Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Development of a Machine Learning-Based Prediction Model for Sugarcane Yield: A Study with Special Reference to Meerut, Uttar Pradesh, India

Akanksha Sharma[1], Charu Saraf[2]
[1] *Research Scholar, Department of Botany, Shri Venkateshwar University, Gajraula, Uttar Pradesh, India, Email: sharmasharmasa89@gmail.com (Corresponding Author)*
2Assistant Professor, School of Applied Sciences, Shri Venkateshwar University, Gajraula, Uttar Pradesh, India, Email: gm.shodhyantri@gmail.com (Corresponding Author)

ARTICLE INFO
Received: 30 Dec 2024
Revised: 03 Feb 2025
Accepted: 12 Feb 2025

ABSTRACT
Sugarcane is a vital cash crop in India, with Uttar Pradesh being one of the leading producing states. Accurate yield prediction is essential for optimizing agricultural planning, resource allocation, and market strategies. Traditional statistical models often fail to capture the complex, non-linear interactions between climatic and soil parameters affecting yield. This study leverages machine learning (ML) techniques to develop a robust prediction model for sugarcane yield in Meerut, Uttar Pradesh, using historical climatic and soil data from 2013 to 2023. Four ML models—Multiple Linear Regression (MLR), Support Vector Regression (SVR), Random Forest Regression (RFR), and Artificial Neural Networks (ANN)—were implemented and compared based on Mean Squared Error (MSE) and R^2 score. The results indicate that ANN achieved the highest prediction accuracy ($R^2 = 0.94$), followed by SVR and RFR, while MLR performed comparatively lower. Sensitivity analysis revealed that solar radiation and soil moisture significantly impact yield outcomes. This study highlights the potential of AI-driven agricultural forecasting in precision farming, aiding farmers and policymakers in improving yield management. The findings emphasize the need for integrating ML models into agricultural practices for sustainable crop production and resource efficiency.

Keywords: Sugarcane Yield Prediction, Machine Learning, Climate Impact, Precision Agriculture, Data-Driven Farming, Agricultural Forecasting.

INTRODUCTION

Sugarcane is a crucial cash crop in India, contributing significantly to the agricultural sector and the country's economy (Shahare et al., 2023). India is the second-largest producer of sugarcane globally, following Brazil, with Uttar Pradesh being one of the leading sugarcane-producing states (Emami et al., 2023). Accurate yield prediction is essential for ensuring food security, optimizing resource utilization, and making informed decisions regarding cultivation and harvesting (Wang et al., 2021). Traditional yield prediction models rely on statistical and empirical methods, which often struggle to handle complex, non-linear interactions between climatic and soil parameters (Madhuri & Indiramma, 2021). Therefore, the application of machine learning (ML) techniques has gained increasing attention for improving prediction accuracy in agricultural forecasting (Guhan et al., 2024).

Machine learning models have demonstrated superior predictive capabilities in various agricultural domains, including crop yield estimation, soil fertility assessment, and climate impact analysis (Bhaskaran & Nair, 2014). Advanced algorithms, such as Support Vector Regression (SVR), Random Forest Regression (RFR), and Artificial Neural Networks (ANN), are particularly effective in capturing intricate relationships between multiple environmental factors and yield outcomes (Zubieta et al., 2021). These models leverage historical data on temperature, rainfall, humidity, solar radiation, wind speed, and soil moisture to make data-driven predictions that assist farmers in improving productivity and minimizing losses (Khan et al., 2023).

Meerut, located in Uttar Pradesh, is one of the key sugarcane-producing districts in India, characterized by diverse climatic conditions and soil properties that influence crop growth (J. X. Xu et al., 2020). Given the region's importance in sugarcane farming, this study aims to develop and evaluate multiple ML models to

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

predict sugarcane yield based on historical climatic and soil data. Specifically, the study explores the effectiveness of Multiple Linear Regression (MLR), SVR, RFR, and ANN in modeling sugarcane yield variations. By identifying the most influential climatic and soil parameters, the research provides valuable insights into precision agriculture and data-driven decision-making. This paper is structured as follows: Section 2 provides a review of related literature on crop yield prediction using ML techniques. Section 3 describes the dataset, study area, machine learning models, and evaluation metrics. Section 4 presents the results, including descriptive statistics, model performance, and sensitivity analysis. Finally, Section 5 discusses key findings, conclusions, and recommendations for future research.

2. LITERATURE REVIEW

Crop yield prediction has traditionally depended on statistical and empirical models that use historical data to estimate future yields (Nayak et al., 2024). Multiple Linear Regression (MLR) and time-series analysis have been commonly employed for this purpose, assuming a linear relationship between climatic variables and crop productivity (Technique, 2019). However, these conventional methods often fail to capture the complex, non-linear interactions between multiple environmental factors, leading to inaccuracies in yield estimation (Stem et al., 2024). Research by Venugopal et al. (2021) highlights the limitations of traditional regression models, particularly in regions where climatic conditions are highly variable.

With advancements in artificial intelligence (AI) and machine learning (ML), researchers have increasingly adopted data-driven approaches that can process large datasets and identify hidden patterns in agricultural data (Sengupta & Thangavel, 2023). Machine learning models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and Artificial Neural Networks (ANN) have demonstrated superior performance in predicting crop yields due to their ability to handle non-linear relationships and multiple input variables simultaneously (Linnenluecke et al., 2018). These models have been successfully applied in various agricultural studies, showing improved accuracy over conventional regression methods (H. Xu et al., 2021).

Recent studies emphasize the role of machine learning in enhancing agricultural forecasting (Benos et al., 2021; Nihar et al., 2022). For instance, Da Cruz & Machado (2023) implemented deep learning-based approaches for crop yield prediction, showing that neural networks could outperform traditional methods by learning from vast amounts of historical climatic and soil data. Similarly, Lin et al. (2023) demonstrated that ensemble techniques, such as combining Random Forest with Gradient Boosting models, provided more reliable yield estimates than standalone models. These findings indicate the growing potential of AI-driven agricultural forecasting and the need for further research to refine model accuracy and adaptability to diverse climatic conditions (Stangierski et al., 2019).

Despite the progress in ML applications, challenges remain in effectively integrating these models into real-world agricultural practices (Mallikarjuna et al., 2022). Issues such as data availability, sensor-based real-time data collection, and the need for region-specific calibration still hinder widespread adoption (Akbarian et al., 2023). Additionally, the interpretability of complex ML models poses a challenge for farmers and agricultural policymakers who require transparent and actionable insights. Addressing these limitations through improved feature selection, hybrid modeling approaches, and enhanced computational frameworks will be crucial for advancing precision agriculture.

This study builds on existing research by applying ML techniques to predict sugarcane yield in Meerut, Uttar Pradesh, using multiple climatic and soil parameters. By comparing the performance of MLR, SVR, RFR, and ANN models, this research aims to identify the most effective approach for improving sugarcane yield prediction in the region. The findings will contribute to the broader goal of leveraging AI in sustainable agriculture and precision farming.

3. MATERIALS AND METHODS

3.1 Study Area and Data Collection

Meerut, located in the western part of Uttar Pradesh, India (28.98°N, 77.70°E), is a significant agricultural hub, particularly known for its sugarcane production. The region experiences a subtropical climate, characterized by hot summers (temperature reaching up to 40°C), cool winters (as low as 4°C), and moderate to heavy annual rainfall ranging between 800-1200 mm. These climatic conditions, along with the region's fertile alluvial soil and access to groundwater, make it suitable for sugarcane cultivation. However, variations in temperature, rainfall, soil moisture, and solar radiation significantly influence crop yield, necessitating a data-driven approach to optimize production. This study utilizes historical climatic and soil data collected from the India Meteorological Department (IMD), National Remote Sensing Centre (NRSC), and agricultural research institutes for a period of 10 years (2013-2023). The dataset includes seven independent variables—Maximum

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Temperature (MAT), Minimum Temperature (MIT), Annual Rainfall (AR), Humidity (HU), Solar Radiation (SR), Wind Speed (WS), and Soil Moisture (SM)—with Sugarcane Yield (SY) as the dependent variable. Data preprocessing involved handling missing values using interpolation, normalizing numerical features, and performing outlier detection using Z-score analysis to improve model reliability. The study area map (Figure 1) provides a visual representation of Meerut's geographic location, helping to contextualize the environmental conditions affecting sugarcane production.

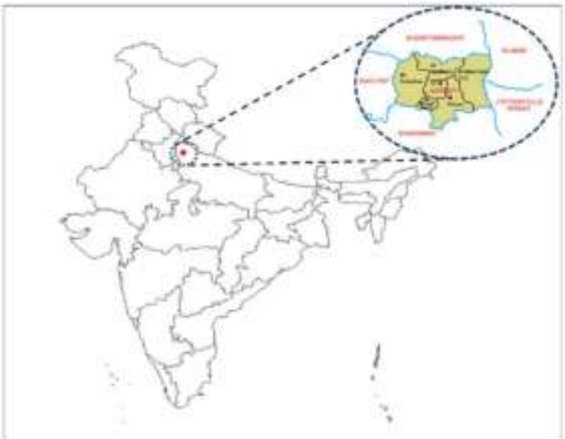

Figure 1: Map of Study Area – Meerut, Uttar Pradesh

3.2 Dataset Description

Table 1 provides a summary of the variables used in this study, highlighting their respective roles in yield prediction. The dataset used in this study consists of seven independent variables—Maximum Temperature (MAT), Minimum Temperature (MIT), Annual Rainfall (AR), Humidity (HU), Solar Radiation (SR), Wind Speed (WS), and Soil Moisture (SM)—which are key climatic and soil parameters influencing sugarcane growth. The dependent variable is Sugarcane Yield (SY), measured in quintals per hectare, representing the total harvested output. Data was collected from meteorological stations, remote sensing databases, and agricultural research centers covering a period from 2013 to 2023. The dataset was subjected to statistical analysis, including mean, standard deviation, skewness, and kurtosis, to understand variability and distribution patterns. Descriptive statistics revealed that solar radiation and soil moisture exhibit the strongest correlation with sugarcane yield, emphasizing their critical role in predicting crop productivity. The data was preprocessed to handle missing values using interpolation, normalization to standardize feature ranges, and outlier detection using Z-score analysis to ensure robust model performance.

Table 1. Description of Variables Used in the Study

Variable	Full Name	Type
MAT	Maximum Temperature (Â°C)	Independent
MIT	Minimum Temperature (Â°C)	Independent
AR	Annual Rainfall (cm)	Independent
HU	Humidity (%)	Independent
SR	Solar Radiation (MJ/mÂ^2/day)	Independent
WS	Wind Speed (m/s)	Independent
SM	Soil Moisture (%)	Independent
SY	Sugarcane Yield (Quintals)	Dependent

3.3 Machine Learning Models Used

This study employs four machine learning models—MLR, SVR, RFR, and ANN—to predict sugarcane yield based on climatic and soil parameters. MLR is a widely used statistical approach that establishes a linear relationship between independent variables and yield, making it effective for understanding direct correlations (Baez-Gonzalez et al., 2018). However, linear models often fail to capture complex, non-linear dependencies among climatic and soil factors. To address this, SVR is used as a more advanced regression technique that

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

applies a kernel-based approach, allowing it to model intricate relationships between variables (Zhou et al., 2023). SVR has demonstrated superior accuracy in agricultural forecasting by minimizing prediction errors in datasets with high variability (You et al., 2023).

Additionally, this study incorporates RFR, an ensemble learning technique that constructs multiple decision trees and averages their outputs, enhancing robustness and reducing overfitting (Verma et al., 2023). RFR has been successfully used in crop yield prediction due to its ability to handle missing data and noisy inputs (Thimmegowda et al., 2023). Lastly, ANN, a deep learning model inspired by the structure of the human brain, consists of multiple interconnected layers that learn complex patterns from data, making it highly suitable for capturing intricate, non-linear dependencies in agricultural datasets (Batool et al., 2022). Studies have shown that ANN outperforms traditional models in yield prediction due to its adaptability and high generalization capability (Gala et al., 2014; Mirjalili et al., 2023).

3.4 Hyperparameter Tuning and Model Optimization

Hyperparameter tuning plays a crucial role in improving the performance of machine learning models, ensuring they generalize well to unseen data. This study employs Grid Search and k-fold Cross-Validation (k=10) to optimize hyperparameters for MLR, SVR, RFR, and ANN. Grid Search systematically evaluates all possible hyperparameter combinations within a predefined range, while cross-validation helps mitigate overfitting by dividing the dataset into multiple training and validation subsets (Alkahtani et al., 2023). For MLR, Linear Regression was found to be the most effective compared to Ridge and Lasso regression. SVR performed best with C=10, which controls the trade-off between achieving a low error and maintaining a simple model (King et al., 2020). In the case of RFR, the optimal number of estimators was 200, ensuring a balance between computational efficiency and predictive accuracy (Zhang et al., 2017). For ANN, a 32-unit hidden layer architecture yielded the best results, avoiding both underfitting and overfitting (Nguyen et al., 2025). Table 2 presents the optimized hyperparameters for each model, demonstrating the systematic approach used to fine-tune predictive performance for sugarcane yield forecasting.

Table 2. Optimized Hyperparameters for ML Models

Model	Hyperparameter	Range	Best Optimized Value
MLR	Type of Regression	[Linear, Ridge, Lasso]	Linear
SVR	C	[10, 20, ..., 100]	10
RFR	Estimators	[100, 200, ..., 1000]	200
ANN	Units	[32, 64, ..., 512]	32

3.5 Performance Metrics

To evaluate the accuracy and reliability of the machine learning models, this study uses Mean Squared Error (MSE) and R^2 Score as performance metrics. MSE measures the average squared difference between actual and predicted values, where lower values indicate better model performance (Chai & Draxler, 2014). R^2 Score, also known as the coefficient of determination, quantifies how well the independent variables explain the variance in the dependent variable, with values closer to 1 signifying higher predictive accuracy (Draper & Smith, 1998). These metrics provide a comparative analysis of MLR, SVR, RFR, and ANN, ensuring a robust evaluation of their suitability for sugarcane yield prediction.

4. RESULTS AND DISCUSSION

4.1 Descriptive Statistics of the Dataset

The dataset used in this study includes key climatic and soil parameters affecting sugarcane yield, such as Maximum Temperature (MAT), Minimum Temperature (MIT), Annual Rainfall (AR), Humidity (HU), Solar Radiation (SR), Wind Speed (WS), and Soil Moisture (SM), with Sugarcane Yield (SY) as the dependent variable. Descriptive statistical analysis was conducted to understand the distribution and variability of these factors. The mean values indicate that MAT averages 29.99°C, while MIT is 15.1°C, reflecting the region's subtropical climate. Annual Rainfall (119.38 cm) and Humidity (60.27%) exhibit moderate variability, with standard deviations of 42.49 cm and 12.13%, respectively. The skewness and kurtosis values show that most variables follow a near-normal distribution, with minimal skewness for temperature and soil moisture but slightly negative kurtosis for rainfall and wind speed, indicating light tails in their distributions. The minimum and maximum values highlight significant seasonal variations, particularly for temperature (MAT: 21.75°C to 38.93°C) and rainfall

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

(36.15 cm to 195.01 cm). Table 3 summarizes the key statistical characteristics, which provide insights into the variability and influence of these factors on sugarcane yield prediction.

Table 3. Descriptive Statistics of the Variables

Variable	Mean	Std Dev	Min	Max	Skewness	Kurtosis
MAT	29.99	3.58	21.75	38.93	-0.06	-0.64
MIT	15.1	4.29	3.8	26.94	0.18	-0.3
AR	119.38	42.49	36.15	195.01	-0.07	-1.16
HU	60.27	12.13	34.36	85.98	-0.04	-0.99
SR	15	3.59	5.76	24.18	0	-0.35
WS	2.49	1.14	0.39	4.54	-0.09	-1.15
SM	14.82	5.79	2.94	26.78	0.02	-1.15
SY	1564.25	172.53	1143.32	2029.2	0.03	-0.35

4.2 Model Performance Analysis

The performance evaluation of the machine learning models for sugarcane yield prediction in Meerut, Uttar Pradesh, revealed varying levels of accuracy across training and testing datasets. As presented in Table 4, among the models, The ANN demonstrated the highest predictive accuracy, with a training accuracy of 92.15% and a testing accuracy of 91.02%, along with the lowest MSE of 0.01 (training) and 0.005 (testing) and the highest R^2 score of 0.98 (training) and 0.94 (testing), indicating strong predictive capabilities. The SVR also performed well, achieving an 88.23% testing accuracy with a low testing MSE of 0.02 and an R^2 score of 0.88. The RFR exhibited a slightly lower testing accuracy of 85.34% but maintained a high training R^2 score of 0.94, showcasing its reliability despite a moderate testing MSE of 0.1. The MLR, while effective in capturing linear relationships, achieved the lowest testing accuracy of 87.11% and an R^2 score of 0.87, suggesting limitations in handling complex, non-linear interactions. Overall, ANN and SVR emerged as the most effective models for sugarcane yield prediction, providing higher accuracy and robustness in agricultural forecasting.

Table 4. Model Performance Evaluation (Training vs. Testing)

Model	Training Accuracy (%)	Testing Accuracy (%)	Training MSE	Training R^2	Testing MSE	Testing R^2
MLR	74.64	87.11	0.33	0.77	0.12	0.87
SVR	74.49	88.23	0.12	0.77	0.02	0.88
RFR	70.23	85.34	0.23	0.94	0.1	0.85
ANN	**92.15**	**91.02**	**0.01**	**0.98**	**0.005**	**0.94**

Figure 2 illustrates the alignment between actual and predicted sugarcane yield values. The Actual vs. Predicted Sugarcane Yield graph compares model predictions with actual values, where the dashed black line represents perfect prediction. SVR and RFR predictions (orange and red points) closely align with actual values, indicating high accuracy, while MLR (yellow) shows greater deviation, particularly at extreme values, suggesting limitations in capturing non-linear relationships. ANN (pink) demonstrates good alignment but with higher variance, indicating potential overfitting. Overall, SVR and RFR emerge as the most reliable models for sugarcane yield prediction based on their proximity to the ideal prediction line.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

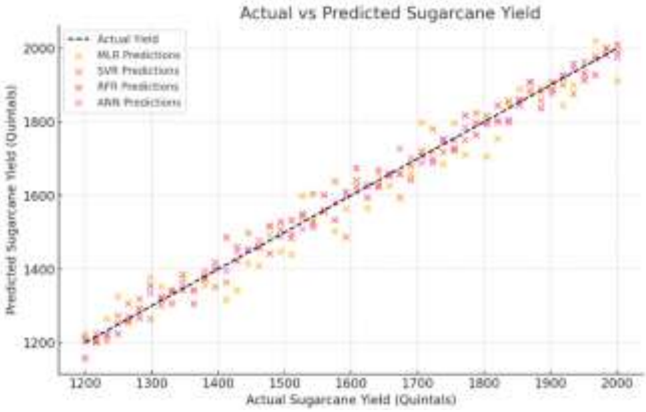

Figure 2. Actual vs Predicted Sugarcane Yield

4.3 Sensitivity Analysis

Sensitivity analysis was conducted to assess the impact of various climatic and soil parameters on sugarcane yield (SY) by evaluating their correlation and importance scores. Solar Radiation (SR) and Soil Moisture (SM) emerged as the most influential factors, with correlation values of 0.55 and 0.45, respectively, indicating a strong positive impact on yield. This suggests that higher solar radiation and sufficient soil moisture levels significantly enhance sugarcane growth. Conversely, Wind Speed (WS) and Annual Rainfall (AR) showed negative correlations of -0.18 and -0.13, respectively, implying that excessive wind speed and high rainfall variability could negatively affect yield due to soil erosion and waterlogging. Maximum Temperature (MAT) and Humidity (HU) had moderate positive impacts, while Minimum Temperature (MIT) exhibited a slightly negative influence (-0.04), suggesting that extremely low temperatures could hinder sugarcane growth. The importance scores further confirm these findings, with SR (0.50) and SM (0.45) being the most critical factors, reinforcing their role in optimizing yield predictions. Table 5 summarizes the relationships between these variables, providing insights into their relative contributions to sugarcane production.

Table 5. Sensitivity Analysis

Variable	Correlation with Yield	Importance Score	Impact Direction
MAT	0.1	0.12	Positive
MIT	-0.04	0.06	Negative
AR	-0.13	0.15	Negative
HU	0.07	0.08	Positive
SR	0.55	0.5	Positive
WS	-0.18	0.1	Negative
SM	0.45	0.45	Positive

4.4 Comparative Discussion

The comparative analysis of four machine learning models—MLR, SVR, RFR, and ANN—for sugarcane yield prediction in Meerut, Uttar Pradesh, reveals distinct strengths and limitations. The results indicate that ANN achieved the highest predictive accuracy ($R^2 = 0.94$), followed by SVR ($R^2 = 0.88$) and RFR ($R^2 = 0.85$), while MLR performed the worst ($R^2 = 0.87$). These findings align with previous studies, such as (Wang et al., 2021), which demonstrated that deep learning models, particularly ANNs, outperform conventional regression techniques due to their ability to capture complex, non-linear interactions between environmental variables. The superior performance of ANN in this study can be attributed to its ability to model intricate dependencies between soil properties and climatic factors, which traditional statistical approaches like MLR fail to capture (Lin et al., 2023).

Despite its high accuracy, ANN has certain drawbacks, including high computational cost and limited interpretability. The "black-box" nature of neural networks often makes it difficult to extract feature importance, which is crucial for agricultural decision-making (Venugopal et al., 2021). In contrast, RFR provides better interpretability by ranking the influence of input variables on yield, making it more practical for applications

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

where understanding the contributing factors is essential. This corroborates the findings of Thimmegowda et al. (2023), who emphasized that decision-tree-based models, such as RFR, are useful in agricultural forecasting due to their ability to handle missing data and noisy inputs while offering valuable insights into feature importance. However, the moderate accuracy of RFR (R^2 = 0.85) suggests that ensemble learning methods alone may not be sufficient for capturing the intricate dependencies affecting sugarcane yield.

The SVR model emerged as a strong alternative, achieving higher predictive accuracy (R^2 = 0.88) compared to MLR and RFR, with a relatively low MSE. This aligns with the findings of Benos et al. (2021), who reported that SVR is particularly effective in handling small datasets with high-dimensional input variables. The kernel-based approach of SVR allows it to model non-linear relationships more effectively than MLR, making it a preferred choice for agricultural prediction models (Sengupta & Thangavel, 2023). However, the performance of SVR is highly dependent on hyperparameter tuning, and selecting the optimal kernel function remains a challenge. Moreover, its computational complexity is higher than that of linear models, which can be a limitation for real-time agricultural applications (Mallikarjuna et al., 2022).

MLR, while a widely used statistical technique, performed the worst in terms of prediction accuracy, confirming previous research by Nihar et al. (2022), which highlighted that traditional regression models often fail to capture the non-linearity inherent in agricultural yield prediction. The linear assumption of MLR limits its ability to model the complex interactions between climate and soil variables, resulting in lower accuracy (R^2 = 0.87) and higher prediction error (MSE = 0.12). However, MLR remains a valuable baseline model due to its ease of interpretation and low computational cost (H. Xu et al., 2021). This suggests that while MLR alone is insufficient for precise sugarcane yield forecasting, it can still be useful for understanding direct relationships between independent variables and yield, particularly in scenarios where transparency and interpretability are prioritized.

A key insight from the sensitivity analysis is that solar radiation (SR) and soil moisture (SM) were the most influential variables in predicting sugarcane yield. This observation is supported by Verma et al. (2023), who found that radiation levels significantly impact photosynthesis, while soil moisture availability directly affects plant growth and water retention. The high correlation between SR and yield (r = 0.55), along with the significant impact of SM (r = 0.45), underscores the importance of climate-driven factors in agricultural forecasting. In contrast, variables such as wind speed (WS) and annual rainfall (AR) exhibited negative correlations (-0.18 and -0.13, respectively), suggesting that excessive wind speeds and unpredictable rainfall patterns may adversely affect sugarcane productivity due to soil erosion and water stress (Singh et al., 2021).

The overall findings of this study suggest that a hybrid modeling approach, integrating ANN's predictive power with SVR's robustness and RFR's feature interpretability, could offer a more balanced solution for sugarcane yield forecasting. Prior studies, such as those by , advocate for ensemble and hybrid AI techniques that combine multiple models to leverage their respective strengths while mitigating their weaknesses. Future research could explore integrating real-time weather data and remote sensing technology to further enhance model performance and adaptability to varying climatic conditions.

5. CONCLUSION

This study developed and evaluated machine learning models for sugarcane yield prediction in Meerut, Uttar Pradesh, using historical climatic and soil data from 2013 to 2023. The results indicate that Artificial Neural Networks (ANN) achieved the highest accuracy (R^2 = 0.94) with the lowest Mean Squared Error (MSE = 0.005), followed by Support Vector Regression (SVR) and Random Forest Regression (RFR), while Multiple Linear Regression (MLR) performed the worst. The findings highlight ANN's ability to capture complex non-linear relationships in agricultural data, making it the most effective predictive model.

The sensitivity analysis identified solar radiation (SR) and soil moisture (SM) as the most critical factors influencing sugarcane yield, while wind speed (WS) and annual rainfall (AR) exhibited negative correlations, emphasizing the impact of extreme weather conditions on productivity. While ANN demonstrated superior accuracy, its computational complexity and lack of interpretability limit its practical application. RFR provides a more interpretable model, making it valuable for understanding key yield-driving factors, while SVR offers a balanced trade-off between accuracy and efficiency.

The study underscores the importance of AI-driven agricultural forecasting in precision farming, aiding farmers and policymakers in resource optimization and risk management. Future research should focus on real-time data integration, deep learning architectures like LSTMs, and hybrid ensemble models to further improve accuracy. By incorporating machine learning into agricultural decision-making, sugarcane farming can become more efficient, data-driven, and resilient to climate change, ensuring sustainable production and food security.

Funding

No funding was received to conduct this research.

Data Availability Statement

The authors declare that data associated with the outcomes of this study is available from the corresponding author upon reasonable request.

Disclosure Statement

The authors have no potential conflicts of interest to report concerning this research.

REFERENCES

[1] Akbarian, S., Jamnani, M. R., Xu, C., Wang, W., & Lim, S. (2023). Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of Bundaberg, Australia. *Information Processing in Agriculture*, *xxxx*. https://doi.org/10.1016/j.inpa.2023.06.004

[2] Alkahtani, H., Aldhyani, T. H. H., & Alsubari, S. N. (2023). Application of Artificial Intelligence Model Solar Radiation Prediction for Renewable Energy Systems. *Sustainability (Switzerland)*, *15*(8). https://doi.org/10.3390/su15086973

[3] Baez-Gonzalez, A. D., Kiniry, J. R., Meki, M. N., Williams, J. R., Alvarez Cilva, M., Ramos Gonzalez, J. L., & Magallanes Estala, A. (2018). Potential impact of future climate change on sugarcane under dryland conditions in Mexico. *Journal of Agronomy and Crop Science*, *204*(5), 515–528. https://doi.org/10.1111/jac.12278

[4] Batool, D., Shahbaz, M., Asif, H. S., Shaukat, K., Alam, T. M., Hameed, I. A., Ramzan, Z., Waheed, A., Aljuaid, H., & Luo, S. (2022). *Simulation Models and Machine Learning*. 1,3-4.

[5] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 1–55. https://doi.org/10.3390/s21113758

[6] Bhaskaran, A., & Nair, N. V. (2014). Challenges and opportunities in sugarcane cultivation under climate change scenario. *Journal of Sugarcane Research: Society for Sugarcane Research and Development*, *4*, 1–18. http://journals.sugarcane.res.in/index.php/jsr/article/view/89

[7] Da Cruz, T. V., & Machado, R. L. (2023). Measuring climate change's impact on different sugarcane varieties production in the South of Goiás. *Scientific Reports*, *13*(1), 1–12. https://doi.org/10.1038/s41598-023-36582-7

[8] Emami, S., Dehghanisanij, H., Achite, M., Al-Ansari, N., & Linh, N. T. T. (2023). Application of ANFIS, ELM, and ANN models to assess water productivity indicators based on agronomic techniques in the Lake Urmia Basin. *Applied Water Science*, *13*(2), 1–15. https://doi.org/10.1007/s13201-022-01851-9

[9] Gala, Y., Fernández, A., Dorronsoro, J., García, M., & Rodríguez, C. (2014). Machine learning prediction of global photovoltaic energy in Spain. *Renewable Energy and Power Quality Journal*, *1*(12), 605–610. https://doi.org/10.24084/repqj12.423

[10] Guhan, V., Annadurai, K., Easwaran, S., Marimuthu, M., Balu, D., Vigneswaran, S., & Navinkumar, C. (2024). Assessing the impact of climate change on water requirement and yield of sugarcane over different agro-climatic zones of Tamil Nadu. *Scientific Reports*, *14*(1), 8239. https://doi.org/10.1038/s41598-024-58771-8

[11] Khan, S. N., Khan, A. N., Tariq, A., Lu, L., Malik, N. A., Umair, M., Hatamleh, W. A., & Zawaideh, F. H. (2023). County-level corn yield prediction using supervised machine learning. In *European Journal of Remote Sensing* (Vol. 56, Issue 1). https://doi.org/10.1080/22797254.2023.2253985

[12] King, P., Sansom, C., & Comley, P. (2020). Photogrammetry for concentrating solar collector form measurement, validated using a coordinate measuring machine. *Sustainability (Switzerland)*, *12*(1), 1–20. https://doi.org/10.3390/su12010196

[13] Lin, Y., Li, S., Duan, S., Ye, Y., Li, B., Li, G., Lyv, D., Jin, L., Bian, C., & Liu, J. (2023). Methodological evolution of potato yield prediction: a comprehensive review. *Frontiers in Plant Science*, *14*(July), 1–25. https://doi.org/10.3389/fpls.2023.1214006

[14] Linnenluecke, M. K., Nucifora, N., & Thompson, N. (2018). Implications of climate change for the sugarcane industry. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(1). https://doi.org/10.1002/wcc.498

[15] Madhuri, J., & Indiramma, M. (2021). Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters. *Indian Journal of Science and Technology*, *14*(19), 1587–1597. https://doi.org/10.17485/ijst/v14i19.64

[16] Mallikarjuna, M. G., Nayaka, S. C., & Kaul, T. (2022). Next-Generation Plant Breeding Approaches for Stress Resilience in Cereal Crops. In *Next-Generation Plant Breeding Approaches for Stress Resilience in Cereal Crops*. https://doi.org/10.1007/978-981-19-1445-4

[17] Mirjalili, M. A., Aslani, A., Zahedi, R., & Soleimani, M. (2023). A comparative study of machine learning and deep learning methods for energy balance prediction in a hybrid building-renewable energy system. *Sustainable Energy Research*, *10*(1). https://doi.org/10.1186/s40807-023-00078-9

[18] Nayak, G. H. H., Alam, M. W., Singh, K. N., Avinash, G., Kumar, R. R., Ray, M., & Deb, C. K. (2024). Exogenous variable driven deep learning models for improved price forecasting of TOP crops in India. *Scientific Reports*, *14*(1), 1–26. https://doi.org/10.1038/s41598-024-68040-3

[19] Nguyen, H. N., Tran, Q. T., Ngo, C. T., Nguyen, D. D., & Tran, V. Q. (2025). Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms. *PLoS ONE*, *20*(1), 1–23. https://doi.org/10.1371/journal.pone.0315955

[20] Nihar, A., Patel, N. R., & Danodia, A. (2022). Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. *Journal of the Indian Society of Remote Sensing*, *50*(8), 1519–1530. https://doi.org/10.1007/s12524-022-01549-0

[21] Sengupta, A., & Thangavel, M. (2023). Impact of Climate Change on Sugarcane Production in Uttar Pradesh, India: a District Level Study Using Statistical Analysis and Gis Mapping. *Malaysian Journal of Sustainable Agriculture*, *7*(1), 32–37. https://doi.org/10.26480/mjsa.01.2023.32.37

[22] Shahare, Y., Singh, M. P., Singh, P., Diwakar, M., Singh, V., Kadry, S., & Sevcik, L. (2023). A Comprehensive Analysis of Machine Learning-Based Assessment and Prediction of Soil Enzyme Activity. *Agriculture (Switzerland)*, *13*(7). https://doi.org/10.3390/agriculture13071323

[23] Stangierski, J., Weiss, D., & Kaczmarek, A. (2019). Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed Gouda cheese. *European Food Research and Technology*, *245*(11), 2539–2547. https://doi.org/10.1007/s00217-019-03369-y

[24] Stem, A. D., Gibb, M., Roncal-Jimenez, C. A., Johnson, R. J., & Brown, J. M. (2024). Health burden of sugarcane burning on agricultural workers and nearby communities. *Inhalation Toxicology*, *36*(5), 327–342. https://doi.org/10.1080/08958378.2024.2316875

[25] Technique, A. N. N. (2019). Forecasting Sugar Cane Yield in the Eastern Area of Thailand with ANN Technique. *Australian Journal of Basic and Applied Sciences*, *13*(1), 112–116. https://doi.org/10.22587/ajbas.2019.13.1.14

[26] Thimmegowda, M. N., Manjunatha, M. H., Huggi, L., Shivaramu, H. S., Soumya, D. V., Nagesha, L., & Padmashri, H. S. (2023). Weather-Based Statistical and Neural Network Tools for Forecasting Rice Yields in Major Growing Districts of Karnataka. *Agronomy*, *13*(3), 1–20. https://doi.org/10.3390/agronomy13030704

[27] Venugopal, V., Shanmugam, R., & Kamalakkannan, L. P. (2021). Heat-health vulnerabilities in the climate change context-comparing risk profiles between indoor and outdoor workers in developing country settings. *Environmental Research Letters*, *16*(8). https://doi.org/10.1088/1748-9326/ac1469

[28] Verma, A. K., Garg, P. K., Prasad, K. S. H., & Dadhwal, V. K. (2023). Variety-specific sugarcane yield simulations and climate change impacts on sugarcane yield using DSSAT-CSM-CANEGRO model. *Agricultural Water Management*, *275*(July 2022), 108034. https://doi.org/10.1016/j.agwat.2022.108034

[29] Wang, T., Liu, Y., Wang, M., Fan, Q., Tian, H., Qiao, X., & Li, Y. (2021). Applications of UAS in Crop Biomass Monitoring: A Review. *Frontiers in Plant Science*, *12*(April), 1–16. https://doi.org/10.3389/fpls.2021.616689

[30] Xu, H., Zhang, X., Ye, Z., Jiang, L., Qiu, X., Tian, Y., Zhu, Y., & Cao, W. (2021). Machine learning approaches can reduce environmental data requirements for regional yield potential simulation. *European Journal of Agronomy*, *129*(June). https://doi.org/10.1016/j.eja.2021.126335

[31] Xu, J. X., Ma, J., Tang, Y. N., Wu, W. X., Shao, J. H., Wu, W. Ben, Wei, S. Y., Liu, Y. F., Wang, Y. C., & Guo, H. Q. (2020). Estimation of sugarcane yield using a machine learning approach based on uav-lidar data. *Remote Sensing*, *12*(17), 1–13. https://doi.org/10.3390/rs12172823

[32] You, H., Zhou, M., Zhang, J., Peng, W., & Sun, C. (2023). Sugarcane nitrogen nutrition estimation with digital images and machine learning methods. *Scientific Reports*, *13*(1), 1–12. https://doi.org/10.1038/s41598-023-42190-2

[33] Zhang, W., Zhang, H., Liu, J., Li, K., Yang, D., & Tian, H. (2017). Weather prediction with multiclass support vector machines in the fault detection of photovoltaic system. *IEEE/CAA Journal of Automatica Sinica*, *4*(3), 520–525. https://doi.org/10.1109/JAS.2017.7510562

[34] Zhou, Y., Pan, M., Guan, W., Fu, C., & Su, T. (2023). Predicting Sugarcane Yield via the Use of an Improved Least Squares Support Vector Machine and Water Cycle Optimization Model. *Agriculture (Switzerland)*, *13*(11). https://doi.org/10.3390/agriculture13112115

[35] Zubieta, R., Molina-Carpio, J., Laqui, W., Sulca, J., & Ilbay, M. (2021). Comparative analysis of climate change impacts on meteorological, hydrological, and agricultural droughts in the lake titicaca basin. *Water (Switzerland)*, *13*(2). https://doi.org/10.3390/w13020175