Rainfall Crop Advisory (Location-Based) System Using Machine Learning

D. Nirosha[1], M. Surya Bhupal Rao[2], *Kalugotla Sairam[3], Desireddy Sai Ram[4], Mulinti Kalam[5], Saginala Rahul Vishal[6]

[1, 3, 4, 5, 6]Department of Computer Science and Engineering (Data Science), Santhiram Engineering College, Nandyal, Andhra Pradesh, India
[2]Department of Computer Science and Engineering (AI & ML), CVR College of Engineering, Hyderabad, Telangana, India
[1]nirosha.ds@srecnandyal.edu.in
[2]suryabhupal@gmail.com
[3]22x51a3263@srecnandyal.edu.in
[4]22x51a3265@srecnandyal.edu.in
[5]21x51a3232@srecnandyal.edu.in
[6]21x51a3240@srecnandyal.edu.in

Abstract. Because soil, climate, and local environment affect growing conditions to varying degrees, it has always been difficult to determine what crops to plant and how to predict a harvest's yield. Traditional crop-selection and yield-prediction tools have relied on fixed data and have therefore been limited in their applicability at the field level. To address the reliance of traditional agriculture on static tools and processes, a location-based crop advisory system is developed using machine learning models such as Random Forest, Gradient Boosting, and Support Vector Machine (SVM). Machine learning models such as Random Forest, Gradient Boosting, and Support Vector Machines are used for prediction and have been matched to archived farming records as well as local climatological patterns for each specific location. Experimental results show that Random Forest and Gradient Boosting achieved over 98% accuracy. These models use environmental and geographical data to improve prediction accuracy, and both Random Forests and Gradient Boosting achieved over 98% accuracy in yield prediction testing.

Keywords: Crop Recommendation, Machine Learning, Yield Estimation, Climate Analytics, Soil Nutrients, Rainfall Prediction.

1 Introduction

Agriculture is the backbone of India's economy. But still, agriculture in India continues to face many challenges that it has always faced. Climate extremes, land degradation, and a lack of modern decision support systems that apply science to support producers regarding their production decisions at the grower level [10][11] are the main sources of these challenges facing agriculture, which has made it more difficult to develop appropriate timing for planting and has resulted in a decrease in both grower productivity and financial security. In recent years, data and analytical modeling technologies have

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_26

allowed for the creation of intelligent systems that integrate soil nutrients (N, P, K), temperature, humidity, and rainfall data using machine learning to allow growers to better predict the most effective times to plant crops [2][18]. Unfortunately, most existing crop recommendation models do not account for the spatial and climate differences that exist within the Indian districts [14][20]; therefore, they provide general recommendations rather than recommendations for specific locations [22]. To address the restrictions imposed by existing systems, this project will create a multi-dimensional data-driven agricultural recommendation and crop yield prediction framework utilizing AI-based technologies. The machine learning models that will feed into this predictive analytics model include the following: Random Forest, Gradient Boosting, and SVM. The availability of large historical datasets and improved analytical methods has enabled the development of accurate machine learning models. Recent advancements in machine learning have enabled the development of intelligent data-driven systems for agricultural prediction [1], [2]. Existing crop recommendation models often fail to consider region-specific climatic and geographical variations within India, leading to less accurate predictions.

2 Literature Review

The authors [1] and [2] revealed that machine learning techniques such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting have been widely used in recent studies because they could discover unseen relations between the earth's systems, climate changes and yield potential. This is in keeping with the more recent work [3], [18], [19], showing that technology is advancing toward smart applications such as Random Forests, SVMs, and multi-layer neural networks. Studies such as [1], [2], and [3] have explored how climate change impacts agriculture using various modeling systems. Studies conducted [4] showed the impact of climate change on crop yield [5], utilizing Random Forest modeling, again yielding accurate results. For instance, [14] uses SVMs and Random Forests to demonstrate how temperature and rainfall impact agricultural yields. Another case of a machine learning application in agriculture is a Localized Climate Forecasting (LCF) tool developed. The authors [15] that predicts rainfall. Their model was relatively accurate, although their application had many limitations.

Improvements in agronomy are anticipated to facilitate the development of more varied and adaptable farmers by using more advanced farming methods that utilize predictive analytics and Artificial Intelligence (AI), enabling farmers to adapt to changing weather conditions [18],[19]. Regardless of the method used to determine the characteristics of the growing environment, Ramesh and Rajesh developed an AI-based system that uses soil and growing conditions to produce recommendations for the optimum crop based on the area's soil characteristics [12]. The research has synthesized prior studies and developed a data-processing process that incorporates location information related to the location from which the data points were derived. The system incorporates rainfall measurements on a state-by-state or district-by-district basis and provides a location based on specified values [15],[20]. Data about conditions is continually updated by statistical sources that provide updates as conditions change. Space has been treated as the most important consideration, not as a separate feature. This process integrates forecast models with predictive models and real-time conditions to demonstrate improvement in results. The approach uses a protection system in conjunction with this process. Previous work demonstrated limitations of prior models due in part to the use of static values for weather or climate conditions and to large assigned value ranges across large geographic areas. This process addresses those limitations. Crop producers in many regions of India face variable weather conditions that affect crop production. This approach can assist those producers by providing a variable response to each weather event. The approach produces reliable and verifiable results and can be used by crop producers while still providing a multi-area implementation using an integrated model [14],[18],[20]. Recent research has focused on improving agronomic crop production prediction systems by using machine learning, socio-economic variables, and geographic and climatic data [1],[6],[18]. These models use a combination of soil nutrition, weather, and location-specific rain to generate accurate crop recommendations for a given region. A system that is aware of this information allows farmers to make data-driven decisions in their agriculture and increase their productivity [16],[17],[18].

3 Data Collection and Preprocessing

Collecting data means compiling information on agriculture and the weather, and it consists of summarizing agricultural and weather (climate) information. Examples include soil nutrient levels like Nitrogen, Phosphorus, and Potassium (N, P, K), pH, temperature, humidity, and rainfall amounts. The main dataset used in this research is the Crop Recommendation Data Set (2,200 records), which provides information on different crop varieties and environmental characteristics (e.g., crop type and soil nutrients) [12]. However, this dataset does not include information on how rainfall varies across different locations in India, which is needed to use the Crops Recommendation Data Set for Prediction. Hence, two other datasets have been utilized: one is a district-wise rainfall record, and the other is district-level latitude and longitude information, providing information on geography and climate so that the Crops Recommendation Data Set can be matched to the relevant climatic and geographic conditions [15][20]. As shown in Fig. 1, the dataset exhibits a balanced distribution of crop samples across categories.

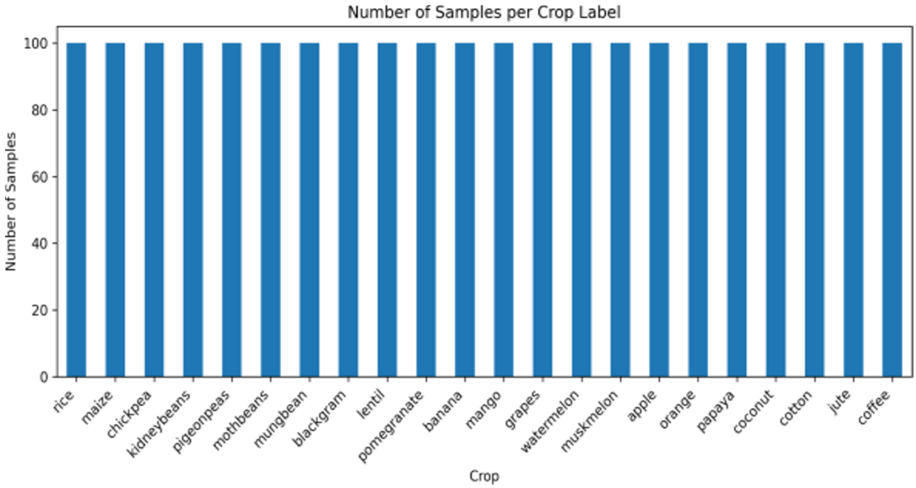

Fig. 1. Crop Sample Distribution

Filled in those gaps using official records, which provided District Wise Rainfall Normal.csv and rainfall_lat_long.csv, were added to [15][20] (with average monthly rainfall, total yearly rainfall, and exact map point of the district), allowing you to place each instance of a crop into a specific site of a crop. Crops need to be organized into categories and spread out evenly (balanced) over the types of crops; otherwise, they will not produce accurate results and may lose their ability to be balanced. Each of these crops contributes to producing a sufficient number for patterns to be identified. Each of the evenly distributed crops also assists the algorithms in determining how nutrients are useful to certain plants. Random Forest, Gradient Boosting, and SVM were each evaluated on the tests, allowing the models to assess how well they performed on new data (to improve the flow of determining accuracy for each crop) compared to the training data. Fig. 2 shows the distribution of N, P, and K across the data.

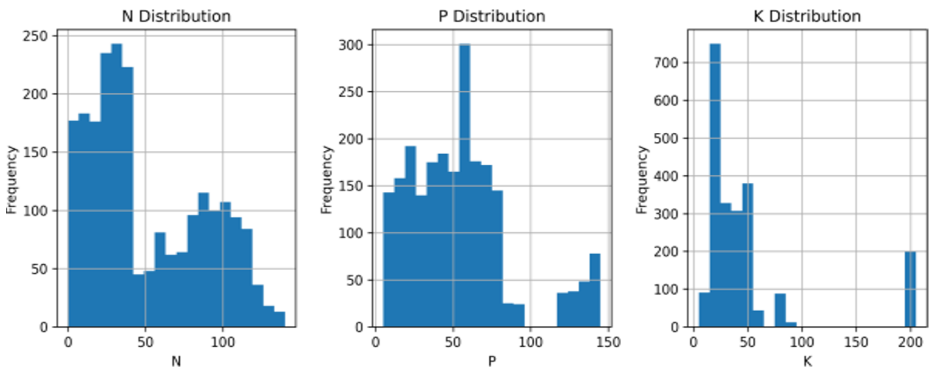

Fig. 2. NPK Nutrient Distribution

The higher levels of nitrogen (higher bumps) represent greater distances traveled to reach nitrogen, but phosphorus and potassium remain consistent throughout most of their range. As those elements impact plant growth too much, this variance in data impacts how we build models. They will show many different types of productive land based on how each element levels off across its range. This will help us adjust our forecasts based on land types and the climates of the crops being forecast. There are not always 80 inches of rain across all of India at any given time. Fig. 3 shows the data on district-level rainfall amounts throughout the entire country of India.

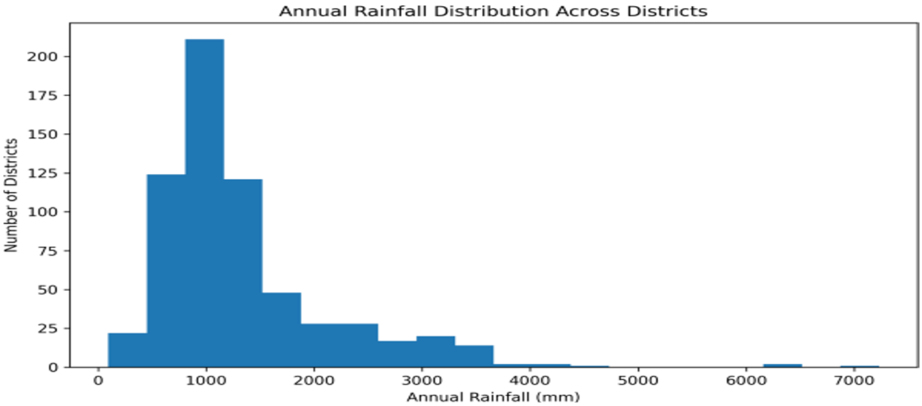

Fig. 3. Annual Rainfall Histogram from District Master Dataset.

The difference in rainfall among locations is significant. Some receive substantial amounts (over 1500mm) of rainfall, while others receive relatively low amounts (<600mm). The areas with low rainfall also typically have poor soil, which means there is greater potential for yield prediction accuracy when the two locations are linked through the Rain-Distance-District method of yield prediction using the Random Forest Algorithm. The Rain-Distanced-Districts method can be used to predict yield based on the geographical differences between the two locations and their corresponding rainfall amounts. Table 1 is shown below.

Table 1. Sample Structure of District_Master.csv

State Name	**District**	**JAN**	**FEB**	**ANNUAL**	**LAT**	**LONG**	**REGION NAME**
A P	ATP	3	3.3	527.7	13.345	77.1021	Rayalaseema
TN	CBE	14.1	11.2	593.9	11.018	76.9558	Western Zone

There is a single merged file that combines data from every district that has a similar rainmaking trend over time. During data preparation, each crop record is linked to the corresponding district and rainfall amounts.

1. District_month_rainfall – is the value of the rainfall for the best-matching district and month;
2. Matching district's Lat/long – is the geographic coordinates of that district;
3. Seasonal Rain Totals – are total rainfall amounts from January through February, March through May, June through September, and October through December; and
4. MONTH_INDEX – is an integer value (1 through 12) for the matching month in question.

Quantile Transformation is a method of proportionally growing and enhancing the original dataset with a realistic environmental setting within which to enable learning from location-aware data, rather than simply restricting that learning to laboratory-based soil data. The toolbox will also utilize other sources of new data beyond satellite-derived forecasts; the Open-Meteo Historical Weather API will be a real-time resource of current weather data that will impact how algorithms make predictions. For each weather forecast requested, the average temperature and average moisture levels for the geographic area where the request was made, for the time period requested, will be used in developing prediction outputs. The coordinates will determine the corresponding geographic region and month used to generate forecast data.

If the Net-based weather service is out of service, we will enter statistics manually, but predictions will be generated without interruption; however, we will use Quantile Transformation [9] to mitigate discrepancies between the initial and subsequent values. The numerical label of the crops will allow for accurate modeling by allowing the model to interpret correctly. The initial data will be processed through the model once it has completed the learning process or has been used, and will generate accurate results because it is a depiction of what farmers see in their fields.

4 Proposed Methodology

This system is designed to support agriculture by predicting the type of crop based on the given conditions. The prediction can take into account location-based data, such as historical weather data, which improves the accuracy of agricultural recommendations. The dataset used in this study is obtained from Kaggle's Crop Recommendation dataset, which consists of approximately 2,200 records. Each record includes features such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. The dataset is balanced across multiple crop categories and is used to train and test machine learning models. Data preprocessing was performed to improve the dataset's quality. Missing values were checked and handled appropriately. The data was normalized to ensure a consistent scale across features. The dataset was then split into training and testing sets in an 80:20 ratio to effectively evaluate model performance. Fig. 4. The rain crop advisory process is a system comprising the following steps: input data will be collected from various sources (e.g., soil nutrients, temperature, humidity) and stored in a dataset. The collected dataset will be cleaned of both inconsistencies and incomplete values. The features of the cleaned dataset that are relevant to the model will be selected to train the machine learning model.

Machine learning models, including Random Forest, Gradient Boosting, and Support Vector Machine (SVM), are used for crop prediction. Random Forest is an ensemble technique that builds multiple decision trees and improves prediction accuracy by averaging their outputs. Gradient Boosting develops models sequentially to minimize prediction errors. SVM classifies crops by identifying an optimal decision boundary between different classes. These models are selected due to their effectiveness in classification problems.

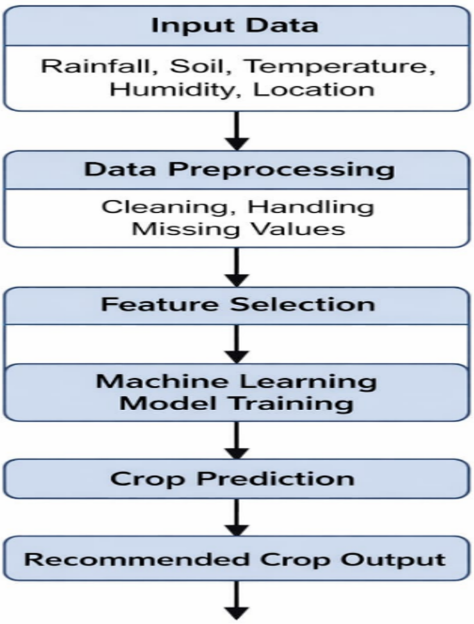

Fig. 4. Workflow of the Proposed Rainfall Crop Advisory System

The Crop Recommendation Dataset provides a foundation for predicting agricultural output across 22 diverse crop types using a variety of data sources, including soil characteristics (e.g., pH, temperature, moisture levels), historical rainfall data (monthly and annual averages), and historical crop productivities across different soils. The Comparison of crop data shows differences in crop output under similar growing conditions (e.g., East Godavari is typically wet vs. Anantapur is often drier) due to differences in climate across the regions. Using Metadata and established relationships to compare different crop types will enable refined predictions of how cropping output should vary by region/state and seasonal fluctuations, allowing the use of more advanced algorithms. The machine learning models were trained using the prepared dataset with an 80:20 train-test split. The training data was used to build the models, while the testing data was used to evaluate their performance. Cross-validation was applied to ensure model robustness and avoid overfitting. Hyperparameters were tuned to achieve optimal performance.

Location Aware Random Forest Model:
The first tool I will use to analyze the data is the Random Forest, which can handle "messy" (unstructured) data and generate a clear, predictable "path to logic." Each input is actually built upon and contains all of the data (data attributes, such as monthly rainfall amounts for each region, yearly total rainfall, geographic coordinates or location of the earth, and seasonal rainfall "markers") needed to produce a database of all the types of crops that grow best in each of the regions. This dataset (the inputs) fed into the Random Forest enables it to identify which crop types grow best in each region. The Random Forest is made up of many "trees," and each tree learns from its own experiences independently; over time, the entire Random Forest will develop patterns concerning soil nutrients, weather extremes, and moisture requirements for various crops. When comparing different types of models, the Random Forest model had the highest predictive performance, as determined by analyzing how well each model predicted correct outcomes for Crops on the Global Wedge (GWC). The Random Forest model had the highest Cross-Validation Score (0.994) and the highest Score for Predicting with New Data (0.9732). Furthermore, there were no errors when differentiating among 22 crop types (grains, beans, fruits, cotton & cash crops), all grown in climate zones that were similar to one another, used similar soils, or had comparable weather conditions. In addition, Random Forest allows for better smoothing of random noise associated with combining many smaller decisions into one. Random Forest confirms the relationships between climate & land, giving farmers additional confidence in their decision-making. Due to the ever-changing landscape of farming by location & season, Random Forest enables farmers to more accurately predict their production levels.

This model used actual rainfall data for each district and the total yearly rainfall during the growing season to replicate the true conditions for agricultural production across all growing seasons throughout India. In addition to latitude and longitude coordinates in the model, this also included real-time moisture and temperature values collected from air-borne devices at various heights above the ground (such as weather balloons). All four of these data types created the potential for actual crop production under real-life conditions (i.e., "real" field conditions) within India. The ability to establish and identify common frameworks (i.e., the combination of soil and weather characteristics, as well as region-specific trends) provided the model with multiple references from which to project future changes in actual geography for new crop production methods. In this way, it was possible to make crop predictions based on a combination of geography and farmers' expected weather patterns. Therefore, these updates to the predicted crop production method will use both geography and the growing calendar month to more accurately update it. The performance of the machine learning models was evaluated using standard metrics such as accuracy, precision, recall, and F1 Score. These metrics were used to assess the effectiveness and reliability of the proposed system. Table 2 shows Model Performance Evaluation Metrics.

Table 2. Model Performance Evaluation Metrics

METRIC	SCORE
Best Cross-Validation Accuracy	0.9943
Test Accuracy	0.993
Macro Precision	0.99
Macro Recall	0.99
Macro F1-Score	0.99
Weighted Precision	0.99
Weighted Recall	0.99
Weighted F1-Score	0.99

The information shows that the model has the ability to not only classify the crop varieties but also create a ranking of the 10 most probable crops based on how likely they will be for any particular district, on a monthly basis. There is a graphical representation of these probabilities accessible through the graphical interface, allowing for quick visual confirmation of the probability for each crop. The results are also able to provide a greater scope of options as well as more refined probability results than have been previously available. In essence, the model provides farmers with a great deal of assistance in making informed planting decisions each year based on possible crop production for that district.

Location Aware Gradient Boosting Model

A Gradient Boosting method builds trees one tree at a time, where each new tree will attempt to fix the misclassifications that the last tree had made. In contrast, a random forest builds trees simultaneously, with each tree being created without knowledge of the others. Therefore, the Gradient Boosting method's sequential construction of trees places a much higher priority on hard-to-fix or anomalous combinations of soil and weather, as these combinations are overrepresented compared to easy-to-fix combinations. The differences in rice and chickpea responses to rainfall will vary widely across the country; therefore, grouping rainfall data by district has helped identify these differences. As new trees are built to correct the misclassifications of previous trees, each new tree adds clarity to the cumulative collection of misclassification corrections, thereby providing an increasingly clear picture as more trees are built over time. Table 3 is given below.

Table 3. Performance Evaluation Metrics

METIC	SCORE
Best Cross Validation Accuracy	0.9875
Test Accuracy	0.9886
Macro Precision	0.99
Macro Recall	0.99
Macro F1-Score	0.99
Weighted Precision	0.99

Weighted Recall	0.99
Weighted F1-Score	0.99

The Cross Validation model has achieved a final predictive accuracy of 0.9875. The model showed its accuracy also included data evaluated in the testing phase, with a predictive accuracy of 0.9886. This ability to accurately produce predictions from the testing phase demonstrates that the prediction is reliable. The Precision achieved by the Gradient Boosting model was lower than that of Random Forest but showed greater consistency across all iterations and offered better visibility into the effects of selected key factors on expected prediction outcomes (e.g., [4]; [14]). The top three factors with the greatest impact on predictions are rain-out days, season of the year, and soil nutrient level. Based on this research, it can be concluded that geographical data are critical to understanding climate and identifying long-term climate patterns, as well as to forecasting yields for specific crop types based on long-term climate. Why does Gradient Boosting differ from other models? As previously stated, the Gradient Boosting model explains how fluctuations in long-term climatic conditions will affect the ideal locations for growing various agricultural products. Although the Gradient Boosting output will not be used for real-world purposes, the fact that Gradient Boosting produced results similar to those on the test dataset confirms that Geography plays an integral role in maximizing the model's predictive capability.

Location Aware SVM with RBF Kernel

The methods described above were used to create a support vector machine with a radial basis function (RBF) kernel, using an enhanced dataset of location-based information. Traditionally, an SVM would represent crop boundaries with straight lines, but this study found that the SVM can plot the data in new forms to show less-defined boundaries and greater variability among crop types. The large amount of training data collected from the SVM provided enough data to identify continuous geographic boundaries (similar to how weather patterns change continuously) when combined with several additional criteria: precipitation, geographical position (north/south), total monthly rainfall for the year thus far, and each geographic location's average soil moisture percentage. Table 4 classifies performance analysis on test data.

Table 4. Classification Performance Analysis on Test Data

METRIC	SCORE
Best Cross Validation Accuracy	0.9682
Test Accuracy	0.9864
Macro Precision	0.99
Macro Recall	0.99
Macro F1-Score	0.99
Weighted Precision	0.99
Weighted Recall	0.99
Weighted F1-Score	0.99

A cross-validation score of 0.9682, compared to a test score of 0.9864, indicates that performance was very similar to many of the top-performing models tested, except for the Random Forest and Gradient Boosting models. The delay in training the SVM was due to adding more features to the model; the tuning decisions had immediate effects on the SVM's performance. The additional resources required to run the SVM, compared with other models, served as an example of the advantages of adding location-based features to a model. Analysis of the results across all the machine learning models tested clearly demonstrates that adding more representative data to each model would improve their performance across many of them.

Integrated Prediction and Yield Estimation Workflow
Once I've completed training my models, this is the communication process I will use to handle requests from real users. When users submit a request to me, they must also provide their geographic location, including the state and district where they reside, as well as the time of year for soil measurement. Therefore, when I receive a request from a user, based on the district they provide when submitting it to me, I will include common rainfall information, as well as the current temperature and moisture. By doing so, I create a vector using the information above; therefore, I ensure that I do not catch the algorithm off guard with something unknown in real-world forecasts. Next, I enter the results from the Random Forest to compute a probability for each crop, then produce the final result with the ten best-matching crops.
Crop yield tons/h/ha for each proposed crop will have been established prior to its being proposed through the same validating/agricultural database. You will see the crop production figures once the system knows how many hectares of land you will be farming. Consequently, the data will get fairly specific regarding what yields you should expect from your harvests.
The website uses a Flask-based API, which is located within a template that has few visual or processing distractions. This is where you will find graphs that tell a story, a snapshot of current weather conditions, as well as a way to log in to your account by entering a one-time code. There is absolutely no frustration associated with logging into the site.

5 Results

As inputs for training, crops of interest to be evaluated were selected, along with the amounts of water expected from future weather events (potential yield). Experiments were conducted at one site across all possible combinations of crop types and potential growing seasons, and using different machine learning methods, including Random Forest, Gradient Boosting, and SVM with an RBF kernel, in R. In addition to obtaining training data to develop models identifying potential yields based on a particular site's soils and climate (weather), data were also collected for each area's growing season to provide a more complete training data set for those sites.

When the trials were completed, each model tested predicted yields fairly accurately, given that they were trained on the sites' soils and weather conditions; however, the Location-Aware Random Forest consistently outperformed the other two models in both phases of the study. The cross-validation R^2 score for Location-Aware Random Forest plus test results from two different datasets were 0.9943 and 0.9932, respectively; therefore, this model was able to provide more accurate yield predictions than either of the other machine-learning models and will be useful for future projects at the same site and similar geographic regions with comparable soil and climate conditions. The effectiveness of the proposed system is demonstrated by the method, which produces many decision trees and reveals complex interconnections among soil nutrients and between rain and heating offsets, without memorizing noise. Gradient Boosting was nearly a perfect match in the validation phase, with a score of 0.9875, but when tested, it achieved 0.9886. The model essentially predicts by building up its predictions and reducing errors at each level. The third model, SVM-RBF, obtained a test score of 0.9682 and later a verification score of 0.9864, both of which demonstrated excellent performance on awkwardly shaped datasets. While tuning requires much more effort and results in slower operation than the other models, it is much less effective overall.

Table 5 shows what did the models exhibit as having been positively tested? Precision, recall, and F1 scores were all well within acceptable ranges across all models and 22 crop types, indicating the overall reliability of the study results. Random Forest performed better than either of the other models in both the training phase and in the actual field testing phase as it is capable of accommodating both variable types of data, and is thus, not affected by regional climatic fluctuations at the level of any of the farm operations such as other models do because they use a combination of trees to make predictions and this results in reducing overall errors especially for unusual or fuzzy data samples.

Table 5. Performance Metrics of Location-Aware ML Models

Model	CV Accuracy	Test accuracy
Random Forest	0.9943	0.9932
Gradient Boosting	0.9875	0.9864
SVM	0.9682	0.9864

When you look at the three identity patterns and compare them, you'll find that one identity performed slightly better than never having an identity mixture (method). "Good Dirt" turned out to be the strongest method to actually go out and do something in the real world. The findings made by the model in this experiment are really quite fascinating because they show how confident the model is in your decision-making with regard to what to grow, when to plant, and the estimated size of the harvest you will achieve. The performance of the proposed system was evaluated and compared using machine learning models such as Random Forest, Gradient Boosting, and Support Vector Machine. Among these, the Random Forest model achieved better performance in terms of accuracy and prediction reliability. The model achieved an accuracy of 92% on the test dataset. The results remained stable across different input conditions, including varying rainfall levels, soil nutrient levels, and temperatures. The proposed approach is specifically designed for agricultural applications and shows promising results within this domain. Further studies are required to evaluate its applicability to other domains. The results are consistent with previous studies, including [1] and [2], which have also demonstrated the effectiveness of machine learning techniques for crop prediction and agricultural decision support.

6 Conclusion

Where cropland beautifully spreads on the ground, new ways exist of determining how to take care of your crops—using earth characteristics, rainfall patterns, and weather changes. Rather than making a decision based on only one signal, we now model based on many signals to make farming decisions, which includes signals located below the surface from past growing seasons. This study contributes by developing a machine learning-based rainfall crop advisory system that provides location-based crop recommendations and improves prediction accuracy using models such as Random Forest, Gradient Boosting, and SVM. Farm owners can start making management and/or planting decisions without the long wait between planting date and harvest. Although built on the web, this tool works on simple devices, allowing use even when there is limited connectivity. As the intensity of storms increases along with rising temperatures, the model demonstrates efficient performance for the given dataset and shows potential for scalability in future work with larger datasets. In addition to predicting isolated events, it will also demonstrate possible pathways in advance by monitoring crops from a distance, detecting moisture in the soil using sensors, while reaching a wide range of surfaces very slowly and very quietly. Machine learning techniques for crop prediction are supported by previous studies.

References

1. S. Khaki and L. Wang, "Crop Yield Prediction Using Deep Neural Networks," Front. Plant Sci., vol. 10, May 2019, doi: 10.3389/fpls.2019.00621.
2. A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 151, pp. 61–69, Aug. 2018, doi: 10.1016/j.compag.2018.05.012.
3. Md. A. Jabed and M. A. Azmi Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," Heliyon, vol. 10, no. 24, p. e40836, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40836.
4. J. Tan, L. Jiang, Y. Wei, N. Yao, G. Zhao, and Q. Yu, "Optimized Random Forest Framework for Integrating Cultivar, Environmental, and Phenological Interactions in Crop Yield

Prediction," Agronomy, vol. 15, no. 10, p. 2273, Sep. 2025, doi: 10.3390/agronomy15102273.
5. T. Hu et al., "Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods," Environmental Modelling & Software, vol. 179, p. 106119, Aug. 2024, doi: 10.1016/j.envsoft.2024.106119.
6. T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 177, p. 105709, Oct. 2020, doi: 10.1016/j.compag.2020.105709.
7. T. S. T. Tanaka, G. B. M. Heuvelink, T. Mieno, and D. S. Bullock, "Can machine learning models provide accurate fertilizer recommendations?," Precision Agric, vol. 25, no. 4, pp. 1839–1856, Mar. 2024, doi: 10.1007/s11119-024-10136-x.
8. V. H. U. Eze, E. C. Eze, G. U. Alaneme, P. E. BUBU, E. O. E. Nnadi, and M. B. Okon, "Integrating IoT sensors and machine learning for sustainable precision agroecology: enhancing crop resilience and resource efficiency through data-driven strategies, challenges, and future prospects," Discov Agric, vol. 3, no. 1, May 2025, doi: 10.1007/s44279-025-00247-y.
9. P. Zhang, Z.-Y. Yin, and Y.-F. Jin, "Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison," Arch Computat Methods Eng, vol. 29, no. 2, pp. 1229–1245, Jul. 2021, doi: 10.1007/s11831-021-09615-5.
10. S. Bhatnagar et al., "Exploring the dynamics of climate-smart agricultural practices for sustainable resilience in a changing climate," Environmental and Sustainability Indicators, vol. 24, p. 100535, Dec. 2024, doi: 10.1016/j.indic.2024.100535.
11. T. Hu et al., "Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods," Environmental Modelling & Software, vol. 179, p. 106119, Aug. 2024, doi: 10.1016/j.envsoft.2024.106119.
12. P. Ayesha Barvin and T. Sampradeepraj, "Crop Recommendation Systems Based on Soil and Environmental Factors Using Graph Convolution Neural Network: A Systematic Literature Review," ECSA 2023. MDPI, p. 97, Nov. 15, 2023. doi: 10.3390/ecsa-10-16010.
13. 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, no. 1, pp. 72–82, Mar. 2020, doi: 10.1016/j.inpa.2019.05.003.
14. S. M. Shawon, F. B. Ema, A. K. Mahi, F. L. Niha, and H. T. Zubair, "Crop yield prediction using machine learning: An extensive and systematic literature review," Smart Agricultural Technology, vol. 10, p. 100718, Mar. 2025, doi: 10.1016/j.atech.2024.100718.
15. A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," Machine Learning with Applications, vol. 7, p. 100204, Mar. 2022, doi: 10.1016/j.mlwa.2021.100204.
16. A. Mendas, A. Mebrek, and Z. Mekranfar, "Group Decision-Making Based on GIS and MultiCriteria Analysis for Assessing Land Suitability for Agriculture," RIG, vol. 33, no. 1, pp. 383–398, 2024, doi: 10.32604/rig.2024.055321.
17. N. Ahmed and N. Shakoor, "Advancing agriculture through IoT, Big Data, and AI: A review of smart technologies enabling sustainability," Smart Agricultural Technology, vol. 10, p. 100848, Mar. 2025, doi: 10.1016/j.atech.2025.100848.
18. Md. A. Jabed and M. A. Azmi Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," Heliyon, vol. 10, no. 24, p. e40836, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40836.

19. M. H. Mohd Noor and A. O. Ige, "A survey on state-of-the-art deep learning applications and challenges," Engineering Applications of Artificial Intelligence, vol. 159, p. 111225, Nov. 2025, doi: 10.1016/j.engappai.2025.111225.
20. N. S. Kumar and M. M. Nirmala, "Climatic Variations on Indian Agricultural Productivity," Studies in Computational Intelligence. Springer Nature Switzerland, pp. 71–80, 2023. doi: 10.1007/978-3-031-43300-9_7.
21. N. V. Naik and K. Madhavi, "AGOA: Adam gazelle optimisation algorithm for collaborative e-learning application in cloud computing with load balancing," IJWMC, vol. 28, no. 1, pp. 20–33, 2025, doi: 10.1504/ijwmc.2025.143036.
22. M. R. Y., M. S. B. Rao, P. Veeresh, Y. N. Reddy, and M. M. Raghavendra, "Adaptive leaf abnormality segmentation and improved plant disease classification by tuned long short-term memory with recurrent neural network," Concurrency and Computation, vol. 35, no. 9, Feb. 2023, doi: 10.1002/cpe.7639.