International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Rainfall-Driven Smart Fertilization System for Sustainable Crop Growth

Dr. Shankar N B[1], Deepthi R Gowda[2], Divya P[2], Gayathri S[2], Keerthi D N[2]

[1]Associate Professor, Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Doddaballapur, Karnataka, India

[2]Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Doddaballapur, Karnataka, India

ARTICLEINFO

Article History:

Accepted : 05 Dec 2025

Published: 20 Dec 2025

Publication Issue

Volume 11, Issue 6

November-December-2025

Page Number

322-334

ABSTRACT

Managing fertilizer application is often a complex challenge for farmers, requiring precise guidance to maximize crop productivity and minimize nutrient waste. While moderate, well-timed precipitation is beneficial for dissolving dry fertilizers and helping nutrients reach the root zone, heavy rainfall can be counterproductive. Excessive moisture often leads to increased surface runoff and the leaching of vital soil components, including primary macronutrients like nitrogen (N), phosphorus (P), and potassium (K), as well as essential micronutrients such as manganese (Mn) and boron (B). This study introduces a sophisticated nutrient recommendation system that utilizes an updated XGBoost algorithm designed for time-series data. By analyzing the relationship between historical rainfall patterns and specific crop fertility requirements, the proposed model provides tailored recommendations to optimize plant growth. Ultimately, this approach seeks to improve soil health and agricultural efficiency by ensuring nutrients are applied in a way that minimizes environmental loss.

Introduction

Agriculture serves as a cornerstone of national economic development, with India notably contributing 17–18% to its GDP and holding the second-highest rank globally for agricultural production. Because crops consistently deplete the topsoil of essential elements, fertilizers are necessary to replenish these nutrients and prevent significant declines in harvest volumes. However, effective fertilization demands high precision; application must be carefully balanced against local rainfall trends and the specific nutritional demands of each plant variety to be effective.

Organic farming plays a key role today as we face climate change, increasing population, and fewer natural resources. While traditional fertilization is necessary for high yields, it frequently results in excessive chemical use, leading to environmental damage and nutrient runoff. To combat these issues, modern agricultural technologies focus on resource optimization. This research presents a Rainfall-Driven Smart Fertilization System, which utilizes predictive weather modeling and environmental sensors to automate and refine nutrient delivery. By synchronizing fertilizer application with forecasted precipitation and soil moisture, the system maximizes plant absorption while significantly reducing the high chance of leaching and ecological harm.

Machine learning offers an effective way to address these challenges by learning from historical soil fertility and climate data. To support farmers with actionable insights, the proposed model utilizes the XGBoost algorithm, requiring only the crop type and geographic location as user inputs. The system then predicts the exact nutrient requirements and identifies the most effective timing for application. To ensure widespread accessibility, the tool is hosted on a platform built with the Flask Python framework, allowing users across various devices to access and share critical agricultural data seamlessly.

LITERATURE SURVEY

The literature survey forms the backbone of this research, offering insights into existing methodologies for Rainfall-Driven Smart Fertilization System for Sustainable Crop Growth. These following studies discusses about the existing model and remarks.

1. Prediction of Crop Fertilizer Consumption the study by Krutika Hampannavar et al. (2018).

The researchers developed a method to pinpoint nitrogen shortages in chili crops, calculate the specific size of the impacted areas, and estimate the necessary fertilizer volume for recovery. The operational flow involves several key stages: capturing digital images, transforming them into grayscale for better analysis, and utilizing histogram data to differentiate between robust plants and those suffering from nutrient gaps. By isolating these distressed regions, the system can then find the correct amount of supplementation required to restore plant health. (a kind of Digital Image Processing).

The significant advantage of this paper is: it provides a cost-effective and efficient solution for detecting nitrogen lack in plants using basic tools like a smartphone and standard image processing techniques. Additionally, it contributes to environmental sustainability by minimizing the overuse of chemical fertilizers, thus protecting soil health and nearby water sources.

Despite its advantages, the system has several limitations. It is currently restricted to detecting only nitrogen deficiency, leaving out other essential nutrients such as phosphorus and potassium. Another drawback is the lack of integration with environmental data like rainfall, temperature, or soil pH, all of which can influence nutrient uptake and fertilizer effectiveness.

To maximize agricultural output and minimize resource loss, it is essential to provide growers with precise recommendations for the most effective use of fertilizers. Several variables influence how much fertilizer is needed, including the total available acreage, the specific crops being grown, local precipitation levels, and the inherent properties of the soil.When plants show signs of nitrogen depletion, common corrective treatments include the application of urea, ammonium nitrate, anhydrous ammonia, or ammonium sulphate. By accurately forecasting fertilizer requirements, farmers can prevent crop damage from the risks of both nutrient deficiency and chemical toxicity, ultimately ensuring a healthy harvest while avoiding unnecessary expenses. [1].

2. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters by M.S. Suchithra et al.(2019).

The study focuses better predicting the nutrient levels in the soil nutrient classification in the North Central Laterite region of Kerala, India, a key step for reducing wasteful expenditure on fertilizers and enhancing soil health. The research tackles five critical classification problems using soil test report values: the village-wise fertility indices for Organic Carbon (OC), Phosphorus (P), Potassium (K), and Boron (B), with Soil Reaction (pH). To achieve faster and more accurate results, the authors employ the Extreme Learning Machine (ELM), an efficient learning algorithm for feedforward neural networks, and optimize its meta-parameters. The optimization centered on tuning the number of hidden neurons and comparing the efficiency of five different activation functions, including gaussian radial basis and hyperbolic tangent.

The core identifying the process of finding the optimal ELM configuration for each classification task. Through cross-validation, the optimal number of hidden neurons was set at 50 for the soil nutrient classifications (OC, P, K, B) and 150 for the pH classification. Critically, the best-performing activation function varied by problem: the gaussian radial basis function (elm_grbf) achieved the highest accuracy for the four soil fertility indices (OC-F: 83.66%, P-F: 90%, K-F: 78.43%, B-F: 88.23%). In contrast, the hyperbolic tangent function (elm_tanh) proved superior for the final classification objective, achieving the highest accuracy for Soil pH (88.59\%).

In conclusion, the prediction successfully demonstrates that by customizing the ELM parameters the specific creation of the agricultural problem, a suitable and accurate estimation model for village-wise soil fertility can be created. This work provides an effective, data-driven analytical tool that could be used to given a machine learning decision system, helping the local government manage nutrient deficiency problems and guide farmers in making accurate decisions about fertilizer application, ultimately improving soil and environmental quality. Overall, the paper emphasizes that such data-driven approaches have the potential to significantly strengthen sustainable soil management and modernize agricultural decision-support systems in nutrient-sensitive regions[2].

3. Crop Yield Prediction Based on Indian Agriculture using Machine Learning the study by Potnuru Sai Nishant et al. (2020).

This work introduces a straightforward approach to crop yield prediction in India by leveraging commonly available agricultural data, allowing accurate yield estimation for any selected year.

The method employed involves collecting agricultural data such as crop type, season, rainfall, temperature, soil type, and location, and using machine learning algorithms like Linear Regression and Decision Tree Regressor to predict crop yields. The model offers an easy-to-use interface that allows users to enter relevant parameters and obtain predicted results, helping them make better and more informed farming decisions.

The primary benefit of this approach lies in its ability to improve agricultural planning, reduce crop failure, and support food security through data-driven insights. Additionally, it enables cost-effective resource management and can be easily adapted to various crops and regions. The model also analyzes past agricultural data collected from various states and regions to estimate crop yield for a specific location.

However, its effectiveness depends on the quality and volume of data available, and the model might be less accurate in areas with inconsistent or incomplete records. Additionally, the paper does not take into consider of several factors like local real-time weather conditions and other parameters which are contingent on the local area.

This study employs a stacked regression approach. Initially, the full training dataset is divided into two parts. The chosen base models are trained on the first part and evaluated on the second (holdout) set. The predictions from this evaluation phase are then used as input features for a higher-level model, known as the meta-model. The final prediction is obtained by averaging these outputs and treating them as meta-features. In this approach, the Lasso Regressor serves as the meta-model. [3].

4. Farmers' risk preference and fertilizer use the study by QIAO Fang-bin et al. (2020).

This study presents a new agricultural decision-support system that leverages machine learning (ML) models to address food security challenges and enhance crop yield predictions in East African countries. The system combines data from three main sources—climate, crop production, and pesticide usage—to develop a predictive tool for farmers and policymakers. It is designed to estimate annual yields for four key crops—potatoes, beans, tea, and coffee—across fourteen East African countries.

The development process began with integrating multiple datasets, followed by preprocessing steps such as data cleaning, handling missing values, and performing exploratory and statistical analyses to better understand the data. Feature engineering methods, including encoding and normalization, were applied to prepare the data for the ML models.

Three supervised ML models were trained and evaluated for regression tasks: Crop Random Forest (CRF), Crop Gradient Boosting Machine (CGBM), and Crop Support Vector Machine (CSVM). The results showed that all three models were reliable and generalizable across the East African region. Among them, CRF achieved the highest accuracy, with an R2R^{2}R2 of 92.27% and a minimal Root Mean Square Error (RMSE) of 0.343, outperforming CGBM(R2=90.19%R^{2}=90.19\%R2=90.19%) and CSVM(R2=86.38%R^{2}=86.38\%R2=86.38%). The authors suggest that averaging the predictions from all three models provides a robust estimate.

For future improvements, the study recommends incorporating additional data features, such as agricultural water usage, wind patterns, pollution levels, meteorological variations, local animal species, and agricultural economic indicators, to further enhance the model's accuracy and applicability across the region[4].

5. A nutrient recommendation system for soil fertilization based on Evolutionary Computation the study by Usman Ahmed et al.(2021).

This paper develops a model to help use nutrients like nitrogen (N), phosphorus (P), and potassium (K) correctly. The model supports a knowledge-based system in an ICT environment and provides recommendations to maintain soil fertility and improve crop production.

The authors developed a nutrient recommendation system that leverages an improved Genetic Algorithm (IGA) to determine optimal fertilizer combinations tailored to specific soil conditions. The methodology involves encoding potential nutrient solutions as chromosomes and employing evolutionary operations like selection, crossover, and mutation to evolve these solutions over successive generations. This paper proposes nutrient recommendations through an improved genetic algorithm (GA) that uses time-series sensor data. One significant advantage of this system is its adaptability; the evolutionary computation framework can efficiently locate complex, multidimensional search spaces to identify effective fertilization plans that might be overlooked by traditional methods. Additionally, the system's ability to integrate various constraints and objectives makes it highly customizable for different agricultural scenarios.

However, a notable limitation of this approach is the computational intensity involved in evolutionary algorithms. These algorithms often require significant processing power and longer execution time, particularly when applied to large-scale systems. Additionally, the overall performance of the system is highly dependent on the quality and accuracy of the input data. Any errors or inconsistencies in soil or crop information may result in less reliable or suboptimal fertilizer recommendations.

Despite these challenges, The proposed system is a promising step toward achieving more sustainable and efficient agricultural practices. Each sensor in the remote area has its set of nutrient levels. The sensor readings are saved on the local system and are periodically uploaded to the database at regular intervals such as weekly, monthly, or yearly. Data from sensors measuring each nutrient are gathered and integrated through the Internet, creating a comprehensive dataset from the remote location. The developed algorithm is then applied to analyze this data and optimize the nutrient sequence to support effective decision-making[5].

6. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters the study by M.S. Suchithra et al.(2022).

The study by M.S. Suchithra and Maya L. Pai (2022) on optimizing Extreme Learning Machines (ELM) for soil nutrient classification found that the optimized ELM model achieved improved prediction accuracy. By fine-tuning the hidden layer size and selecting suitable activation functions, the model demonstrated high computational efficiency and faster training. This approach can support precision agriculture by enabling informed fertilizer management and reducing reliance on chemical soil tests. However, the study's results may be limited to the specific geographic region, and the ELM model's black-box nature may limit interpretability. Overall, the optimized ELM model shows promise for enhancing soil nutrient classification and contributing to sustainable farming practices.

This study introduces an approach aimed at improving soil nutrient classification by using Extreme Learning Machines (ELM). Soil samples were collected from the Davangere district, and key parameters such as available phosphorus, potassium, organic carbon, boron, and pH were analyzed.

One major advantage of this methodology is its high computational efficiency and faster training compared to traditional neural networks, making it suitable for large-scale agricultural datasets. The optimized ELM model also enhanced prediction accuracy, allowing the support more informed and cost-effective fertilizer management. Additionally, this technique aids in precision agriculture by minimizing the reliance on chemical soil tests.

However, a notable disadvantage is the sensitivity of the ELM model to its parameter settings, which requires careful tuning to avoid poor performance. Furthermore, the study's results may not applied well outside the specific geographic region where the data was collected. Another drawback is the model's interpretability, as neural networks like ELM can act as black boxes, offering limited insight into the reasoning behind their predictions[6].

7. Crop Yield Prediction based on Indian Agriculture using Machine learning by Siddana Gowda S M et al.(2022).

The paper focuses on predicting crop yield and recommending suitable crops for Indian agriculture, which is considered the backbone of the country's economy. The work aims to assist farmers, especially those who are dilettante, in choosing the right crop bases of soil requirements to avoid setbacks in productivity. This system attempts to be novel by using simple factors such as State, district, season, and area to estimate the crop yield for a desired year. The advanced regression techniques employed for yield estimation include Kernel Ridge, Lasso, and ENet algorithms, which are enhanced using Stacking Regression for better accuracy. Furthermore, the system is designed to provide fertilizer suggests based on soil parameters to achieve greater precision in yield prediction.

The system uses a data mining approach and a Support Vector Machine (SVM) Classification algorithm. This approach improves the efficiency of the Crop Recommendation System by analyzing soil and crop data to predict a suitable list of crops for cultivation. And also details about nutrients, which are deficient in the soil for a particular crop, allowing the user to make the final decision on what to sow. The system's architecture takes inputs such as pH value and site/location from the user. This data is processed by controllers to determine factors like weather, temperature, soil type, nutrient value, rainfall, and soil alkalinity (based on pH value and nutrient percentages like N, P, K, etc.). These results are then compared with predefined "nutrients" and "crop" data stores to ultimately predict the crop yield and provide fertilizer recommendations if needed. The data set for analysis includes parameters like precipitation, temperature, area, production, and yield for the seasons from January to December for the years 2000 to 2018. Farming is a major part of India's economy, and many farmers struggle to choose the right crop due to limited knowledge about soil and environmental requirements. The study presents a machine-learning-driven system designed to predict crop yield using basic inputs such as state, district, season, and cultivated area. By applying advanced regression techniques, including Kernel Ridge, Lasso, and ElasticNet, and improving them through a stacking approach, the system achieves greater prediction accuracy. The system also recommends suitable crops and fertilizer ratios by analyzing soil pH, nutrient content, weather conditions, and rainfall using external APIs.

The literature review shows that several machine learning techniques—such as K-Means, KNN, SVM, Bayesian networks, and decision trees—have been applied in agriculture for soil segmentation, weather forecasting, pest analysis, and yield prediction. Building on these studies, the proposed model integrates soil data, environmental parameters, and crop requirements to generate precise yield predictions. Diagrams in the paper illustrate how the system processes user inputs through multiple controllers to compare soil characteristics with predefined crop-nutrient datasets.

In conclusion, the paper emphasizes that accurate crop selection is essential for improving productivity and supporting farmers financially. The proposed model enables farmers to make better decisions by predicting crop yield and offering fertilizer recommendations. The authors suggest improving the system further by expanding datasets and incorporating additional soil attributes to enhance prediction accuracy[7].

8. Eco-growth: Sustainable Fertiliser Solutions the study by Yash Pabari et al.(2024).

The paper finds that leveraging machine learning—specifically the Random Forest Regression algorithm—can significantly improve the precision of fertilizer application in agriculture. By analyzing key environmental and crop-specific variables such as temperature, humidity, rainfall, land area, and crop type, the model accurately predicts the required quantities of essential nutrients like nitrogen, phosphorus, and potassium. These predictions are then used to determine the optimal dosages of common fertilizers such as urea, DAP, and potassium chloride. The model demonstrated high predictive accuracy, with low error margins, and was further enhanced through the integration of real-time weather data to issue timely alerts for extreme conditions like heavy rainfall. These findings highlight the potential of AI-driven tools in reducing environmental degradation caused by fertilizer runoff and in improving agricultural productivity by promoting efficient and sustainable nutrient management practices.

The system integrates real-time weather data through an API and is deployed as a web application using Flask, making it useful for farmers on multiple platforms. Data preprocessing metods such as outlier removal and normalization ensure the robustness of the model.

The main benefits of the proposed method include improved accuracy in nutrient recommendations, reduced fertilizer wastage, better soil fertility, and minimized environmental risks such as nutrient runoff and leaching. It also supports decision-making in agriculture, helping farmers boost productivity efficiently.

However, the system does have certain limitations. Its effectiveness depends heavily on the accuracy of weather data, and its reliance on digital infrastructure may pose challenges in rural areas. Additionally, the model does not consider pest or disease factors, which can also influence crop nutrient requirements.

Overall, the study promotes sustainable agricultural practices by optimizing fertilizer use, reducing environmental impact, and improving crop productivity and farmer profitability[8].

METHODOLOGY

3.1 Data Collection: The dataset containing soil parameters, crop types, and corresponding NPK values is collected from reliable agricultural sources. This data serves as the foundation for training the Machine Learning model.

3.2 Data Preprocessing: The collected dataset is cleaned by removing missing, duplicate, and inconsistent values. Categorical features such as crop type and soil type are converted into numerical form using label encoding. Feature selection and normalization are also applied to improve model performance.

3.3 Model Training: The cleaned dataset is utilized in two main modules: the User Interface and the training of Machine Learning models for predicting Nitrogen (N), Phosphorus (P), and Potassium (K) values. The XGBoost regression algorithm is used due to its high prediction accuracy and performance. Separate regression models are trained for each nutrient.

3.4 Model Evaluation: The trained models are tested using standard evaluation metrics such as:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- R^2 Score

These evaluation metrics provide insights into the accuracy and reliability of the predicted outcomes.

3.5 Model Deployment: The trained models are saved using Pickle and integrated into a Flask-based web application. This allows real-time prediction of NPK values based on user inputs.

3.6 User Input Processing: The user enters crop type, soil type, soil pH, and location through a web interface. The input values are encoded and converted into machine-readable format before being passed to the ML model.

3.7 NPK Prediction: The processed input values are fed into the trained ML model to predict the optimal NPK fertilizer values required for the selected crop.

3.8 Weather Forecast Integration: A Weather API is integrated to fetch real-time 3-day weather forecast based on the user's location. This helps in providing rainfall-based irrigation advice.

3.9 Result Visualization: The predicted NPK values and weather information are displayed on the dashboard using interactive charts and tables for better user understanding.

3.10 Final Recommendation: Based on the rainfall forecast, the system generates irrigation advice to help farmers make better farming decisions.

3.11 System Architecture The overall system architecture of the proposed smart fertilizer recommendation system is designed using a modular approach to ensure scalability, efficiency, and ease of integration. The system consists of four major modules, namely the User Interface (Frontend), Backend Server (Flask Framework), Machine Learning Prediction Engine, and the Weather Forecasting Module using a Weather API. These modules work together seamlessly to provide real-time NPK prediction and rainfall-based irrigation advisory to the users.

The process begins when the user enters the required inputs such as crop name, soil type, soil pH, and location through the web-based user interface developed using HTML, CSS, and JavaScript. These inputs are transmitted to the Flask backend server through secure HTTP POST requests. Once the backend receives the input data, it performs preprocessing by encoding the crop and soil parameters using pre-trained label encoders. The encoded input is then passed to the trained machine learning prediction engine, where the NPK regression models process the input features and generate the predicted Nitrogen (N), Phosphorus (P), and Potassium (K) values.

Simultaneously, the backend server communicates with the Weather Forecasting Module by making API requests to fetch real-time weather data based on the user's location. The system retrieves critical environmental parameters such as rainfall and temperature. Using the predicted rainfall values, the system generates an irrigation advisory by analyzing whether irrigation is safe or should be avoided to prevent nutrient loss and water wastage. Finally, the predicted NPK values, real-time rainfall forecast, temperature data, and irrigation advisory messages are sent back to the frontend, where they are displayed clearly on the dashboard along with visual charts for better understanding. This integrated architecture ensures real-time decision support for farmers and promotes efficient, data-drivenagricultural practices.

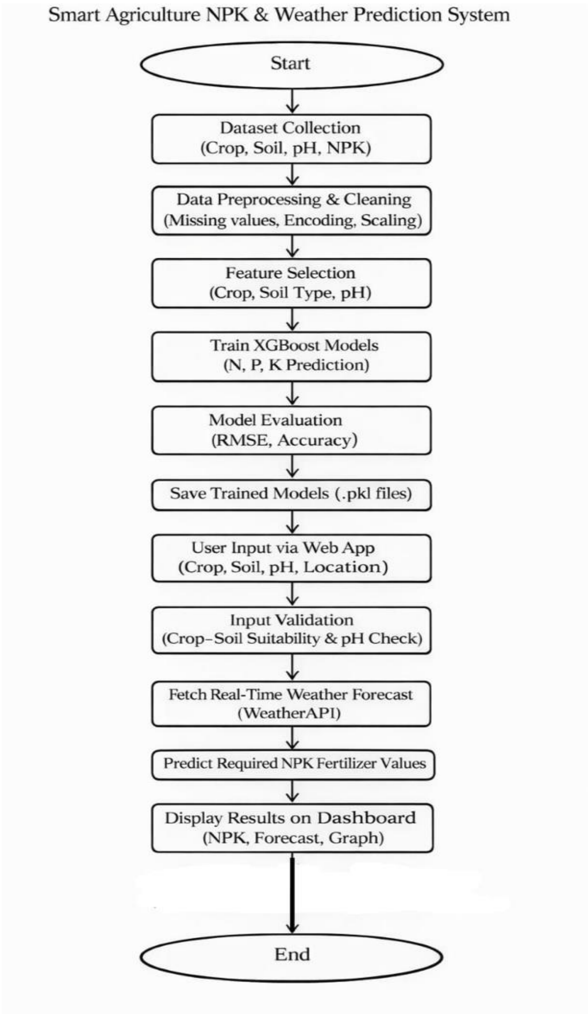

Fig. 1. System Design Flowchart

IMPLEMENTATION

4.1 Workflow

The proposed Smart Agriculture system follows a structured workflow to predict essential soil nutrients and provide weather-based decision support for farmers.

The workflow begins with dataset collection, where historical agricultural data consisting of crop type, soil type, soil pH, and corresponding nitrogen (N), phosphorus (P), and potassium (K) values are gathered from reliable sources. This dataset forms the foundation for training the prediction models.

Next, data preprocessing and cleaning is performed to improve data quality. This stage includes handling missing values, encoding categorical variables such as crop and soil types into numerical form, and scaling numerical features where required. Preprocessing ensures that the dataset is suitable for machine learning model training.

Following preprocessing, feature selection is performed to identify the most influential input parameters. In this system, crop type, soil type, and soil pH are selected as primary features for making predictions with the model required NPK values.

The system then proceeds to model training, where separate XGBoost regression models are trained for predicting nitrogen, phosphorus, and potassium values. XGBoost is chosen due to its high accuracy, ability to handle non-linear relationships, and robustness against overfitting.

After training, model evaluation is performed using performance metrics such as Root Mean Square Error and accuracy to assess prediction reliability. Once satisfactory performance is achieved, the trained models are saved as serialized (.pkl) files for reuse during deployment.

In the deployment phase, the user provides input through a web application, including crop type, soil type, soil pH, and location. Before prediction, the system performs input validation, ensuring that the soil pH lies within the valid range (0–14) and that the selected crop is suitable for the chosen soil type. This validation step prevents agronomically invalid predictions.

The system then fetches real-time weather forecast data using a Weather API based on the user's location. This enables the application to provide rainfall-aware recommendations.

Using the validated inputs, the trained XGBoost models predict the required NPK fertilizer values. The predicted nutrient values are then combined with weather insights to generate irrigation-related advice.

Finally, the outcomes are displayed on the dashboard, including predicted NPK values, weather forecast details, irrigation suggestions, and graphical visualizations. This enables farmers and other users to make informed decisions that promote efficient and sustainable farming practices. The workflow concludes at this stage.

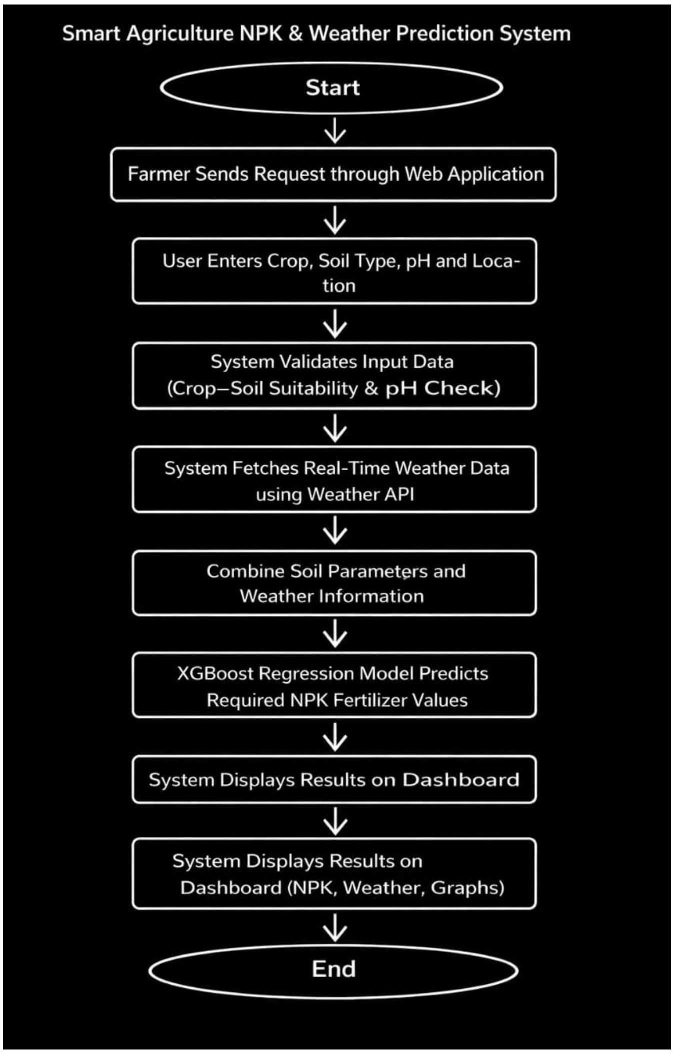

Fig. 2. Overall Workflow

4.2 Technologies used

1. **Python** – Core language for backend development, preprocessing, and model execution.
2. **NumPy & Pandas** – For numerical computation, dataset handling, and CSV operations.
3. **Scikit-learn (Random Forest Classifier)** – Used for building, training, and deploying the crop predictionmodel.
4. **Flask** – Backend framework providing REST APIs connecting the UI, model, datasets, and weather API.
5. **Requests Library** – Used to fetch real-time data from the OpenWeatherMapAPI.
6. **OpenWeatherMap API** – Provides temperature and rainfall data for location-based predictions.
7. **HTML, CSS, JavaScript** – Used for building the farmer and tester user interfaces.
8. **CSV-basedRuntimeStorage** (farmerrequests.csv, linked-tester-data.csv, predictions.csv, feedback.csv) – Maintains logs for requests, tester inputs, predictions, and feedback.
9. **Model Evaluation Metrics** -RMSE (Root Mean Square Error), Accuracy (for suitability validation)These metrics are used to assess the execution of the trained models.
10. **Model StoragePickle (.pkl files)**- Used to save trained machine learning models and enables reuse of models without retraining
11. **Deployment Environment**- Localhost / Flask Development ServerUsed for testing and demonstration of the system

RESULT AND DISSCUSSIONS

A. Performance Matrix

To analyze the performance of the proposed rainfall-driven smart fertilization system for sustainable crop growth, five widely-used machine learning algorithms were evaluated by Decision Tree (DT), Random Forest (RF) and XGBoost. The performance comparison according to the accuracy achieved from the trained model on the integrated agricultural dataset.

Table 1. Comparison of Classification Accuracies

Algorithm	Accuracy(%)
Decision Tree	87.88
Random Forest	92.22
XGBoost	92.34

From the performance matrix, it is proven that the Random Forest classifier achieves the highest accuracy among all evaluated models.

Decision Tree

The Decision Tree classifier achieved an accuracy of 87.88%. This model works by recursively partitioning the dataset based on feature values to form a tree-like structure. Although Decision Trees are easy to interpret and computationally efficient, they are highly prone to overfitting, especially when dealing with complex and high-dimensional agricultural data such as soil nutrients, temperature, rainfall, and humidity. Minor variations in inputs taken by user can affect the tree structure, leading to reduced generalization performance.

Random Forest

Random Forest model achieved an accuracy of 92.22%, significantly outperforming the standalone Decision Tree. Random Forest is a machine learning method that builds several decision trees from randomly sampled subsets of the data and combines their outputs using a majority voting process to produce the final prediction.

Its superior performance can be attributed to:

- Ensemble bagging strategy, which reduces model variance
- Robustness to noisy and missing data
- Ability to handle multi-dimensional agricultural features effectively
- Reduced overfitting compared to a single Decision Tree

These properties make Random Forest a reliable and stable model for crop recommendation tasks involving complex environmental and soil interactions.

XGBoost

XGBoost recorded the highest accuracy of 92.34%, marginally outperforming Random Forest. XGBoost is a powerful gradient boosting algorithm that builds trees sequentially, where each new tree corrects the errors of the before done tress. Its strong performance is due to:

- Gradient boosting framework, enabling better error minimization
- Regularization techniques, which control overfitting
- Efficient handling of non-linear relationships
- Optimized computation and feature importance learning

XGBoost is particularly effective in capturing intricate patterns in agricultural datasets, making it well-suited for predictive modeling in crop recommendation systems.

B. Graphical Representation

A graphical comparison of the NPK ratio and rainfall pattern is shown in Fig. 3 and Fig. 4 respectively.

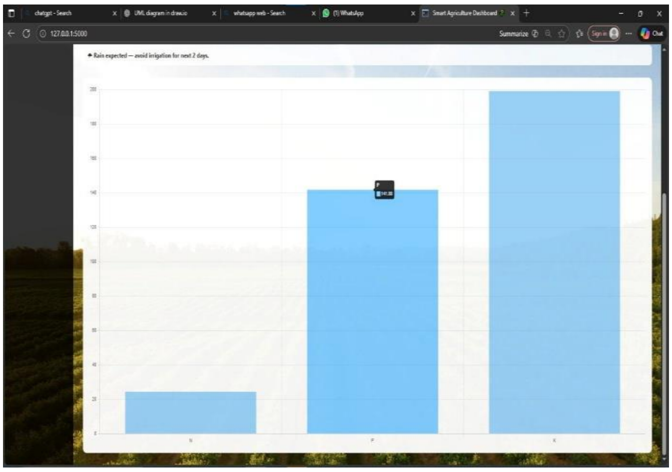

Fig. 3. NPK ratio

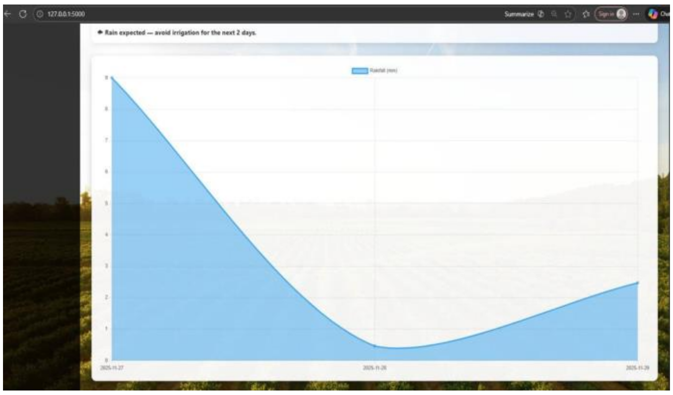

Fig.4. Rainfall pattern

The graphical outputs in this project focus on the analysis of soil nutrients (NPK values) and rainfall patterns, which serve as the basic data input to the ML models. These visual representation help in understanding the relationship between soil fertility, climatic conditions, and crop suitability, rather than comparing algorithm accuracies.

NPK Nutrient Distribution Analysis Graphs representing Nitrogen (N), Phosphorus (P), and Potassium (K) values illustrate the variability of soil nutrient composition across different samples. These plots assist in:

- Identifying nutrient-rich and nutrient-deficient soil conditions
- Understanding nutrient balance required for optimal crop growth
- Supporting feature selection for machine learning models

The variation observed in NPK levels confirms the need for intelligent models capable of handling multi-dimensional agricultural data.

Rainfall Pattern Analysis

Rainfall distribution graphs depict seasonal and regional rainfall variations that significantly influence crop yield and suitability. These visualizations help:

- Capture the impact of rainfall on crop recommendation
- Highlight patterns such as low, moderate, and high rainfall zones
- Support the learning process of models such as Random Forest and XGBoost
- Rainfall analysis is critical as it interacts with soil nutrients to affect crop growth conditions.

Graph Interpretation

The NPK and rainfall visualizations provide strong exploratory data analysis (EDA) support for the proposed system. They demonstrate:

- Clear variability in agricultural parameters
- Non-linear relationships between features
- The necessity of advanced ensemble models for accurate prediction

CONCLUSION

The Flask application is designed to predict the required NPK (Nitrogen, Phosphorus, and Potassium) levels for different crops based on soil type, soil pH, and crop selection. It uses a trained XGBoost model along with encoded values for both soil and crop types. When the user submits the form, the application encodes the crop using a label encoder and reads the soil type directly from the numeric mapping. These processed inputs are then passed to the model to generate accurate NPK recommendations.

In addition to nutrient prediction, the application also integrates a weather forecasting feature using the WeatherAPI service. When a user enters a location, the app retrieves a 3-day weather forecast, including average temperature, expected rainfall, and general weather conditions. This weather data helps the system analyze the total predicted rainfall.

Based on the rainfall forecast, the application provides intelligent farming advice. If the total rainfall expected over the next three days is more than 5 mm, it advises the farmer to avoid irrigation for the next two days. If rainfall is low, the system informs the user that it is safe to irrigate. This helps farmers plan irrigation more efficiently and avoid over-watering.

The application also preserves user inputs such as selected crop, soil type, pH value, and location, ensuring that the form remains filled after submission for better user convenience. Additionally, error handling is implemented to prevent the app from breaking if any issue occurs, such as invalid input or an API failure. Finally, the corrected Flask execution condition ensures the application runs smoothly when executed, completing a fully functional and interactive smart farming support system.

FUTURE ENHANCEMENT

Although the proposed Rainfall-driven smart fertilization system for sustainable crop growth performs efficiently using machine learning several enhancements can significantly broaden its applicability, accuracy, and usability. The following future improvements are recommended:

1. Inclusion of Soil Health Parameters

Additional soil attributes such as soil moisture, organic carbon, electrical conductivity, and micronutrient levels can be incorporated for more precise fertilizer recommendations.

2. Crop Growth Stage-Based Recommendation

Fertilizer suggestions can be customized according to different crop growth stages such as sowing, vegetative, flowering, and harvesting stages.

3. Long-Term Weather and Rainfall Prediction

Extending the system to support weekly and monthly rainfall forecasting will help farmers in seasonal crop planning.

4. IoT Sensor Integration

Real-time data from soil sensors (moisture, pH, temperature) can be integrated to enable continuous monitoring and dynamic fertilizer recommendations.

5. Multilingual Support

Adding regional language support will make the system more user-friendly and accessible to a wider farming community.

6. Government and Market Data Integration

Integrating government soil health card data and fertilizer market costs will help farmers make cost-effective decisions.

References

[1]. K. Hampannavar, A. N. Madhu, and S. R. Patil, “Prediction of Crop Fertilizer Consumption Using Image Processing,” International Journal of Engineering Research & Technology (IJERT), vol. 7, no. 6, pp. 1–5, 2018.

[2]. M. S. Suchithra and M. L. Pai, “Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters,”Information Processing in Agriculture, vol. 7, pp. 72–82, 2020. doi: 10.1016/j.inpa.2019.05.003.

[3]. P. S. Nishant, N. S. Kumar, M. S. Tejaswini, and K. D. Reddy, “Crop Yield Prediction Based on Indian Agriculture Using Machine Learning,” International Journal of Research in Engineering and Technology (IJRET), vol. 9, no. 4, pp. 234–240, 2020.

[4]. F.-b.Qiao and J.-k. Huang, "Farmers' risk preference and fertilizer use,"Journal of Integrative Agriculture, vol. 20, no. 7, pp. 1987–1995,2021. doi: 10.1016/S2095-3119(20)63450-5.

[5]. U. Ahmed, S. Q. Ali, and M. A. Khan, “A Nutrient Recommendation System for Soil Fertilization Based on Evolutionary Computation,” IEEE Access, vol. 9, pp. 112340–112352,2021.

[6]. M. S. Suchithra and M. L. Pai, “Improving the Prediction Accuracy of Soil Nutrient Classification by Optimizing Extreme Learning Machine Parameters,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 5, pp. 210–218, 2022.

[7]. Siddana Gowda S M, Mounika C H, Abhishek N, Swathi N M, Prof. Sheela B P, and Dr. B. Sreepathi, “Crop Yield Prediction based on Indian Agriculture using Machine Learning,” Int. J. Adv. Res. Sci., Commun. Technol. (IJARSCT), vol. 2, no. 2, pp. 53–57, 2022, doi: 10.48175/IJARSCT-5781.

[8]. Y. Pabari, R. Mehta, A. Shah, and K. Patel, "Eco-Growth: Sustainable Fertiliser Solutions Using Machine Learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), vol. 10, no. 1, pp. 45–52, 2024.