Crop Recommendation Using Soil Characteristics And Historical Rainfall Data

Chaitanya Ashish Khot
MSc in Computing (Data Analysis)
Dublin City University
Dublin, Ireland
chaitanya.khot2@mail.dcu.ie

Yash Manoj Modi
MSc in Computing (Data Analysis)
Dublin City University
Dublin, Ireland
yash.modi2@mail.dcu.ie

***Abstract*—To maximize agricultural productivity and ensure food security, crop selection is crucial. In order to help farmers make decisions for sustainable agricultural practices, this study investigates the use of historical rainfall data and soil properties for crop recommendation. The research primarily focuses on the rainfall factor as a key determinant in crop suitability and takes into account ideal soil conditions for different crops. We utilize sophisticated data analysis methods and machine learning algorithms, including ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory), for rainfall forecasting. By leveraging vast datasets on soil qualities and historical rainfall made accessible by meteorological departments, we identify the most suited crops for particular locations, taking into consideration factors such as soil pH, moisture content, and past precipitation patterns.**

The suggested strategy provides farmers with a useful tool to optimize their farming practices and increase productivity while maintaining economic stability and protecting the environment in the agricultural industry. By combining the insights from historical rainfall data with soil properties, this research aims to provide accurate and informed crop recommendations, contributing to sustainable and efficient agricultural practices.

***Index Terms*—Crop recommendation, Soil characteristics, Historical rainfall data, Agriculture, Data analysis, Machine learning, Sustainability, Agricultural productivity.**

I. Introduction

Agriculture has been the backbone of India's economy, supporting a significant portion of the population. Ensuring sufficient crop yields is crucial for farmer welfare and overall economic health. Accurate crop forecasting can alleviate the challenges faced by farmers and promote sustainable agricultural practices. This literature review aims to explore the factors influencing crop recommendations, with a focus on incorporating soil characteristics and historical rainfall data. Parameters like soil composition, moisture content, and temperature significantly impact crop performance. Access to reliable data and rigorous analysis techniques are essential for making informed decisions. In the extensive data preprocessing stage of the study process, we displayed the data, found trends, and looked at moving averages to comprehend

weather patterns. To help with exact crop selection, we developed the seasonal features kharif, rabi, and zaid in the feature engineering part. Moving on to the modeling stage, we utilized earlier research and continued to employ ARIMA and LSTM models while fine-tuning them for the best rainfall-predicting outcomes. The user interface section demonstrates how predicted precipitation is used to suggest appropriate crops to farmers based on their unique subdivisions. The interface shows typical kharif-season crops, including details on the qualities of the soil and the necessary amount of rainfall. Additionally, it suggests the top three crops based on crop preferences and projected rainfall for the kharif season the top three closest matches.

The performance of the ARIMA and LSTM models for rainfall forecasting is discussed in the results section. We analyze their usefulness in supporting trustworthy crop recommendations by measuring their accuracy using measures such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). This study aims to give a complete and data-driven approach to crop recommendation, enabling farmers to make informed decisions for sustainable agriculture and higher productivity.

II. Literature Review

The development of accurate crop recommendation systems based on soil parameters and historical rainfall data has been the focus of extensive research, with researchers employing diverse methods, including machine learning and deep learning approaches. In a recent study [1], machine learning and deep learning techniques such as ARIMA, Artificial Neural Networks (ANN), and Support Vector Machines (SVM) were utilized for precise rainfall forecasting. Another study [2] concentrated on predicting rainfall using Generative Adversarial Networks (GAN) with Long Short-Term Memory (LSTM) as a generator and Convolutional Neural Network (CNN) as a discriminator, achieving over 90% accuracy in rainfall prediction.

Preprocessing methods have played a vital role in ensuring data accuracy and model reliability. For instance, [3] employed decision trees and k Nearest Neighbor (kNN) in machine learning to predict rainfall, emphasizing their advantages over traditional techniques. Similarly, [4] conducted time series analysis using Seasonal ARIMA and SARIMA to identify yearly and seasonal trends and patterns in rainfall data.

Assessing the effectiveness of crop recommendation models is essential, and various evaluation metrics have been utilized. [6] employed feature selection techniques like the Boruta algorithm, Sequential Forward Feature Selection (SFFS), and Recursive Feature Elimination (RFE), along with multiple classifier methods, to assess model performance using criteria like F1 score, error rate, kappa, and mean absolute error (MAE). Likewise, [7] used Logistic Regression, Support Vector Machine, Random Forest Classifier, and K Nearest Neighbor to identify influential factors in crop forecast and achieved around 90% accuracy when measured by R-Squared and Root Mean Squared Error (RMSE).

Additionally, the aim of the study in [18] was to determine the best-fit probability distribution for monthly rainfall in Navsari and develop an appropriate ARIMA model for rainfall forecasting. The study utilized various continuous probability distributions and trend analysis using the Mann Kendall test and goodness of fit tests. The results indicated specific distributions for different months and highlighted an appropriate ARIMA model for monthly rainfall prediction.

Furthermore, in [19], the aim was to improve rainfall data forecasting accuracy in Makassar, Indonesia, by comparing two forecasting methods: Autoregressive Integrated Moving Average (ARIMA) and Kalman Filter. The study evaluated the models based on the Mean Absolute Percentage Error (MAPE) value, with the Kalman Filter algorithm outperforming the ARIMA model, providing more precise rainfall forecasts with a lower MAPE value of 47.00 for the given dataset.

Adding to the literature, in [20], a rainfall forecasting model called RSDF-AM-LSTM was developed, focusing on regional scale division forecasting using attention mechanisms and Long Short-Term Memory (LSTM) networks. The proposed model aimed to improve the accuracy and efficiency of rainfall predictions. Attention mechanisms were employed to capture important spatial features, while LSTM networks handled temporal modeling. The study compared the RSDF-AM-LSTM model's performance against existing forecasting methods and found that it outperformed them, providing more accurate rainfall forecasts for regional-scale divisions. This novel approach to rainfall forecasting, leveraging attention and LSTM mechanisms, has the potential to benefit various applications in water resource management, agriculture, and disaster preparedness.

III. Data

A. Data Description

1) Rainfall Data: The spatial granularity of this dataset [16] is one of its main advantages. It allows for a thorough investigation of spatial variability in precipitation patterns and has 4188 rows and 19 columns. It provides rainfall data for several meteorological subdivisions across India. Such research can be used to pinpoint areas that are particularly susceptible to severe weather events like droughts, high rain, or unpredictable climatic behavior. Therefore, depending on historical rainfall patterns in their individual regions, farmers can optimize their crop choices and irrigation practices, which is highly beneficial for developing targeted strategies for agricultural planning.

The dataset's temporal and spatial coverage also makes it an effective tool for determining how climate change will affect rainfall patterns. The historical data can be used as a starting point for researchers to examine whether there have been any notable changes or variances in precipitation throughout time. This can help in determining places that may be more susceptible to extreme weather occurrences as a result of shifting climatic conditions and in understanding how climate change impacts various parts of India.

The literature review identifies the unmet research needs in the realm of deep learning and machine learning methods for predicting rainfall. The requirement for ensemble approaches, resolving data imbalance and seasonal trends, spatial and temporal integration, examining model transfer-ability, taking climate change considerations into account, and creating decision support systems are a few of these gaps. Exploring these topics can help rainfall prediction models become more accurate and useful, which will be good for water resource management, agriculture, and disaster preparedness.

TABLE I
RAINFALL DATASET DATATYPES

Column	Datatype
SUBDIVISION	object
YEAR	int64
JAN	float64
FEB	float64
MAR	float64
APR	float64
MAY	float64
JUN	float64
JUL	float64
AUG	float64
SEP	float64
OCT	float64
NOV	float64
DEC	float64
ANNUAL	float64
JF	float64
MAM	float64
JJAS	float64
OND	float64

This information is significant for many different fields other than those studying agriculture and climate change. The thorough knowledge of rainfall patterns can be used to better manage water resources, assisting authorities in properly allocating water resources, especially in water-scarce areas. By researching heavy rainfall events and locating flood-prone locations, disaster preparedness and response planning can be improved, enabling authorities to create focused mitigation strategies.

Additionally, by using this dataset, urban planners and engineers can create more resilient infrastructure that accounts for past rainfall patterns and severe weather events. This preventative approach can reduce infrastructure damage and guarantee local communities' safety and well-being during inclement weather.

2) Crop Data: For the research project "Crop Recommendation Using Historical Rainfall Data," the dataset "What Crop to Grow" [17] on Kaggle offers crucial data. This dataset provides insightful information on numerous crops, including their unique characteristics and needs, which may be combined with historical rainfall data to produce well-informed crop recommendations.

The information provides specifics on many crops, such as their nutrient needs (N, P, K), optimal climatic conditions (temperature, humidity, pH), and historical average yields. Understanding the ideal environmental circumstances and nutrient requirements for each crop depends on this information.

The research effort can create a comprehensive model that suggests acceptable crops based on the current climatic circumstances by fusing the historical rainfall data from the prior dataset with this crop-specific data.

TABLE II
CROP DATASET DATATYPES

Column	Datatype
N (Nitrogen)	int64
P (Phosphorus)	int64
K (Potassium)	int64
temperature	float64
humidity	float64
pH	float64
rainfall	float64
label (crop names)	object

This information makes it possible to identify crops that prosper under certain environmental circumstances, guaranteeing that farmers receive precise and unique crop suggestions specific to their areas. Additionally, by evaluating each crop's prospective productivity with the help of average yield statistics, farmers may make more informed crop selection selections.

A comprehensive approach to crop recommendation is made possible by the integration of these two datasets, which considers both historical rainfall data and crop-specific features. By utilizing this enormous dataset, the research initiative can provide farmers with insightful analysis and helpful suggestions, promoting improved crop cultivation and sustainable agricultural practices.

3) Subdivision and Crop data (Custom dataset): The two existing datasets "What Crop to Grow" and "Rainfall Data from 1901 to 2017 for India" were combined to produce a bespoke dataset for this research project. The unique dataset acts as a link between historical rainfall data and crop-specific features, allowing for the accurate creation of crop recommendations for various Indian subdivisions.

The "Rainfall Data from 1901 to 2017 for India" dataset's subdivisions are all included in the custom dataset's "Subdivision" column. Each entry in this column refers to a distinct Indian subdivision.

A 'Crop' column in the custom dataset also includes a list of every crop from the "What Crop to Grow" dataset. This column provides useful details on the kinds of crops that can be grown in particular regions by matching each state (or subdivision) with the crops that are grown there.

TABLE III
CUSTOM DATASET DATATYPES

Column	Datatype
SUBDIVISION	object
crops	object

Utilizing the open-source data on Kaggle for all data integration and mapping, careful study was done to guarantee the veracity and accuracy of the data in the bespoke dataset. To prevent any discrepancies and guarantee that the recommendations presented are based on reliable and trustworthy data sources, the procedure involves meticulous validation and cross-referencing of the data.

The study effort guarantees that the crop recommendation system is founded on a firm foundation of precise rainfall data, crop-specific features, and well-researched information by developing this bespoke dataset. By combining several datasets, the research is better equipped to give farmers in India insightful explanations and practical advice that will enable them to choose crops and design their farms in an informed manner.

IV. METHODOLOGY

A. Data Preprocessing

The dataset's null values and the absence of any negative values were addressed in the first stage. A pattern developed as a result of this approach, which involved identifying the rows and columns that had missing values. It was noted that a sizable percentage of null values were present in data entries made before the year 1970, most likely as a result of possible shortcomings in data retention during that time.

Different imputation strategies, such as filling with 0, mean, forward fill, and interpolation methods, were taken into consideration to manage the missing values. Forward fill (ffill) was chosen as the best suitable method after these strategies were compared. This option was chosen because it can maintain temporal order and reduce distortions in the original dataset.

The missing values in the seasonal and yearly columns were supplied using the entire monthly rainfall data that was received during the imputation method. This action improved the dataset's comprehensiveness and readiness for further analysis.

EDA, or exploratory data analysis, was done to glean important insights from the dataset. Analysis of the rainfall's monthly distribution revealed a significant concentration (about 87

Visualizing the annual rainfall average by subdivision for all years was a further step in the investigation. The results showed that West Rajasthan had the lowest rainfall levels, while Arunachal Pradesh had the greatest average rainfall. Additionally, individual extreme occurrences were found to have a substantial impact on India's overall rainfall patterns, including the 1965–66 drought and the 1961–1962 floods in Pune.

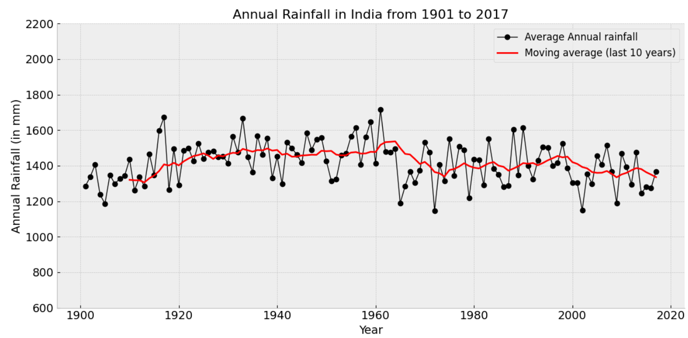

Fig. 1. Moving average of the Annual rainfall in India (1901-2017)

The seasonality of the rainfall data was also analyzed using the dataset. The majority of the rainfall occurred during the monsoon season (Jun-Sep), which was confirmed by the visualization of subdivision-by-subdivision seasonal rainfall data. Pre- and post-monsoon plots for all years were compared to get additional understanding of the seasonal trends.

Moving average plots were created during the peak rainfall months of July and August in order to look for any potential long-term trends or variations in rainfall. The findings showed essentially constant rainfall patterns over time, with a very modest downward trend in recent decades. Additionally, the analysis suggested that there may have been a decrease in rainfall since the 1960s, which was also a time of global warming. As a result of climate change, signs of a shorter monsoon season have also been detected.

In order to get new insights into the historical rainfall data, this research project's data pretreatment step involved managing missing values, using the necessary imputation techniques, and performing exploratory data analysis. The obtained information was thought to be appropriate for further research targeted at creating crop recommendation algorithms based on rainfall patterns.

B. Feature Engineering

A crucial phase in the data analysis process is feature engineering, which aims to turn unstructured data into useful characteristics that might improve the performance of predictive models. In order to extract critical seasonal and monthly features from the original dataset in this study on "Crop Recommendation Using Historical Rainfall Data," feature engineering was used.

The complete dataset was split into three distinct dataframes: monthly, seasonal, and yearly rainfall dataframes in order to more clearly distinguish the columns and make additional analysis easier. This division made it possible to more clearly distinguish between the various kinds of rainfall data and allowed for more precise feature extraction.

There are three major agricultural seasons in India: Kharif (June–Sept), Rabi (October–March), and Zaid (Mar–June) [15]. These different seasons' average and cumulative rainfall were estimated and included to the dataset as additional attributes. The ability to take advantage of the various climatic conditions in different months to cultivate a variety of crops all year long is a key benefit of this information for Indian farmers.

The specific timing of these agricultural seasons is essential for crop planning and selection as it enables farmers to modify their farming techniques in accordance with the seasonal weather patterns.

As a test dataset for assessing the effectiveness of rainfall forecasts, a second dataframe for the year 2017 was developed in addition to the seasonal features. When forecasting rainfall patterns for subsequent years, this test dataset is crucial for evaluating the precision and dependability of crop recommendation algorithms.

All derived dataframes, including the monthly, seasonal, yearly, and 2017 rainfall dataframes, were exported and saved as copies in order to preserve the preprocessed data for subsequent research and replication..

Using a Principal Component Analysis (PCA) test, the significance of the dataset's features was assessed. The PCA analysis showed that as compared to when they were compiled seasonally or annually, the monthly rainfall data had a more substantial impact. This conclusion emphasizes how important monthly rainfall patterns are in shaping agricultural outcomes and how crucial it is to take them into account individually in crop recommendation models. In order to assist better analysis and feature extraction, the feature engineering method in this research study involves developing unique dataframes for monthly, seasonal, and annual rainfall data. In certain locations of India, agricultural planning could be more effectively planned thanks to the identification and inclusion of crucial seasonal elements. The test dataset for 2017 also offered a way to assess the precision of rainfall predictions. The PCA analysis highlighted the importance of monthly rainfall data as key drivers in crop recommendation models in its final point. The foundation is set by these characteristics for the later development and assessment of crop recommendation algorithms based on historical rainfall patterns in India.

V. Modeling

Various strategies were used to acquire insights into the underlying patterns in the rainfall data and prepare it for modeling during the pre-modeling phase.

To determine whether seasonality, trend, and residual components were present in the data, multiplicative and additive decomposition were used. A seasonal component of 0 in the additive decomposition signifies the lack of recurrent patterns in the time series.Similar to this, a residual component of 0 indicates that the trend and seasonal patterns can account for all changes in the data. A residual component of 1 indicates that all data variations can be explained by the trend and seasonal patterns alone, without the need for additional random fluctuations, while a seasonal component of 1 in the multiplicative decomposition implies that seasonal patterns scale proportionally with the trend.

The decomposition analysis's findings showed that there was little seasonal variation in the historical rainfall data throughout time. This discovery helped to clarify the dataset's temporal dynamics and provided direction for the ensuing modeling strategy.

Using the Augmented Dickey-Fuller (Adfuller) test, stationarity in the dataset was evaluated. The null hypothesis, which presupposed that the data weren't stationary, was rejected when it was discovered that the test's p-value was less than 0.05. This suggested that the data had a steady behavior, allowing time series forecasting algorithms to be used.

The main purpose of this research project's modeling phase was to create precise and efficient rainfall forecasting models, with the ultimate objective of improving crop recommendation systems. Initially, a number of well-known regression methods were built and assessed, including Multiple Linear Regression, Random Forest Regression, K Neighbors Regression, Support Vector Regression, and Decision Tree Regression. Multiple Linear Regression among these models performed the worst, while Support Vector Regression (SVM) displayed the most encouraging results. However, a significant flaw in these earlier models was found. They used all the available input features, but without taking into account future data points, they simply relied on existing features and previous data to forecast results for the test data. Due to this limitation, the models were unable to accurately estimate future years' rainfall, which limited their usefulness in predicting seasonal rainfall that could influence agricultural operations.

The emphasis switched to using time series forecasting models, which are especially made to handle temporal data and produce forecasts for future time points, in order to overcome this intrinsic restriction. The three time series forecasting models that were selected were LSTM (Long Short-Term Memory), ETS (Exponential Smoothing State Space Model), and ARIMA (AutoRegressive Integrated Moving Average).

As a traditional time series model, ARIMA's autoregressive (AR) and moving average (MA) components can capture autocorrelation and seasonality in the data. ETS, on the other hand, uses an exponential smoothing technique to detect seasonality and trends in the data. Finally, LSTM, a deep learning-based model, has demonstrated outstanding performance in a variety of time series forecasting tasks and is ideally suited to capture complicated temporal patterns in sequential data.

These time series models were initially used to forecast monthly data, but the results were found to be unreliable. The model's attempt to forecast values for all 12 months across different subdivisions presented difficulties in managing such complexity. To increase forecast accuracy, efforts were undertaken to fine-tune the models and investigate the application of AIC/BIC and regularization. These efforts, however, did not produce the anticipated improvements and, in some instances, even led to a drop in performance.

A method that used the forecast from prior months as input for forecasting the following month was also tested. This method, however, turned out to be less successful because it was unable to fully capture the data's underlying temporal dynamics. When the researchers recognized that concentrating on seasonal forecasts may simplify the models and produce more relevant results than projecting monthly values, they made a huge advancement. As a result, the focus of the research changed to predicting India's three main harvesting seasons: Kharif, Rabi, and Zaid. These seasons are important to Indian farmers because they provide certain times for growing crops and correspond to the various meteorological conditions throughout the year. So, for this seasonal forecast, ARIMA and LSTM, two time series forecasting models, were used. While LSTM, a deep learning-based model, has the potential to learn complex temporal patterns from the data, ARIMA, a traditional time series model, gives the advantage of interpretability and simplicity of implementation. When compared to ARIMA, baseline models using LSTM performed better, suggesting that LSTM has the capacity to capture the intricate correlations found in seasonal rainfall data. Cross-validation techniques were used to further improve the accuracy of the findings, and other combinations of the ARIMA parameters (p, d, and q) were investigated to optimize the models. These initiatives were made with the intention of creating precise and dependable models to predict seasonal rainfall patterns, which might greatly improve crop recommendation systems and well-informed agricultural planning in India.

This research project's modeling phase witnessed a significant change from time series forecasting techniques to conventional regression models. The transition from monthly forecasts to seasonal forecasts, together with the use of ARIMA and LSTM, contributes significantly to solving the research challenge. These initiatives attempted to create reliable forecasting models for the Kharif, Rabi, and Zaid seasons, offering insightful information for improving agricultural operations and promoting sustainable crop farming methods in India.

VI. User Interface

Using Python and the 'ipywidgets' package, the crop recommendation system's user interface was designed on Google Colab to be interactive and user-friendly. Farmers can choose their subdivision from a drop-down menu, which then displays the kharif season crops that are often cultivated there. Farmers are informed about the appropriate soil qualities and quantity of rainfall needed for a particular crop's successful production after making their choice.

The interface also offers information on recent seasonal precipitation for the past two years (the kharif, rabi, and zaid seasons), as well as a forecast for the future year. This makes it possible for farmers to better plan their agricultural efforts based on past trends and forecasts.

The interface compares predicted rainfall for the kharif season with crop choices in order to further improve suggestions. It recommends the top three crops whose average rainfall needs are most in line with the projections. With the use of this data-driven methodology, farmers are better equipped to choose crops, maximize yields, and advance sustainable farming techniques.

In conclusion, the user interface is a useful tool that provides tailored and practical crop suggestions based on local data and predicted meteorological circumstances. It seeks to raise agricultural output, promote economic stability, and encourage environmentally friendly farming methods.

VII. Results and Discussion

We give the evaluation findings for the two time series forecasting models, ARIMA and LSTM, in this section, both before and after adjusting the corresponding parameters. The Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) evaluation metrics are used to evaluate the performance of the models. It is preferred to use the Mean Absolute Percentage Error (MAPE) when forecasting rainfall since it provides a general indicator of the model's accuracy for continuous numerical predictions like rainfall amounts by calculating the percentage difference between projected and actual values.However, because rainfall forecasting works with continuous numerical variables, the F1 score may not be applicable because classification tasks are better suited to them. As a result, in this situation, MAPE is the better statistic to use to evaluate forecasting accuracy. The results are summarized in the table below:

TABLE IV
Comparison of Evaluation Metrics for ARIMA and LSTM Before and After Tuning

Model	MAPE (%)	MAE	RMSE
ARIMA (Before Tuning)	35.11	20.12	20.12
ARIMA (After Tuning)	26.08	53.33	53.33
LSTM (Before Tuning)	50.23	25.72	25.72
LSTM (After Tuning)	33.35	19.07	19.07

For the ARIMA model:

Before tuning, the MAPE was 35.11%, reflecting an average prediction error of 35.11%. The MAE and RMSE were 20.12, indicating the average absolute and squared errors, respectively. After tuning, the ARIMA model's performance significantly improved. The MAPE decreased to 26.08%, demonstrating a reduction in prediction errors. However, the MAE and RMSE increased to 53.33, suggesting that the model's predictions had larger absolute and squared errors after tuning. Regarding the LSTM model:

Before tuning, the LSTM model exhibited a higher MAPE of 50.23% compared to the ARIMA before tuning. The MAE and RMSE were 25.72, indicating the average absolute and squared errors, respectively. After tuning, the LSTM model's performance showed improvement, as indicated by the lower MAPE of 33.35%. The MAE and RMSE also decreased to 19.07, signifying a reduction in average absolute and squared errors after tuning.

Interpretation of the results: The evaluation results reveal that both the ARIMA and LSTM models performed better in terms of MAPE, which reflects superior forecast accuracy, after parameter adjustment. The untuned LSTM model originally beat the untuned ARIMA model, however after tuning, the ARIMA model's relative accuracy was greatly improved. However, the tuned LSTM model outperformed the tuned ARIMA model in terms of performance. In conclusion, the findings imply that, among the models examined, the tweaked LSTM model offers the best precise projections. Given its capability to capture intricate temporal trends in the historical rainfall data, this points to its potential for helping crop recommendation systems and agricultural planning. It is important to remember that rainfall serves as the key factor in this study's focus on the prediction of rainfall patterns. Future research should investigate the incorporation of additional pertinent parameters, such as temperature or location terrain, to boost the accuracy of the predictions, even if the current study offers insightful information about rainfall forecasting using specific techniques. Incorporating more variables might provide a more thorough understanding of the intricate relationships that affect rainfall, improving predicting accuracy and dependability. Future study has the potential to advance the science of rainfall forecasting and contribute to better water resource management, agricultural practices, and disaster preparedness by taking a wider range of elements into account.

VIII. Conclusion and Future Work

In this study, we used machine learning algorithms and data analytic tools to anticipate rainfall and suggest appropriate crops for sustainable agriculture practices. In order to help farmers make better decisions, historical rainfall data and soil attributes have been integrated. We identified crucial elements including soil pH, moisture content, and previous precipitation patterns that have a major impact on crop output by utilizing enormous datasets made available by meteorological departments.

Our analysis revealed that accurate crop forecasting is essential to address the challenges faced by farmers, especially in light of the urgent problems affecting agricultural productivity and farmer welfare. Poor yields have led to serious difficulties, necessitating precise crop selection to ensure food security and economic stability. By employing ARIMA and LSTM models for rainfall prediction, we were able to improve forecast accuracy and reduce relative prediction errors (MAPE) to 33.35%. The tuning process further enhanced the LSTM model's performance, resulting in reduced absolute and squared errors (MAE and RMSE) of 19.07.

The most typical harvesting season, the kharif season, is the main focus of this study's crops. For a wider application, future study can investigate other agricultural seasons like rabi or zaid. Researching how climate influences various cropping seasons will result in more reliable crop recommendation systems that will help with sustainable agriculture and water resource management..

We have given farmers a useful tool through our research to help them improve their agricultural methods and increase crop yields. Farmers may make wise choices for efficient crop production by taking important aspects like soil properties and historical rainfall data into account. In addition to increasing agricultural productivity, this will support the agricultural sector's long-term growth and sustainability.

REFERENCES

[1] C. Z. Basha, N. Bhavana, P. Bhavya, and Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 2020, pp. 92–97.

[2] R. Venkatesh, C. Balasubramanian, and M. Kaliappan, "Rainfall prediction using generative adversarial networks with convolution neural network," *Soft Comput.*, vol. 25, no. 6, pp. 4725–4738, 2021.

[3] S. Biruntha, B. S. Sowmiya, R. Subashri, and M. Vasanth, "Rainfall Prediction using kNN and Decision Tree," in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, 2022, pp. 1757–1763.

[4] U. Ashwini, K. Kalaivani, K. Ulagapriya, and A. Saritha, "Time Series Analysis based Tamilnadu Monsoon Rainfall Prediction using Seasonal ARIMA," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1293–1297.

[5] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop yield prediction using machine learning algorithms," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, 2019, pp. 125–130.

[6] A. Suruliandi, G. Mariammal, and S. P. Raja, "Crop prediction based on soil and environmental characteristics using feature selection techniques," *Math. Comput. Model. Dyn. Syst.*, vol. 27, no. 1, pp. 117–140, 2021.

[7] R. Kavitha, M. Kavitha, and R. Srinivasan, "Crop recommendation in precision agriculture using supervised learning algorithms," in *2022 3rd International Conference for Emerging Technology (INCET)*, 2022, pp. 1–4.

[8] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, 2017, pp. 32–36.

[9] Anjana, A. Kedlaya, A. Sana, B. A. Bhat, S. Kumar, and N. Bhat, "An efficient algorithm for predicting crop using historical data and pattern matching technique," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 294–298, 2021.

[10] A. Sharma, M. Bhargava, and A. V. Khanna, "AI-Farm: A crop recommendation system," in *2021 International Conference on Advances in Computing and Communications (ICACC)*, 2021, pp. 1–7.

[11] M. Viviliya, P. Vaidhehi, and R. Srinivasan, "The design of hybrid crop Recommendation System using machine learning algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 4305–4311, 2019.

[12] A. Mahadware, A. Saigiridhari, A. Mishra, A. Tupe and N. Marathe, "Rainfall Prediction using Different Machine Learning and Deep Learning Algorithms," *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, 2022, pp. 1-8, doi: 10.1109/ASIANCON55314.2022.9908857.

[13] Benny Antony, "Prediction of the production of crops with respect to rainfall," *Environmental Research*, Volume 202, 2021, 111624, ISSN 0013-9351, https://doi.org/10.1016/j.envres.2021.111624.

[14] M. S. Pathan, M. Jain, A. Nag, T. A. Skaif and S. Dev, "Analyzing Impact of Time on Early Detection of Rainfall Event," *2021 Photonics & Electromagnetics Research Symposium (PIERS)*, 2021, pp. 1985-1989, doi: 10.1109/PIERS53385.2021.9695104.

[15] Wikimedia Commons. (n.d.). Different cropping seasons in India [Photograph]. Retrieved from https://commons.wikimedia.org/wiki/File:Different_cropping_seasons_in_India.jpg

[16] Government of India, Ministry of Electronics and Information Technology. (n.d.). Sub-Divisional Monthly Rainfall (1901-2017) [Data file]. Retrieved from https://data.gov.in/resource/sub-divisional-monthly-rainfall-1901-2017

[17] Atharva Ingle. (n.d.). What Crop to Grow [Data file]. Retrieved from https://www.kaggle.com/code/atharvaingle/what-crop-to-grow/input

[18] D. K. Dwivedi and P. K. Shrivastava, "Rainfall Probability Distribution and Forecasting Monthly Rainfall of Navsari using ARIMA Model," Indian Journal of Agricultural Research, 2022, doi: 10.18805/IJARe.A-5793.

[19] Sukarna, E. Y. P. Ananda, and M. S. Wahyuni, "Rainfall Forecasting Model Using ARIMA and Kalman Filter in Makassar, Indonesia," Journal of Physics: Conference Series, vol. 2123, no. 1, p. 12044, 2021, doi: 10.1088/1742-6596/2123/1/012044.

[20] W. Li, P. Zhang, H. Dong, Y. Jia, and W. Cao, "RSDF-AM-LSTM: Regional Scale Division Rainfall Forecasting Using Attention and LSTM," ACM/IMS Transactions on Data Science, vol. 2, no. 4, pp. 1–27, 2021, doi: 10.1145/3498333.