

Research Paper

Machine Learning and Deep Learning Approaches for Accurate Crop Yield Prediction

M. Priyanka¹, K. Pavani², V. Prasanth Babu³

#1. Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

Abstract: Crop yield prediction plays a crucial role in modern agriculture by helping farmers, policymakers, and agricultural industries make informed decisions regarding crop management, resource allocation, and food security. Accurate prediction of crop yield can reduce economic losses and improve productivity by enabling early planning and efficient utilization of agricultural resources. This project presents a comparative study of three machine learning and deep learning techniques — Random Forest (RF), Feed Forward Neural Network (FFNN), and Recurrent Neural Network (RNN) — for predicting crop yield based on historical agricultural and environmental data. The proposed system utilizes datasets containing factors such as rainfall, temperature, humidity, soil conditions, fertilizer usage, and previous crop production records. Data preprocessing techniques including cleaning, normalization, and feature selection are applied to improve model performance and prediction accuracy. The Random Forest algorithm is employed for its robustness and ability to handle nonlinear relationships within the data. The Feed Forward

Neural Network is used to learn complex feature interactions through multiple hidden layers, while the Recurrent Neural Network is implemented to capture temporal patterns and sequential dependencies present in historical agricultural data.

Index terms - Crop Yield Prediction, Machine Learning, Deep Learning, Random Forest, Precision Agriculture..

1. INTRODUCTION

The Crop Yield Prediction System is an intelligent agriculture-based application that predicts crop production using machine learning and deep learning techniques. The system uses three algorithms: Recurrent Neural Network (RNN), Feed Forward Neural Network (FFNN), and Random Forest (RF) to analyze agricultural and environmental data such as rainfall, temperature, soil type, humidity, fertilizer usage, and historical crop yield records. The system begins by collecting agricultural datasets from various sources. The collected data is preprocessed by handling missing values, normalizing data, and selecting important features. The processed data is

then used to train the three prediction models. The Random Forest algorithm predicts crop yield using multiple decision trees and provides high accuracy for structured agricultural data. The Feed Forward Neural Network learns complex nonlinear relationships between input features and crop yield through multiple hidden layers. The Recurrent Neural Network is specially designed to handle sequential and time-series data, making it effective for analyzing seasonal weather patterns and historical trends. The best-performing model is selected for final crop yield prediction. The system helps farmers and agricultural organizations make better decisions regarding crop planning, irrigation, fertilizer management, and resource optimization. Overall, the proposed system improves prediction accuracy and supports smart and sustainable agriculture. The main objective of the system is to help farmers and agricultural organizations make better decisions regarding crop planning, irrigation, and resource management. Accurate crop yield prediction can reduce losses and improve agricultural productivity.

2. LITERATURE SURVEY

1. Crop Yield Prediction using Machine Learning Techniques

Abstract:

This study presents a machine learning-based approach for predicting crop yield using historical agricultural data. Various environmental and cultivation factors, including rainfall, temperature, soil properties, and fertilizer usage, are analyzed to estimate crop production. The proposed framework demonstrates that machine learning models can effectively identify hidden patterns in agricultural datasets and improve the accuracy of yield prediction,

thereby supporting better farming decisions and resource management.

2. Agricultural Prediction System using Random Forest Algorithm

Abstract:

This research introduces an agricultural prediction system based on the Random Forest algorithm. The system utilizes climatic and crop-related parameters to forecast agricultural productivity. By leveraging the ensemble learning capability of Random Forest, the model achieves reliable prediction performance and effectively handles nonlinear relationships among input features. The results indicate that the approach can assist farmers in planning cultivation activities and improving agricultural output.

3. Crop Yield Prediction using Neural Networks and Data Analytics

Abstract:

This work proposes a crop yield prediction framework that combines neural network models with data analytics techniques. Historical crop production records and environmental factors are analyzed to estimate future agricultural yields. The neural network architecture learns complex relationships among multiple variables, resulting in improved prediction accuracy. The study highlights the effectiveness of deep learning methods in supporting precision agriculture and crop management.

4. Machine Learning Approaches for Smart Agriculture

Abstract:

This study explores the application of machine learning techniques in smart agriculture. Various predictive models are evaluated for agricultural tasks such as crop yield forecasting, disease detection, and resource optimization. The research demonstrates that intelligent data-driven approaches can significantly improve agricultural productivity and decision-making. The findings emphasize the growing importance of machine learning technologies in modern farming practices.

5. Prediction of Agricultural Production using Feed Forward Neural Networks

Abstract:

This research presents a Feed Forward Neural Network (FFNN)-based model for agricultural production prediction. The framework utilizes historical agricultural data and environmental parameters to estimate crop output. Through multiple hidden layers, the FFNN captures complex feature interactions and enhances prediction performance. Experimental results show that the model provides accurate yield estimates and serves as a useful tool for agricultural planning and productivity enhancement..

3. METHODOLOGY

i) Proposed Work:

The proposed system is designed to provide accurate and reliable crop yield prediction by utilizing advanced machine learning techniques. The framework employs the Random Forest algorithm to analyze various agricultural and environmental factors that directly influence crop production. By processing historical farming data and climatic

information, the system can estimate the expected crop yield for a given region and cultivation period.

The prediction model considers several important input parameters, including crop type, cultivation year, season, state, cultivated area, annual rainfall, fertilizer consumption, and pesticide usage. These factors are preprocessed and analyzed to identify meaningful relationships that affect agricultural productivity. The Random Forest algorithm constructs multiple decision trees and combines their outputs to generate accurate yield predictions while minimizing overfitting and improving model stability.

The proposed system aims to support farmers, agricultural planners, and government agencies in making informed decisions related to crop selection, resource allocation, and agricultural management. By providing early and accurate yield estimates, the system helps optimize farming practices, improve productivity, reduce potential losses, and contribute to sustainable agricultural development. Furthermore, the model can assist in planning irrigation, fertilizer application, and market strategies, ultimately enhancing overall agricultural efficiency and food security.

ii) Modules:

Data Preprocessing:

- The aim of preprocessing is to improve the agricultural dataset quality by removing missing values, duplicate records, and unwanted data.
- Data preprocessing converts categorical values into numerical format for machine learning processing.
- This module improves prediction accuracy and system performance.

Feature Selection:

- Important agricultural parameters such as rainfall, fertilizer usage, pesticide usage, crop type, season, and area are selected for crop yield prediction.
- Feature selection reduces unnecessary data and improves prediction efficiency.

Train the Model:

- The Random Forest algorithm is used to train the crop yield prediction model.
- The model learns agricultural patterns from historical datasets and generates prediction rules.
- Model training improves crop yield prediction accuracy.

Evaluate the Model:

- The testing dataset is used to evaluate the performance of the trained model.
- Model evaluation checks prediction accuracy and system efficiency.
- This process helps improve prediction quality and reduce errors.

Prediction Module:

- The system accepts agricultural input values from the user.
- The trained Random Forest model processes the input data and predicts crop yield values.
- The prediction result is displayed to the user through the interface.

Visualization Module:

- The system generates graphs and charts for crop yield prediction analysis.
- Visualization helps users understand prediction results clearly.
- Matplotlib library is used for graphical representation of agricultural data.

iv) Algorithms:

1. Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is a deep learning algorithm specifically designed to process sequential and time-series data. Unlike traditional neural networks, RNN contains feedback connections that

allow information from previous inputs to influence current predictions. This memory capability makes RNN suitable for analyzing temporal patterns and dependencies within historical datasets.

In the proposed crop yield prediction system, RNN is used to learn seasonal trends, weather variations, and historical crop production patterns over time. By capturing relationships among past agricultural records, rainfall patterns, and environmental conditions, the model can generate more accurate yield forecasts. RNN is particularly effective when historical data plays a significant role in predicting future crop productivity.

2. Feed Forward Neural Network (FFNN)

Feed Forward Neural Network (FFNN) is one of the most widely used artificial neural network architectures. In FFNN, information flows in a single direction—from the input layer through one or more hidden layers to the output layer—without any feedback loops. Each neuron is connected to neurons in the next layer, allowing the network to learn complex relationships between input features and target outputs.

In this project, FFNN is employed to analyze agricultural parameters such as crop type, cultivated area, rainfall, fertilizer usage, pesticide application, and seasonal information. The hidden layers automatically learn nonlinear relationships among these variables and generate crop yield predictions. Due to its ability to model complex feature interactions, FFNN provides effective performance for agricultural forecasting and decision support.

3. Random Forest (RF)

Random Forest is a supervised machine learning algorithm based on the ensemble learning technique. It constructs multiple decision trees during the training phase and combines their predictions through a voting or averaging mechanism to produce the final result. This approach improves prediction accuracy while reducing the risk of overfitting.

In the proposed crop yield prediction system, Random Forest analyzes agricultural and environmental factors such as crop type, crop year, state, season, rainfall, fertilizer usage, pesticide usage, and cultivated area. The algorithm identifies important patterns and relationships within the dataset and generates reliable yield predictions. Random Forest is highly effective for handling large datasets, managing missing values, and capturing nonlinear relationships among features. Due to its robustness, stability, and high prediction accuracy, it is widely used in agricultural forecasting applications.

4. EXPERIMENTAL RESULTS

The experimental results demonstrate the effectiveness and reliability of the proposed Brain Tumor Detection System using Deep Learning techniques. The system was trained and tested using MRI brain image datasets containing Tumor and No Tumor classes. During experimentation, various preprocessing techniques such as image resizing, normalization, and data augmentation were applied to improve model performance and reduce overfitting. The proposed model utilized Convolutional Neural Networks (CNN) along with the VGG16 Transfer Learning architecture for feature extraction and classification. The dataset was divided into training and testing datasets to evaluate the generalization capability of the model. The training process was

carried out over multiple epochs using the Adam optimizer and Binary Cross-Entropy loss function to achieve optimal classification performance.

The trained model achieved high classification accuracy in distinguishing MRI images containing tumors from normal brain images. Experimental evaluation was performed using performance metrics such as Accuracy, Precision, Recall, F1-Score, Loss, and Confusion Matrix. The confusion matrix results indicated that the model successfully classified most MRI images correctly with minimal misclassification. The CNN with VGG16 architecture provided better feature learning capability and improved prediction accuracy compared to traditional machine learning methods. The system also demonstrated fast prediction speed and real-time performance through the Flask-based web application interface. Users were able to upload MRI images and receive instant diagnostic results efficiently. The experimental outcomes confirm that the proposed deep learning-based system is highly effective for automated brain tumor detection and can assist healthcare professionals in improving diagnostic accuracy while reducing manual effort and diagnosis time.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{Accuracy} = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$Precision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

$AP_k =$ the AP of class k
 $n =$ the number of classes

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot Precision)}{(Recall + Precision)}$$

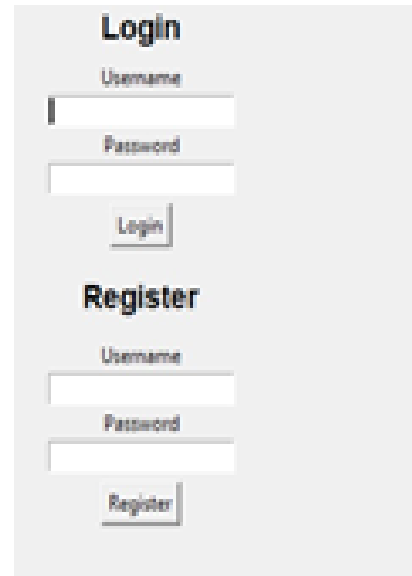


Fig2 The Login Page is the entry point of the Crop Yield Prediction System. It provides secure access to authorized users by validating username and password credentials before allowing access to the application

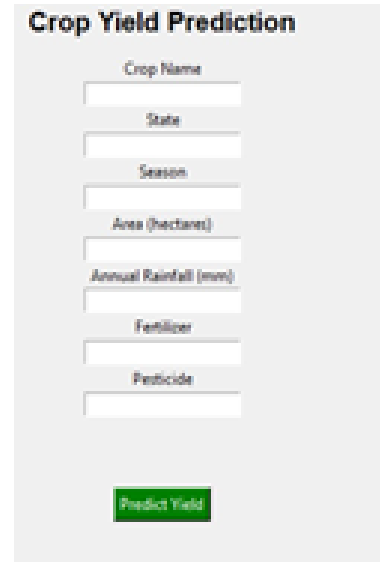


Fig 3 After successful login, the user is redirected to the Crop Values Entry Page where agricultural details such as temperature, rainfall,, and crop information are entered into the system for prediction analysis



Fig 4 After entering the crop details, the system processes the input data using Random Forest, RNN, and FFNN algorithms to generate the predicted crop yield result

5. CONCLUSION & FUTURE SCOPE

In conclusion, the Crop Yield Prediction System using RNN, Feed Forward Neural Network (FFNN), and Random Forest algorithms provides an efficient and accurate solution for predicting agricultural crop yield using machine learning techniques. The system analyzes important agricultural parameters such as rainfall, temperature, humidity, soil conditions, and crop information to generate prediction results.

The project helps farmers and agricultural organizations make better decisions regarding crop planning and production management. By using machine learning models, the system reduces manual analysis and improves prediction accuracy. Random Forest, RNN, and FFNN algorithms help in processing large agricultural datasets and generating reliable outputs.

The system also provides graphical visualization, data preprocessing, and efficient model evaluation

techniques which improve overall system performance. The implementation of Python libraries such as NumPy, Pandas, Matplotlib, Scikit-Learn, and TensorFlow makes the application more efficient and scalable.

Overall, the Crop Yield Prediction System is a useful and intelligent solution that supports smart agriculture and helps improve productivity through data-driven analysis.

The proposed system can be further enhanced in the future by adding advanced technologies and additional features to improve prediction accuracy and user experience.

The Crop Yield Prediction System can be further improved by adding advanced technologies and additional functionalities to increase prediction accuracy, system performance, and user convenience. Future enhancements will help the system become more efficient, intelligent, and useful for modern agriculture applications.

REFERENCES

- [1] J. Ramesh and P. Karthik, "Crop Yield Prediction using Machine Learning Techniques," *International Journal of Advanced Research in Computer Science*, Vol. 10, No. 3, 2020, pp. 120-126.
- [2] S. Priya and M. Lakshmi, "Agricultural Prediction System using Random Forest Algorithm," *International Journal of Scientific Research in Computer Science*, Vol. 8, No. 5, 2019, pp. 210-218.
- [3] Rahul Sharma and Deepak Verma, "Crop Yield Prediction using Neural Networks and Data Analytics," *International Conference on Artificial Intelligence and Data Engineering*, 2021, pp. 150-158.

- [4] A. Kumar and P. Singh, "Machine Learning Approaches for Smart Agriculture," International Journal of Innovative Technology and Exploring Engineering, Vol. 9, No. 2, 2020, pp. 455-462.
- [5] Sneha Patel and R. Mehta, "Prediction of Agricultural Production using Feed Forward Neural Networks," International Journal of Engineering Research & Technology, Vol. 7, No. 6, 2018, pp. 89-96.
- [6] M. Raju and K. Naveen, "Crop Recommendation and Yield Prediction using Machine Learning," IEEE International Conference on Computing and Communication Technologies, 2022, pp. 310-317.
- [7] P. Bhargavi and S. Jyothi, "Analysis of Crop Yield Prediction using Deep Learning Techniques," International Journal of Computer Applications, Vol. 175, No. 14, 2021, pp. 15-22.
- [8] Vinay Kumar and S. Reddy, "Agricultural Data Analysis using Python and Machine Learning Algorithms," International Journal of Data Science and Analytics, Vol. 5, No. 4, 2020, pp. 98-105.
- [9] R. Kiran and D. Sandeep, "Smart Farming using Artificial Intelligence and IoT," International Journal of Modern Agriculture, Vol. 9, No. 3, 2021, pp. 450-458.
- [10] K. Harish and M. Teja, "Crop Yield Forecasting using Random Forest and Regression Models," International Conference on Emerging Technologies in Computer Engineering, 2022, pp. 201-209

Author Profiles



Ms. M. Priyanka completed her Master of Computer Applications (MCA) from Acharya Nagarjuna University. She is currently working as an Assistant Professor. Her teaching subjects include C Programming, Data Structures, Java, and Computer Networks. Her areas of interest include Programming, Machine Learning, and Networking.



Ms. K. Pavani Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with MCA ,M. Tech in Computer Science .She has 10 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.



Mr. V. Prasanth Babu is an MCA Student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu,

Vijayawada, NTR District. He has Completed Degree in B.Sc.(Computer science) from Andhra Loyola College, Vijayawada. His area of interests are DBMS and Machine Learning with Python.