more than half of the Indian population and greatly contributes to the country's economy. Global food production relies heavily on crop yield predictions. In fact, India has the highest net cropped area in the world so it is important that the land under cultivation is used efficiently and extract maximum crop yield, it is important to make efforts to use the resources effectively and sustainably. Precision agriculture is one such concept that has revolutionized agriculture. Precision agriculture is an agriculture management strategy that helps in improving crop yields and assists in decision-making using large amounts of sensorial data and information as well as analysis tools. Technologies like IoT, Artificial Intelligence, and Cloud Computing are deployed for the collection of data related to land, weather conditions, fertilizers water management, and soil fertility. This paper largely covers the application of machine learning as part of precision agriculture for crop yield prediction. A crop yield survey is important for financial and management decision-making regarding the choice of crops, as well as, accurate predictions can help in making timely import and export decisions to improve national food security. In this paper, we have implemented and presented an analysis of several ML algorithms for yield prediction.

II. LITERATURE REVIEW

Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou [1], Simon Pearson, and Dionysis Bochtis reviewed 40 articles in total. On analysing these articles, it was found that a total of eight ML models have been implemented. For crop management, Artificial Neural Networks were the most popular, for livestock management, SVMs (Support Vector Machines) were the most popular for water management, ANNs were the most frequently implemented models and for soil management, ANNs were the most popular models [1]. This shows that by applying machine learning to sensor data, farm management systems are evolving into real artificial systems, providing richer recommendations and insights with the aim to improve production. But at the moment, individual approaches and solutions are not adequately connected to the decision-making process as seen in other application domains.

Prof. M.D Tambakhe, Dr V.S. Gulhane, Prof. J.S. Karnewar [2] have reviewed various applications of machine learning in the farming sector. The growing number of machine learning techniques in agriculture require large amounts of data that can be available from many sources and can be analysed to find hidden knowledge.

©IJRASET: All Rights are Reserved

Machine Learning can be very well implemented in fields where input and output variables have complex relationships. Machine Learning algorithms have boosted the accuracy of AI machines used in precision farming [2].

Venugopal, Aparna, Jinsu Mani, Rima, Prof. Vinu [3] focused on the prediction of crops and the calculation of their yield with the help of machine learning techniques.

Crop prediction for a chosen district was done from the collection of past data using Random Forest Classifiers on the basis of area, production, temperature, humidity, rainfall, and wind speed. The proposed technique would help the farmers in decision-making for the cultivation of crops.

III. MACHINE LEARNING USE CASES IN THE INDIAN AGRICULTURE DOMAIN

A. Species Selection

A difficult task lies ahead when it comes to choosing the right species. Climate change adaptation is essential for species as well as it should stand against various diseases [4]. Additionally, it should provide a variety of nutrients. Farm field data can be collected for several years and machine learning algorithms can be used to predict the correct species genes to assist farmers.

B. Plant Classification

There are thousands of species of plants. The ability to recognize and classify all plants are not practical with the traditional human approach. A variety of ML algorithms are used to extract leaf vein features and classify plants [4]. For any plant classifier, the shape is the most universal feature. Additionally, features like color, texture, and veins are utilized to classify plants.

C. Plant Disease Detection

Crop diseases are the main challenge in the agriculture domain as smallholder farmers whose lives depends on healthy crops. Agriculture experts directly identify the disease in plants. But for larger farm areas, this human approach takes time and the availability of skilled experts.

Agriculture-based ML algorithms are showing promising results. Popular Convolutional Neural Network (CNN) architecture takes disease datasets of different plants and shows excellent disease classification accuracy [5]. In addition to the use of mobile application and internet usage worldwide, disease diagnosis is made online for farmers.

D. Precision Irrigation Management

With advances in machine learning, irrigation decisions can now be made based on the concept of predicting a crop's water requirements based on the forecast of weather and soil conditions rather than relying on previous experience. Most widely used supervised ML algorithms (K nearest neighbor (KNN), support vector machine (SVM), decision trees (DT), random forest (RF)) can guide to optimize irrigation time, monthly water schedule, soil moisture prediction, and weather predictions. Also, it is observed that farmer experience low yield because of pests and insufficient irrigation supply [6].

IoT and machine learning scenarios can, therefore, be leveraged to implement a high-efficiency irrigation monitoring and control system via mobile and web applications, saving significant amounts of water, energy, and manpower by implementing a comprehensive monitoring and control system for agricultural irrigation.

E. Soil Management

As a result of poor crop and soil management strategies in recent years, soil quality has been heavily degraded. By using different machine learning algorithms, different chemical features and micronutrients present in the soil are analysed, and the soil's fertility is classified as well as soil moisture, and soil nutrient content is predicted [7]. As a result of the ML method, more accurate soil fertility predictions can be made, which will streamline farmers' difficulties and act as a medium for farmers to gain a more efficient result.

F. Yield Prediction

In order to forecast a higher yield for the coming season, farmers relied on their years of farm field experience with specific crops. Having accurate information about the crop yield minimizes loss on the part of the producers. Farmers will be able to decide what crop to grow based on whether, rains, environmental components, and other factors, using the prediction made by machine learning algorithms [8].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

IV. YIELD PREDICTION IMPLEMENTATION

Fig. 1 shows the implementation block diagram of yield Prediction.

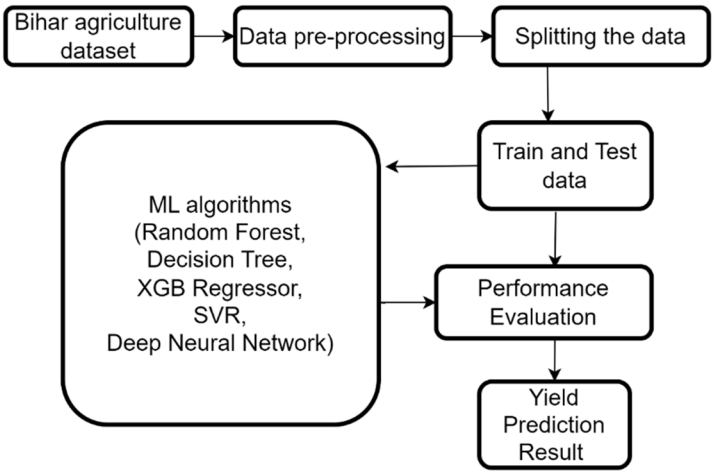


Fig. 1 Implementation diagram

A. Dataset and Features

Data is essential for any ML algorithm. And any model can only be efficient if it is fed the right amount of data. For our research work, we gathered agriculture data from the website of the Government of India. This data set is in CSV format and contains information from Bihar for a period of ten years. It has 18885 datapoints and 8 attributes like State Name, District name, crop year, season, Area, and Production and Yield. This dataset includes 38 districts of Bihar state. Out of them, 4 are categorical attributes and 2 are real.

B. Data Pre-Processing

It is necessary to have a large dataset for machine learning applications. So data pre-processing has to be performed on CSV file. By using a one-hot encoding technique, categorical data is encoded. In order to normalize the data, the Min-Max scalar is applied to each column, and 'NA'(missing) values are replaced with the mean of that column. Pre-processing is performed using the Python pandas library. The dataset, after preprocessing, is split into a training and testing dataset. The training data will comprise 70% and the testing data will comprise 30%.

C. Regression Algorithms

Machine Learning is a branch of Artificial Intelligence and computer science that uses data and algorithms to learn to do tasks. Algorithms are trained on data to come up with mathematical models that make predictions or decisions without being programmed to do so. A machine learning model gives us the relationship between different parameters of the data. Machine Learning has a wide range of applications in Medical Diagnosis, Stock Market Trading, Online Fraud Detection, Agriculture etc,. Machine Learning Regression algorithms used in this paper are discussed follow.

1) *Random Forest:* Random Forests consist of a large number of decision trees that perform as an ensemble. Each tree in a random forest gives out a class prediction and the prediction with the most votes is taken as the model's output. All decision trees of a random forest are uncorrelated. The fundamental idea behind Random Forest classifiers is that a large number of uncorrelated models will perform much better than individual models [9].
2) *SVR:* One of the most popular supervised learning algorithms, support vector machines are used to solve classification and regression problems. SVR algorithms are used to build models that assign new examples to different categories. Algorithms plot examples as points in n-dimensional space to find a hyperplane that differentiates the two categories the best [9].
3) *XGBRegressor:* Effective and efficient implementation of gradient boosting is available in the (XGBRegressor) open-source library. Kaggle competitions consistently place this method among the top contenders [9]. The advantage of XGBoost over gradient boosting algorithms is that it does not overfit data. It does this by using more accurate approximations as the basis of finding the optimal model to control overfitting [10].

©IJRASET: All Rights are Reserved

4) *Decision Tree:* Decision Trees are supervised learning algorithms used to solve classification and regression algorithms. Decision tree algorithms use tree-like representations which consist of the root node, branches, internal nodes, and leaf nodes [9]. The internal node represents feature values, the branches represent the conjunction of different features and the leaf node represents the decision class label. Decision trees are built using algorithmic approaches that identify ways to split datasets based on certain conditions.
5) *Deep Neural Network:* The characteristic of machine learning models is that they view the output (crop yield) as implicitly determined by the input variables. So input-output relationships are non-linear in nature. Supervised ML algorithms cannot be used for non-linear input-output relationships. The development of an advanced crop yield prediction model using deep learning has been made possible by technological advances in recent years. Feature extraction using deep learning approaches performs better than traditional machine learning approaches. Deep learning is able to extract features from available data with great accuracy since crop yield predictions depend on factors affecting crop growth. Due to the fact that this is a regression problem, in general, we have applied different algorithms, including Support Vector Regression(SVR), Random Forest, and Deep Neural Network.

D. Performance Measure

During performance evaluations, three factors are taken into consideration.

1) *Mean Absolute Error (MAE):* Mean absolute error is the average value of absolute errors for predicted and observed values. Absolute error is the difference between the predicted value and the true value. It is also referred to as the L1 loss function[11].
2) *Mean Square Error (MSE):* Mean squared error is calculated as the average of the square of the difference between actual value and the predicted value[11]. If MSE is large, it indicates that the data points are widely dispersed around the mean while a smaller MSE indicates less dispersion.
3) *R2 Score:* R2 Score, or R-squared, also called the coefficient of determination, is calculated as the variance of a dependent variable shown by the model over total variance. It is calculated as a percentage and the higher the percentage, the lesser is variance and the better is the mode [11].

V. RESULT

During the course of this research, machine learning frameworks played an important role. An experiment was run on an AMD Ryzen 5 4500U 2.38 GHz,8 GB of RAM system with ANACONDA software and Python programming language. Having successfully trained and tested the dataset, we moved on to find the model's performance on the basis of MAE, MSE, and R2 score. In Fig. 2, individual attributes are depicted using a heat map.

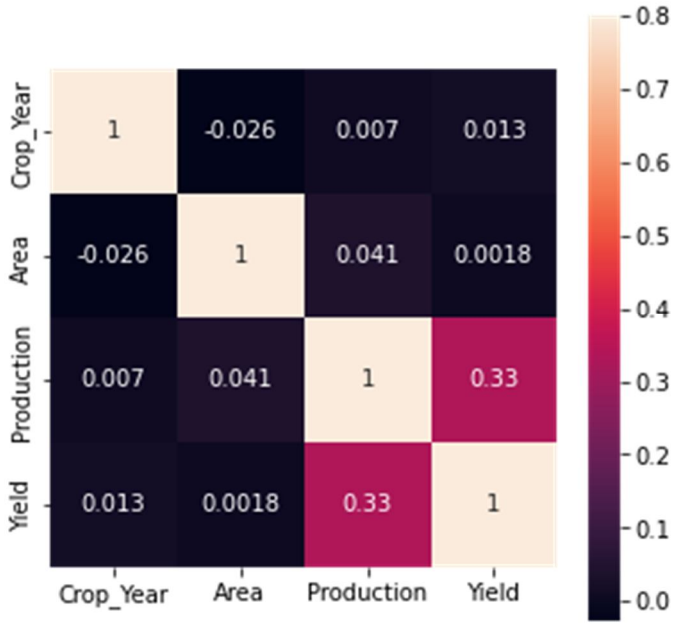


Fig. 2 plot of Heat map

TABLE I Performance Comparison Of Different Algorithms

Regression Algorithm	Performance evaluation		
	MSE	MAE	R2 score
Random Forest	21.3858	2.1307	0.6158
Decision Tree	12.0440	0.8283	0.8306
SVR	25.9422	1.5861	-0.4132
XGBRegressor	9.6624	0.80284	0.8661

	Performance evaluation			
Deep Neural network	Validation mean absolute error	MAE	Validation loss	loss
	0.3313	0.7470	0.7470	0.3313

As we can see from the table, Deep Neural Networks have the lowest Mean Absolute Error of all models for predicting crop yields. Hence it is the best model for our application. Same can be depicted in the below graph.

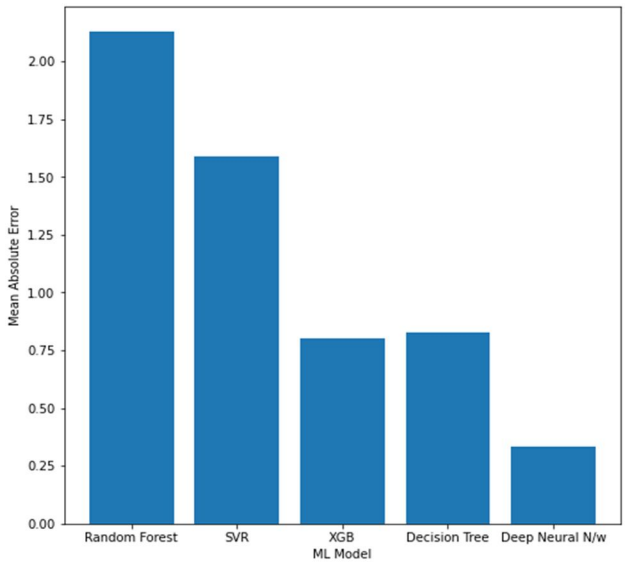


Fig. 3 comparison of different ML models on the basis of MAE

For the Deep neural network, based on the graph of epochs vs MAE as shown in Fig. 4, it is evident that as the epochs increase the error decreases, and in the end, we are able to create a model with a minimum MAE and maximum accuracy.

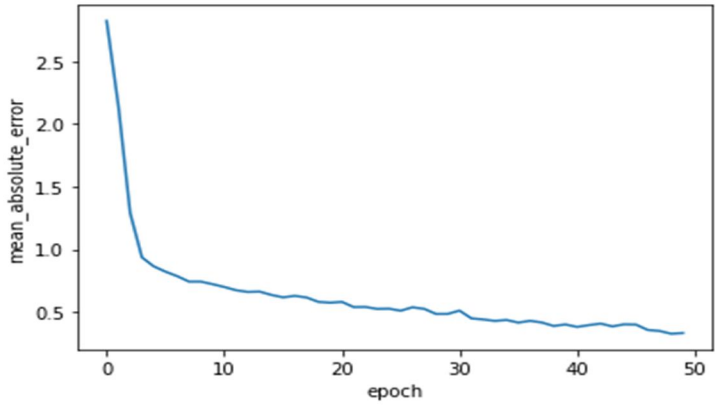


Fig. 4 epochs v/s MAE for Deep Neural Network

The scatter plot in Fig. 5 shows the predicted vs actual production value of the XGB regressor. Analyzing results using this method shows you at a glance how the predicted value deviates from the actual.

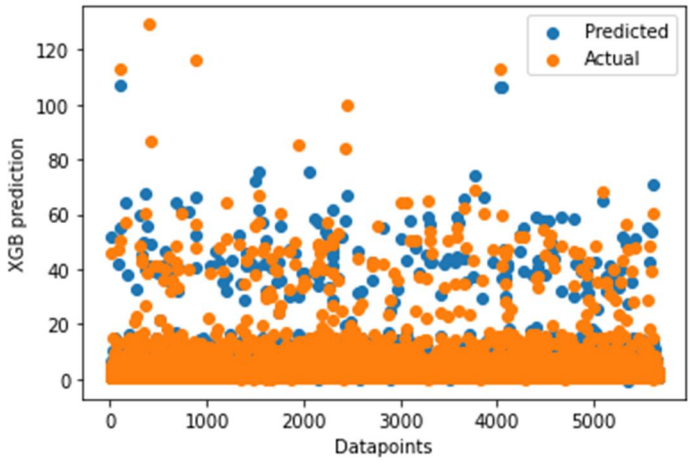


Fig. 5 Scatter plot of predicted v/s actual value of XGBRegressor

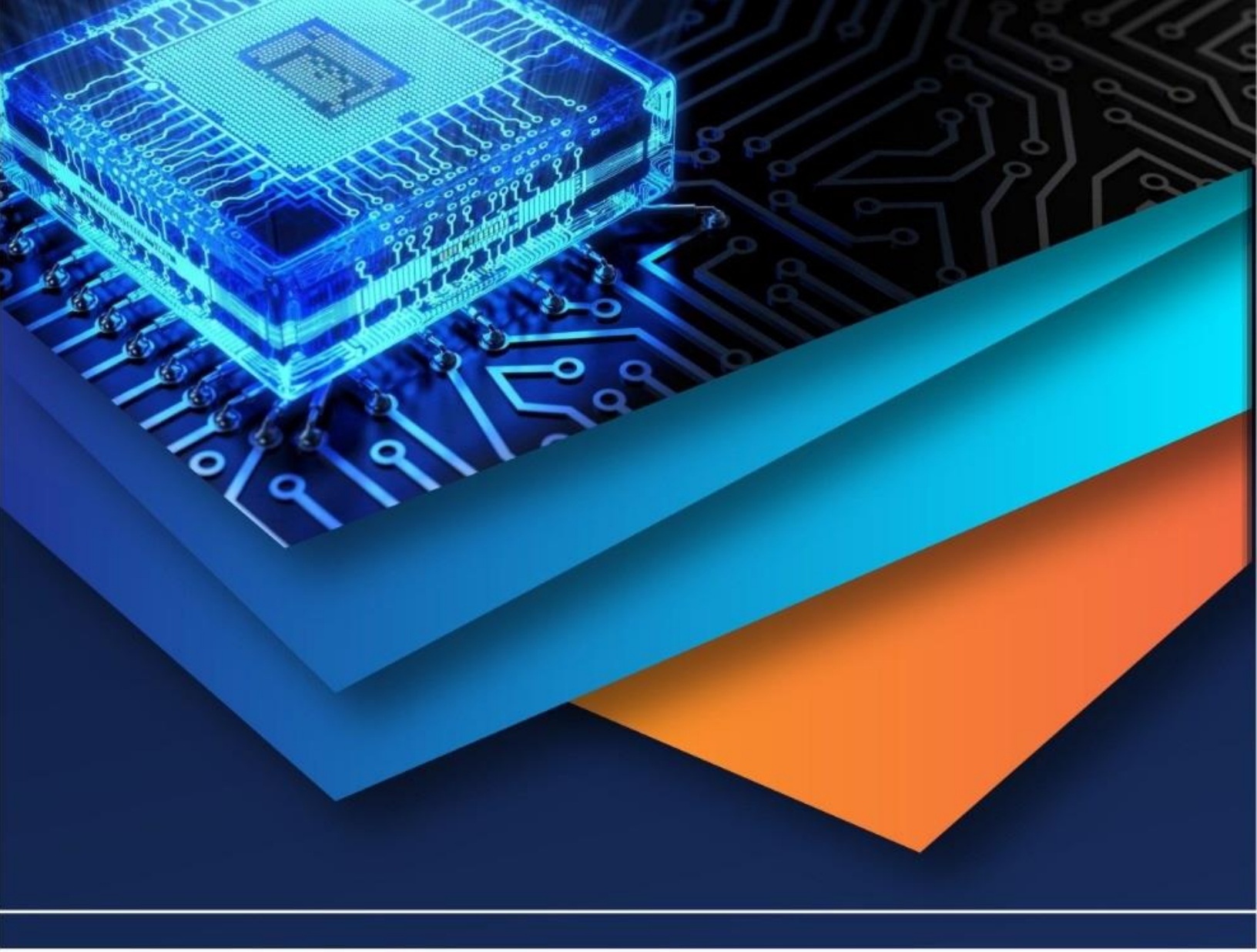
VI. CONCLUSION AND FUTURE SCOPE

The purpose of this paper is to utilize different machine learning techniques to predict crops and calculate their yield in 38 districts of Bihar state agriculture data. We have implemented and evaluated 5 different machine-learning algorithms which were trained on past data from Bihar for the years 1997-2014. Out of which Deep Neural Network has shown better performance. In farming, the proposed techniques in crop yield prediction help in efficient decision-making regarding what kind of crops to grow, harvesting activities, and matching crop supply with demand.

In this paper, we covered 8 features. However, this work can be extended with more features like soil quality, rainfall, and weather data. An Android app can be developed to predict the crop and calculate the yield. Such a system will help to maximize crop production in Indian agriculture and to raise farmer's revenue.

REFERENCES

[1] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, "Machine Learning in Agriculture: A Review", Sensors 2018, 18, 2674; doi:10.3390/s18082674.
[2] Prof. M.D Tambakhe, Dr V.S. Gulhane, Prof. J.S. Karnewar, "Machine Learning Application in the field of Agriculture: A Review", International Journal of Research in Advent Technology, Vol.7, No.4, April 2019.
[3] Venugopal, Aparna, Jinsu Mani, Rima, Prof. Vinu, "Crop Yield Prediction using Machine Learning Algorithms", International Journal of Engineering Research & Technology, Volume 9, Issue 13, Special Issue – 2021.
[4] Grinblat, G.L.; Uzal, L.C.; Larese, M.G.; Granitto, P.M. Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. 2016, 127, 418–424
[5] Jasmeet Kaur, Er. Ramanpreet Kaur, "Plant Disease Detection using SVM Algorithm and Neural Network Approach", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2016
[6] Abiodun Abioye, Oliver Hense, Travis J. Esau, "Precision Irrigation Management Using Machine Learning and Digital Farming Solutions", AgriEngineering 2022, 4, 70–103. https://doi.org/10.3390/agriengineering4010006
[7] Blesslin Sheeba,1 L. D. Vijay Anand, Gunaselvi Manohar, "Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms", Hindawi Journal of Nanomaterials Volume 2022, Article ID 5343965
[8] Pallavi Kamath, Pallavi Patil, Shrilatha S, "Crop yield forecasting using data mining" Global Transitions proceedings 2(2021), page No. - 402-407
[9] D.T. Larose, Discovering knowledge in data: An Introduction to data mining, John Wiley & Sons, (2005)
[10] Tianqi Chen and Carlson Guestrin, "Xgboost: A scalable tree boosting system", in Proceedings of the 22nd ACM SIGKDD international conference knowledge discovery and data mining, KDD '16, pages 785-794, New York, NY, USA, 2016
[11] S. Sridhar, M. Vijayalaxmi: Machine Learning, Oxford University Press, (2021)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)