International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 05, May 2026)

Yield Prediction Using Machine Learning

N. Aravindhan[1], R. Hariprasad[2]

[1]Assistant Professor, [2]II MCA, Department of Master of Computer Applications, Er. Perumal Manimekalai College of Engineering, Hosur, Tamil Nadu, India.

***Abstract* -- Agriculture plays a major role in the economy and food production of many countries. Accurate crop yield prediction is important for farmers and agricultural planners to improve productivity and manage resources effectively. Traditional prediction methods mainly depend on historical records and manual observations, which often produce inaccurate results due to changing environmental conditions. This project presents a Machine Learning-based Crop Yield Prediction System that analyzes agricultural data such as rainfall, temperature, humidity, soil type, and crop information to generate accurate yield predictions. The system uses machine learning algorithms including Random Forest, Decision Tree, and Support Vector Machine (SVM) to improve prediction performance. The application is developed using Python, Flask, Scikit-learn, Pandas, NumPy, and MySQL. The system also provides data visualization and a user-friendly interface for prediction analysis. By automating the prediction process, the proposed system improves accuracy, reduces manual effort, and supports better agricultural decision-making.**

***Keywords* -- Crop Yield Prediction, Machine Learning, Random Forest, Support Vector Machine, Agriculture, Python, Flask, Scikit-learn, Data Analysis, Smart Farming.**

I. Introduction

Agriculture plays a vital role in the economic development of many countries and is essential for ensuring food security. Accurate prediction of crop yield is important for farmers, agricultural planners, and policymakers to make effective decisions regarding crop selection, resource management, and market planning. Traditional methods of yield prediction mainly depend on historical records, manual observations, and farmer experience, which often produce inaccurate and inconsistent results due to changing environmental conditions.

With the rapid growth of Machine Learning (ML) technologies, advanced prediction systems can now analyze large volumes of agricultural data efficiently. Machine learning algorithms are capable of identifying hidden patterns and relationships among factors such as rainfall, temperature, humidity, soil type, and crop conditions. These techniques improve prediction accuracy and support better agricultural planning.

To address the limitations of traditional methods, the proposed system "Yield Prediction Using Machine Learning" is developed as an intelligent prediction platform that uses machine learning algorithms to estimate agricultural crop yield accurately. The system utilizes historical and real-time agricultural datasets to train predictive models such as Random Forest, Decision Tree, and Support Vector Machine (SVM). By analyzing environmental and soil-related parameters, the system generates reliable yield predictions that help farmers improve productivity and reduce risks.

1.1 Importance of Crop Yield Prediction

Crop yield prediction is important for improving agricultural productivity and ensuring efficient resource utilization. Accurate prediction helps farmers select suitable crops, manage fertilizers and irrigation effectively, and reduce financial losses caused by unfavorable climatic conditions. It also supports governments and agricultural organizations in food supply planning and policy-making.

Machine learning-based prediction systems provide faster and more accurate results compared to traditional estimation methods. These systems can process large datasets efficiently and generate predictions that assist farmers in making informed agricultural decisions.

1.2 Challenges in Traditional Yield Prediction

Traditional crop yield prediction methods involve several limitations, including:

- Dependence on manual observations and historical records
- Inaccurate prediction due to changing climate conditions
- Difficulty in analyzing multiple environmental factors simultaneously
- Time-consuming data analysis process
- Higher chances of human error
- Limited use of advanced technologies for prediction

II. Existing System

The existing crop yield prediction system mainly depends on traditional agricultural practices and manual estimation methods.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 05, May 2026)

Farmers generally predict crop production based on personal experience, historical records, rainfall conditions, soil fertility, and climatic observations. In some cases, basic statistical methods and spreadsheet-based calculations are used for yield estimation.

These traditional methods are often time-consuming and less accurate because they cannot efficiently analyze multiple environmental factors simultaneously. Factors such as temperature, humidity, soil nutrients, rainfall patterns, and seasonal variations greatly influence agricultural productivity, but conventional systems fail to process these complex relationships effectively.

Existing systems also have limited automation and do not support real-time data analysis. Predictions are mostly based on manual calculations and assumptions, which may lead to inconsistent and unreliable results. In addition, traditional methods are not scalable for handling large agricultural datasets collected from different regions and farming conditions.

The lack of advanced technologies such as Machine Learning and intelligent data analysis reduces prediction accuracy and affects decision-making for farmers and agricultural planners. These limitations create the need for a more accurate, automated, and data-driven crop yield prediction system.

2.1 Limitations of Existing System

- Depends mainly on manual observations and historical records
- Less accurate prediction results
- Time-consuming data analysis process
- Unable to analyze multiple factors efficiently
- Higher chances of human error
- Limited automation and scalability
- No real-time prediction capability
- Difficult to handle large agricultural datasets

III. Proposed System

To overcome the limitations of traditional crop yield prediction methods, the proposed system introduces a Machine Learning-based Crop Yield Prediction System that provides accurate and efficient agricultural yield estimation. The system uses advanced machine learning algorithms to analyze agricultural and environmental data such as rainfall, temperature, humidity, soil type, nutrients, and crop information.

The proposed system collects historical and real-time agricultural datasets from different sources and processes the data through preprocessing and feature selection techniques.

Machine learning models such as Random Forest, Decision Tree, and Support Vector Machine (SVM) are trained using these datasets to identify hidden patterns and relationships between environmental factors and crop productivity.

The system provides a user-friendly interface where users can enter agricultural parameters and obtain instant crop yield predictions. Data visualization techniques such as graphs and charts are also integrated to help users understand prediction results and agricultural trends more effectively.

The proposed system automates the prediction process, reduces manual effort, and improves prediction accuracy compared to traditional methods. It also supports efficient handling of large datasets and can be extended with advanced technologies such as Deep Learning and IoT-based real-time data collection systems.

3.1 Advantages of Proposed System

- Provides accurate crop yield prediction
- Automates agricultural data analysis
- Reduces human error and manual effort
- Supports real-time and large-scale data processing
- Improves decision-making for farmers
- User-friendly and easy to use
- Scalable and flexible system architecture
- Supports advanced machine learning algorithms

IV. Related Work

Several research studies have explored the use of Machine Learning techniques in agriculture to improve crop yield prediction accuracy and support smart farming practices. Traditional statistical methods were initially used for agricultural forecasting, but these methods often failed to analyze complex relationships between environmental and crop-related factors.

Recent studies have applied machine learning algorithms such as Linear Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM) for yield prediction. Researchers found that Random Forest and Decision Tree algorithms provide better prediction accuracy due to their ability to process non-linear and large-scale agricultural datasets efficiently.

Many researchers have also implemented Artificial Neural Networks (ANNs) and Deep Learning models for crop yield estimation. These models are capable of analyzing complex agricultural patterns and improving prediction performance when compared to traditional machine learning approaches. Deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are increasingly used with satellite imagery and remote sensing data for large-scale agricultural analysis.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 05, May 2026)

Several studies have integrated IoT devices and real-time weather monitoring systems with machine learning models to improve prediction reliability and provide dynamic agricultural recommendations. Visualization tools and data analytics techniques are also widely used to help farmers understand agricultural trends and make informed decisions.

The proposed system builds upon these existing research works by implementing efficient machine learning algorithms, data preprocessing methods, and user-friendly prediction interfaces to provide accurate and reliable crop yield prediction for modern agricultural applications.

V. METHODOLOGY

The proposed system works through a machine learning-based prediction platform where users can analyze agricultural parameters and obtain crop yield predictions efficiently. The system follows a client-server architecture in which the frontend communicates with the backend through APIs. Machine learning models are trained using historical agricultural datasets to generate accurate yield predictions based on environmental and soil-related factors.

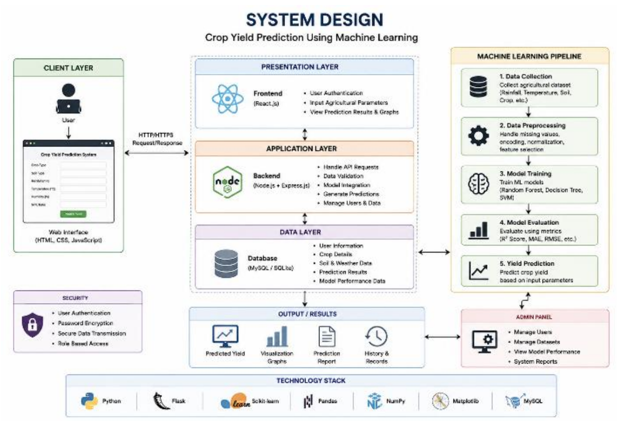
Users can provide inputs such as rainfall, temperature, humidity, soil type, and crop information through the web interface. The backend processes the input data, performs preprocessing and feature selection, and applies machine learning algorithms to generate yield predictions. The system also provides visualization graphs and prediction reports for better agricultural analysis.

5.1 Requirement Analysis

The system requirements include:

- Secure user authentication and login system
- Agricultural dataset collection and preprocessing
- Machine learning model training and prediction
- Crop yield prediction interface
- Data visualization and report generation
- Database management for agricultural records
- Real-time prediction and analysis support

5.2 System Design



The system follows a three-tier architecture consisting of frontend, backend, and database layers.

The frontend provides a user-friendly interface for entering agricultural parameters and viewing prediction results. The backend processes input data, handles machine learning operations, and generates predictions. The database stores agricultural datasets, user inputs, prediction results, and model-related information.

5.3 Implementation

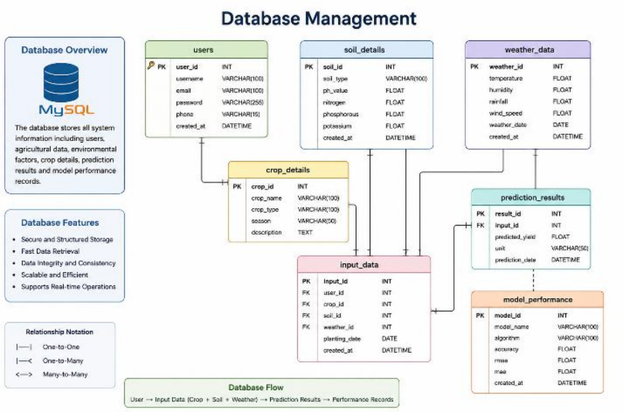
The system is implemented using Python-based machine learning technologies:

- Python – Core programming language
- Flask – Backend web framework
- Scikit-learn – Machine learning algorithms
- Pandas and NumPy – Data preprocessing and analysis
- Matplotlib and Seaborn – Data visualization
- MySQL / SQLite – Database management system

Users can input agricultural parameters through the web interface, and the system generates crop yield predictions using trained machine learning models such as Random Forest, Decision Tree, and Support Vector Machine (SVM).

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 05, May 2026)

5.4 Database Management



MySQL / SQLite is used as the primary database for storing system data. The database maintains tables for:

- User information
- Agricultural datasets
- Soil and weather records
- Crop details
- Prediction results
- Model performance records

Structured data storage and indexing techniques ensure efficient retrieval, scalability, and reliable data management.

5.5 Testing and Validation

Testing is performed to verify that the system functions correctly and generates accurate predictions. Different test cases are used to validate data preprocessing, model training, prediction accuracy, user inputs, and database operations. Performance evaluation metrics such as accuracy score and prediction analysis are used to validate the reliability of machine learning models. The testing process ensures that the system operates efficiently, accurately, and securely.

VI. System Architecture

6.1 Client (Frontend)

The frontend of the system is developed using HTML, CSS, JavaScript, and Flask templates. It provides a user-friendly interface where users can enter agricultural parameters such as rainfall, temperature, humidity, soil type, and crop information to obtain crop yield predictions. The frontend also displays prediction results, graphs, and visualization reports for better analysis.

6.2 Server (Backend)

The backend server is implemented using Python and Flask. It handles user requests, data preprocessing, machine learning model integration, prediction processing, and communication with the database. The backend is responsible for training machine learning models, processing input data, and generating accurate crop yield predictions.

6.3 Database (MySQL / SQLite)

MySQL / SQLite is used as the primary database for storing agricultural datasets, user inputs, prediction results, and model-related information. The database ensures efficient data storage, retrieval, and management for machine learning operations and prediction analysis.

VII. System Modules

7.1 Data Collection Module

This module collects agricultural datasets from different sources, including weather records, soil information, and crop production data. The collected data is used for training and testing machine learning models.

7.2 Data Preprocessing Module

This module performs data cleaning, handling missing values, normalization, and transformation of agricultural datasets. It improves the quality of input data before applying machine learning algorithms.

7.3 Feature Selection Module

This module identifies important agricultural parameters such as rainfall, temperature, humidity, and soil nutrients that significantly influence crop yield prediction. It helps improve model accuracy and reduces computational complexity.

7.4 Model Training Module

This module trains machine learning algorithms such as Random Forest, Decision Tree, and Support Vector Machine (SVM) using historical agricultural datasets. The trained models learn patterns and relationships between environmental factors and crop productivity.

7.5 Prediction Module

This module generates crop yield predictions based on user-provided agricultural parameters. The trained machine learning models analyze the input data and produce accurate yield estimation results.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 05, May 2026)

7.6 Visualization Module

This module displays prediction results using graphs, charts, and statistical visualizations. It helps users understand agricultural trends, prediction accuracy, and model performance more effectively.

VIII. Results And Analysis

8.1 System Performance

The system provides accurate and efficient crop yield prediction using machine learning algorithms. The prediction process is fast and capable of handling large agricultural datasets efficiently.

8.2 Prediction Accuracy

Machine learning models such as Random Forest and Support Vector Machine provide reliable prediction accuracy by analyzing multiple environmental and agricultural factors simultaneously.

8.3 Data Analysis Efficiency

The system efficiently preprocesses and analyzes agricultural datasets using Python libraries such as Pandas and NumPy. This improves data quality and overall model performance.

8.4 User Experience

The system provides a simple and interactive interface where users can easily input agricultural parameters and obtain prediction results with graphical analysis and reports.

IX. Project Implementation

The implementation of the Yield Prediction Using Machine Learning system is carried out using Python, Flask, Scikit-learn, Pandas, NumPy, and MySQL / SQLite. The frontend of the system is developed using HTML, CSS, and JavaScript to provide an interactive and responsive user interface. Users can input agricultural parameters such as rainfall, temperature, humidity, and soil type through the web interface and obtain crop yield predictions.

The backend is implemented using Python and Flask, which handle data preprocessing, machine learning model training, prediction generation, and database communication. Scikit-learn is used to implement machine learning algorithms such as Random Forest, Decision Tree, and Support Vector Machine (SVM) for accurate yield prediction.

The implementation process begins with data collection and preprocessing, where agricultural datasets are cleaned, normalized, and transformed into suitable formats for machine learning analysis.

Feature selection techniques are then applied to identify important agricultural parameters influencing crop productivity.

The model training module trains machine learning models using historical datasets and evaluates prediction performance using accuracy metrics. After training, users can provide new agricultural input values, and the system generates predicted crop yield results instantly.

Visualization libraries such as Matplotlib and Seaborn are used to display graphical analysis and prediction reports. The system is tested using different datasets and prediction scenarios to ensure accuracy, reliability, and efficient system performance.

X. Conclusion

The Yield Prediction Using Machine Learning system provides an effective solution for improving agricultural productivity through intelligent prediction techniques. The system uses machine learning algorithms to analyze agricultural and environmental data and generate accurate crop yield predictions.

The proposed system overcomes the limitations of traditional prediction methods by automating data analysis, reducing manual effort, and improving prediction accuracy. By integrating machine learning models, data preprocessing, visualization techniques, and a user-friendly interface, the system supports better decision-making for farmers and agricultural planners.

The project demonstrates the importance of machine learning technologies in modern agriculture and highlights their potential for improving sustainable farming practices and resource utilization. Future enhancements may include integration with IoT devices, real-time weather forecasting systems, mobile application support, and advanced deep learning techniques for even more accurate prediction results.

JOURNAL REFERENCES

[1] Angelin Rosy M, Chaya M, Felix Xavier Muthu M, "Artificial Intelligence in Machine Learning Techniques for Clustering and Classifications," International Journal of Computer Science and Engineering, vol. 7, no. 17, pp. 71–75, May 2019, ISSN: 2347–2693.

[2] Patel R., & Kumar S., "Machine Learning Techniques for Crop Yield Prediction Using Agricultural Data," International Journal of Agricultural Science and Technology, vol. 14, no. 2, pp. 112–120, 2024.

[3] Sharma P., & Verma R., "Crop Yield Prediction Using Random Forest and Decision Tree Algorithms," Journal of Data Science and Intelligent Systems, vol. 11, no. 4, pp. 210–218, 2023.

[4] Singh A., & Kaur H., "Artificial Intelligence Approaches for Smart Agriculture and Yield Prediction," International Journal of Computer Applications, vol. 182, no. 15, pp. 25–31, 2024.

[5] Zhang Y., Li X., & Chen J., "A Comparative Study of Machine Learning Models for Agricultural Yield Forecasting," Journal of Artificial Intelligence Research, vol. 19, no. 1, pp. 55–67, 2025.

[6] Gupta N., & Sharma D., "Machine Learning Models for Precision Agriculture: A Review," International Journal of Smart Farming, vol. 9, no. 1, pp. 45–53, 2024.

[7] Khan A., & Ali M., "Predictive Analytics in Agriculture Using AI Techniques," Journal of Agricultural Informatics, vol. 13, no. 2, pp. 78–86, 2025.

[8] Iyer S., & Nair V., "Crop Yield Estimation Using Neural Networks and Climate Data," International Journal of Advanced Computing Research, vol. 12, no. 3, pp. 150–158, 2023.

WEB REFERENCES

[1] Python Official Documentation https://docs.python.org/3/

[2] Scikit-learn Documentation https://scikit-learn.org/

[3] TensorFlow Documentation https://www.tensorflow.org/

[4] Pandas Documentation https://pandas.pydata.org/docs/

[5] NumPy Documentation https://numpy.org/doc/

[6] Matplotlib Documentation https://matplotlib.org/stable/

[7] Flask Documentation https://flask.palletsprojects.com/

[8] Kaggle Datasets https://www.kaggle.com/datasets

BOOK REFERENCES

[1] Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition, O'Reilly Media, 2022.

[2] Sebastian Raschka and Vahid Mirjalili, Python Machine Learning, 3rd Edition, Packt Publishing, 2019.

[3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, 1st Edition, MIT Press, 2016.

[4] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[5] Andriy Burkov, The Hundred-Page Machine Learning Book, 1st Edition, Andriy Burkov Publications, 2019.

[6] Kevin P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

[7] François Chollet, Deep Learning with Python, 2nd Edition, Manning Publications, 2021.

[8] Ethem Alpaydin, Introduction to Machine Learning, 4th Edition, MIT Press, 2020.