IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1116~1129
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1116-1129 ❒ 1116

Optimizing potato crop productivity: a meteorological analysis and machine learning approach

Md. Jiabul Hoque[1,2], Md. Saiful Islam[2], Abdullah Al Noman[1], Md. Abrarul Hoque[1], Irfan A. Chowdhury[1], Mohammed Saifuddin[1]
[1]Department of Computer and Communication Engineering, International Islamic University Chittagong, Chattogram, Bangladesh
[2]Department of Electronics and Telecommunication Engineering, Chittagong University of Engineering and Technology, Chattogram, Bangladesh

Article Info

Article history:

Received Dec 11, 2023
Revised Nov 9, 2024
Accepted Nov 14, 2024

Keywords:

Correlation analysis
Early yield forecasting
Enhance potato production
Feature importance analysis
Feature selection
Machine learning
Random forest

ABSTRACT

Motivated by the critical need to enhance potato production in Bangladesh, particularly in the face of a changing climate, this study investigates the significant impact of weather on potato yield. This research employs various statistical and machine-learning approaches to identify key weather factors influencing potato crops. We utilize ANOVA F regression and random forest (RF) with feature importance analysis to pinpoint crucial monthly weather variables. Additionally, a correlation study employing Pearson's and Spearman's coefficients alongside p-values is conducted to determine the relationships between weather conditions and crop yield. Seaborn's bivariate kernel density estimation is then used to visualize ideal weather conditions for optimal harvests. Furthermore, to predict future yields, the study implements thoroughly trained and validated machine learning models including k-nearest neighbors (KNN), RF, and support vector regressor (SVR). Our analysis reveals that the RF model emerges as the most reliable predictor, achieving a high correlation coefficient (R^2=0.9990), and minimal error values (mean absolute percentage error (MAPE)=0.70, mean absolute error (MAE)=0.0803, and root mean square error (RMSE)=0.1114). These findings provide valuable insights to guide informed agricultural decisions and climate-related strategies, particularly for resource-limited countries like Bangladesh.

Corresponding Author:

Md. Saiful Islam
Department of Electronics and Telecommunication Engineering
Chittagong University of Engineering and Technology
Chattogram, Bangladesh
Email: saiful05eee@cuet.ac.bd

1. INTRODUCTION

Unpredictable weather patterns and resource limitations in Bangladesh pose a significant barrier to the continuous and sustainable production of potatoes, a vital commodity for millions of people worldwide [1]. Securing food supply for the growing nation's population depends on overcoming these issues. Researchers must focus on the sustainable growth of potato farming and the creation of new technologies to increase yields due to a growing population and limited arable land [2]. Strategic planning and technical innovation are crucial to ensure sustainable food security and to improve agricultural production in the long term. Despite recent developments, accurately predicting potato production early in the growth phase is still essential. This knowledge enables farmers to make informed decisions about resource distribution, use proactive management techniques, reduce possible production losses, and improve food security [3].

Ongoing studies have investigated different techniques to forecast potato crop output. However, these endeavors often face constraints [4], [5].

Improvements in potato yield forecasting are hindered by issues related to the precision and scope of the data. Contemporary studies frequently use past data to train machine learning algorithms, but encounter challenges due to the need for more data, limited coverage, and variations in data sources [6]–[8]. This study suggests improving the accuracy of the prediction in Bangladesh by collecting comprehensive information from reliable sources such as the Bangladesh Bureau of Statistics (BBS) and the Bangladesh Meteorological Department (BMD). Integrating various data collection methods and employing thorough cleaning established a cohesive dataset for more precise forecasts specific to Bangladesh.

Potato cultivation in Bangladesh is more difficult due to erratic weather patterns [9]. The prediction of potato yield traditionally occurs late in the growing season, which hinders farmers' ability to make quick decisions regarding resource allocation and mitigating techniques. Early yield prediction is vital to tackle these difficulties by providing information on elements that may contribute to crop performance. Afterward, farmers can take proactive steps, such as modifying irrigation techniques or using pest control strategies, to improve resource distribution and reduce adverse effects. Early projections provide for better risk management and agricultural planning, including harvest planning and market negotiation [10]. Although early prediction is crucial, several research studies concentrate on estimating production shortly before harvest, providing little time to implement remedial actions. This impedes critical decision-making and limits the potential effect of improving yield [11], [12]. Our study suggests a new method that combines sophisticated data collection techniques, machine learning models designed for early prediction, and feature relevance analysis to overcome these restrictions.

Climate, especially precipitation and temperature, significantly impacts potato production [13]. Studies show that excessive rainfall or high temperatures could reduce agricultural yields [14], [15]. The complex connection between weather conditions and crops highlights the challenges that climate change poses to food security. Although predictive models have advanced, it is still essential to understand the meteorological conditions vital for crop productivity. Many studies use historical data and machine learning for crop prediction but often need more specificity in pinpointing crucial weather variables [16], [17]. This study aims to fill this gap using feature significance analysis to pinpoint significant monthly weather elements. We want to use machine learning and historical data to reveal how weather affects potato yield. This will help stakeholders by providing specific monitoring and data-driven methods to improve prediction accuracy and agricultural decision-making.

It is essential to understand the complex connections between meteorological factors and potato production to guide agricultural methods, especially in regions such as Bangladesh, where erratic weather conditions have a notable impact on agricultural results. Recent studies have investigated several approaches to evaluate this connection, encountering constraints such as restricted scope, regional disparities, and the absence of temporal detail [18], [19]. This study uses correlation analysis to investigate the links between meteorological variables and potato yield in Bangladesh to improve agricultural practices and improve predictive models for better food security.

Developing accurate and timely forecasts of potato production is crucial to making informed decisions about resource distribution, using proactive management tactics, and reducing possible yield reductions [20]. Recent developments in machine learning provide opportunities for accurate crop yield forecasting, despite constraints in current initiatives. The issues involve a propensity to concentrate on a restricted range of machine learning methods, insufficient evaluations of model performance, and concerns about applicability in various geographical contexts [21]. This study examines the effectiveness of three well-known machine learning models, k-nearest neighbors (KNN), support vector regression (SVR), and random forest (RF), to overcome these constraints. The goals are to evaluate several machine learning models using a thorough dataset, to perform a detailed comparative study, and to ensure applicability using data from various geographical regions and types of potatoes. The study attempts to determine the most efficient model for predicting early potato yield to provide farmers with precise forecasting tools.

Recent studies have investigated many approaches to improve potato yield [22], [23]. However, there is a significant requirement for specific guidance designed for farmers in Bangladesh. Current studies on enhancing potato yield need to pay more attention to the difficulties of Bangladeshi farmers. They typically suggest general solutions that can be applied in many areas or provide technically sound but unfeasible suggestions due to limited resources [24]. Furthermore, there is frequently a restricted emphasis on the actual execution facets of suggested solutions [25]. The study intends to close the gap between research and practical implementation by offering context-specific solutions, emphasizing concrete tactics, and bridging the knowledge-practice gap. We aim to provide farmers with practical knowledge to reduce the impact of unfavorable weather, improve farming techniques, and sustainably increase potato yield by translating the study findings into clear and straightforward advice.

Predictive models have potential, but each crop thrives under certain weather conditions [26]. Accurate yield projections require suitable climate conditions and a deep understanding of how weather factors impact growth phases [27]. The specific influence of climate on potato yield in Bangladesh remains uninvestigated despite the acknowledged influence of climate on crop production. This study aims to identify the critical weather variables that influence potato cultivation in this region. We use several methods to achieve this goal, including empirical studies, sophisticated data processing, and machine learning. This study investigates the significant impact of meteorological conditions on potato yield to improve production in Bangladesh. The study uses ANOVA F regression and RF with the importance of features to determine important monthly weather components. A correlation analysis employing Pearson's and Spearman's coefficients and p-values establishes the relationships between weather conditions and crop yield. Seaborn's bivariate kernel density estimation determines the optimal weather conditions for achieving the highest harvest yields. Well-trained and validated machine learning models such as KNN, RF, and SVR are used to predict future yields. The RF model is highly reliable, with a strong correlation coefficient of 0.9990 and low error values: mean absolute percentage error (MAPE) of 0.70%, mean absolute error (MAE) of 0.0803, and root mean square error (RMSE) of 0.1014.

This study offers six key contributions: first, early yield prediction: predicts potato yield early in the cultivation phase, providing valuable information on crop success factors. Second, comprehensive data collection: uses reliable data from various Bangladesh organizations, enhancing the robustness of research. Third, analysis of the importance of features: identifies key monthly weather factors impacting yield, aiding in data improvement and modification. Fourth, correlation analysis: confirms the relationships between the weather variables (rainfall, temperature) and yield, deepening the understanding of potato farming in Bangladesh. Fifth, selection of machine learning models: RF is identified as the most effective model for early yield prediction after empirical investigation, and sixth practical recommendations: offers actionable advice to Bangladeshi farmers, including strategies for managing heavy rainfall, adapting to temperature variations, and utilizing precision agriculture technologies.

The article is organized as follows. Section 2 describes the models used, the technique applied, and the extent of data analysis. Section 3 examines the research results, highlighting the importance of characteristics, correlation analysis, and regression models. Section 4 concisely presents the study's findings and suggests areas for future research.

2. MATERIALS AND METHODS

2.1. Proposed crop-weather model

This study aims to develop a novel crop weather model that can predict early crop production by analyzing current weather conditions. The model will also identify the best weather conditions for maximizing crop yield and provide strategies to minimize the negative impact of adverse weather conditions and improve overall crop yield. Figure 1 shows the proposed novel crop-weather framework.

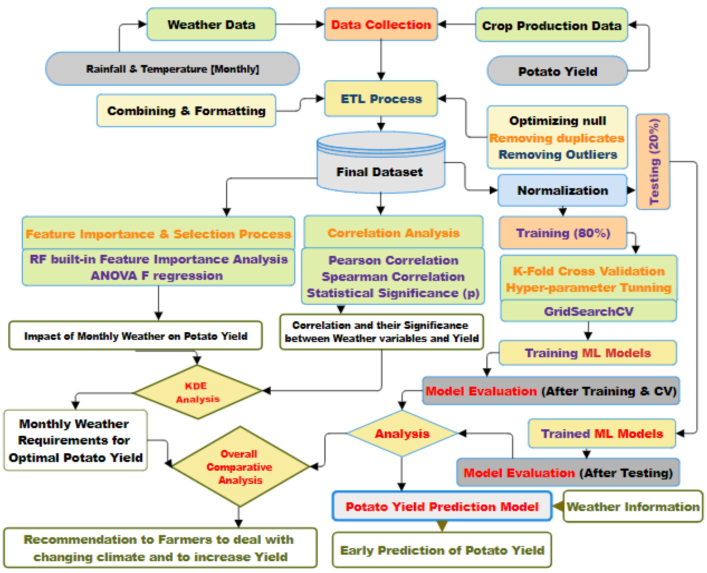


Figure 1. Proposed crop-weather framework

2.2. Dataset description

The study obtained data from reliable institutions, such as the Bangladesh Climate Information Management System (CIMSB) and the BMD, to collect information on rainfall and temperature. In addition, agricultural production data were acquired from the BBS and the Bangladesh Agricultural Research Institute (BARI). The extensive data set obtained from these credible sources enabled a meticulous analysis of the relationship between climatic variables and agricultural production, improving the reliability of the study's results. Table 1 summarises monthly rainfall, temperature, and potato yield data from 2,144 observations. Rainfall and temperature fluctuate monthly, with rain_1m averaging 34.62 mm and temp_4m peaking at 23.58 °C. The mean potato production is 12.04 tons per hectare, ranging from 7.83 to 19.89 tons per hectare. These data allow for investigating optimal growing conditions and early yield forecasts.

Table 1. Statistical summary of the data set

	rain_1m	rain_2m	rain_3m	rain_4m	temp_1m	temp_2m	temp_3m	temp_4m	Yield
count	2144	2144	2144	2144	2144	2144	2144	2144	2144
mean	34.62	9.01	7.37	19.87	23.58	19.63	18.36	20.81	12.04
std	37.38	12.65	6.90	13.98	0.68	0.73	0.73	1.01	3.20
min	0.00	0.00	0.00	0.00	22.02	17.95	16.55	18.50	7.83
25%	6.46	0.51	1.10	7.13	23.12	19.07	17.84	20.04	9.65
50%	21.44	2.95	4.84	18.72	23.60	19.66	18.33	20.73	10.98
75%	49.711	15.73	11.91	29.67	24.09	20.23	18.91	21.47	13.70
max	141.19	44.58	23.68	51.10	25.39	21.31	20.61	23.89	19.89

2.3. Feature-importance techniques

Feature importance techniques evaluate the extent to which each variable impacts the correctness of a model. These methods assign numerical values to the variables, focusing on the most significant ones to achieve precise predictions. This aids in improving our understanding of both the data and the model. We can determine the most significant and least significant aspects and enhance or fine-tune the data. Understanding feature relevance allows us to utilize feature selection approaches to retain only the most essential variables in the model. This optimizes the modelling process, accelerates it, and enhances the model performance [28]. This study used the built-in variable importance technique of the RF regression model with the help of the (1) to (4), and ANOVA F regression using (5) to determine the impacts of the monthly average rainfall and temperature predictors on the forecast of potato yield.

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{1}$$

$$f_i = \frac{\sum j{:}node\ j\ splits\ on\ feature\ i.ni_j}{\sum k \in all\ nodes.ni_k} \tag{2}$$

$$Normalized\ f_i = \frac{f_j}{\sum j \in all\ features.f_j} \tag{3}$$

$$RF\ f_i = \frac{\sum j \in all\ trees.Normalized\ f_{ij}}{T} \tag{4}$$

$$ANOVA\ F = \frac{(R_r - R_f)/(p-k)}{R_f/(n-p-1)} \tag{5}$$

2.4. Correlation between weather variables and crop yield

The study aimed to measure the connection between potato production and individual meteorological variables to determine their influence and to determine whether this impact is positive or negative. To satisfy this objective, this work employed Pearson's correlation using (6) to analyze linear relationships and Spearman's correlation with (7) to investigate non-linear relationships between meteorological factors and potato yield [29]. The statistical significance of these correlations was determined by calculating the p-values using (8).

$$\text{Pearson Correlation Coefficients } (r) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}} \tag{6}$$

$$Spearman\ Correlation\ (R_s) = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{7}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (8)$$

2.5. Machine learning models

Several research studies have consistently demonstrated the critical importance of machine learning as a tool to aid decision-making and predict agricultural productivity. Machine learning, an advanced technological tool, can help farmers reduce agricultural losses by offering precise crop recommendations and crucial insights [30]. This study focuses mainly on three machine learning models: RF, SVR, and KNN. The selection of these techniques is determined by the numerical nature of the forecast rather than its category aspect and the size of the data set.

3. RESULTS AND ANALYSIS

This section introduces a series of experimental results, starting with the analysis of the importance of characteristics to discover and study the essential aspects that affect potato yield. The study investigates linear and nonlinear connections and the statistical significance (p-value) between rainfall, temperature, and potato yield. It also examines the identification and evaluation of the optimal weather conditions necessary to achieve the highest potato output. Moreover, the section optimizes, trains, and verifies three prominent machine learning models, KNN, SVR, and RF, to predict early potato yield using weather parameters. The section closes with an offer for the most effective machine learning model for reliably predicting potato yield after a comprehensive study and comparison.

3.1. Importance analysis of characteristics

We evaluated the influence of changing rainfall and temperature on potato yields in Bangladesh during the October-January growing season by combining the importance of RF features and ANOVA approaches. Figures 2 and 3 provide a visual representation of the influence of monthly rainfall and monthly temperature, respectively, on the potato yield, elucidating the nuanced dynamics associated with each cultivation month. Figure 2 shows that December rainfall (rain_3m) in Bangladesh significantly impacts potato yields (score: 0.34). December marks a critical growth stage for tubers, which requires consistent moisture for optimal development. Lower temperatures in this month improve water retention, making tuber absorption more efficient. In contrast, rainfall in November (rain_3m) is less critical, with a score of 0.24, as attention turns to root and leaf growth, making the crop less responsive to changes in moisture levels. This knowledge guides customized water resource management practices in agriculture to maximize crop yields. Meanwhile, Figure 3 illustrates the essential elements of the potato production schedule related to temperature changes. October and January focus on root development and harvesting, while November is vital for tuber formation. December's (temp_3m) temperature has the highest impact (score: 0.33) on tuber growth and maturity, ultimately determining yield. Using adaptive management strategies to address changing temperature needs in potato farming is crucial to achieving the best possible production results.

3.2. Correlation between weather variables and potato yield

The effect of monthly rainfall on potato yield from October to January was analyzed using Pearson (linear) and Spearman (nonlinear) correlation coefficients, as shown in Table 2. The p-values assess the statistical significance of these connections. Table 2 shows the correlation between rainfall and potato yield from October (rain_1m) to January (rain_4m), showing persistent patterns. Higher rainfall negatively correlates with potato production, particularly in the first month (rain_1m), leading to a significant decline in yield caused by heavy rainfall in October. The third month (rain_3m) exhibits a less pronounced negative association, indicating a reduced impact of December (rain_3m) rainfall on yield. The statistical significance of all p-values, approaching 0, emphasizes the dependability of these results. The results highlight the need to control rainfall during the growing season, especially in October, a crucial early stage. Farmers can use this information to adjust their methods and improve potato yield in response to changing weather patterns. Consistent monitoring is essential for the successful implementation of the strategy in Bangladesh's monsoonal environment, particularly during the first month.

The relationship between monthly temperature and potato yield (October-January) was also analyzed using Pearson (linear) and Spearman (nonlinear) correlation coefficients in Table 3. The p-values determine the statistical significance of these connections. The correlation between temperature and potato yield fluctuates for four months. November (temp_2m) shows a robust positive association that is statistically significant. October (temp_1m) and January (temp_4m) exhibit a weak positive association that is statistically significant. In December (temp_3m), there is a statistically significant weak negative association. In summary, higher temperatures in November (temp_2m) are ideal to maximize potato production.

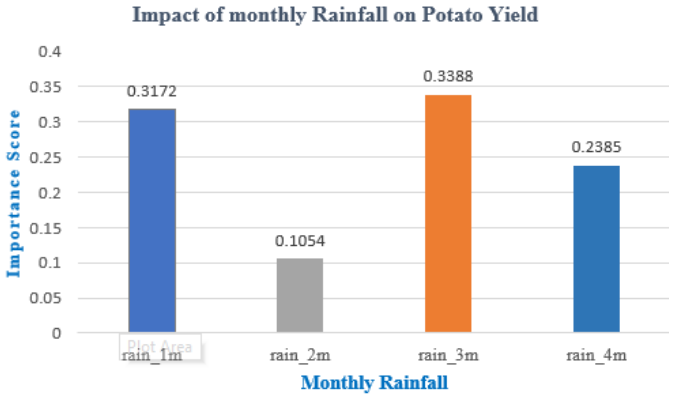


Figure 2. Impact of monthly rainfall on potato yield

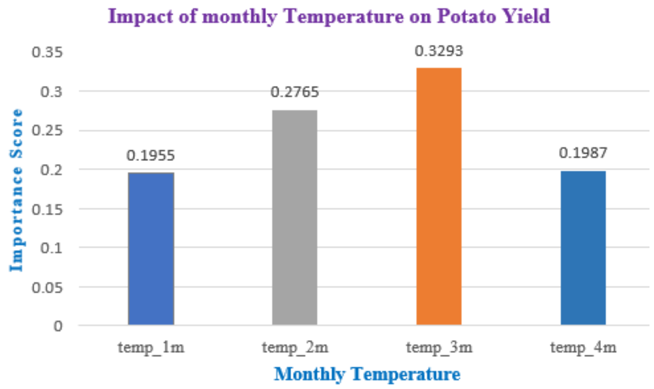


Figure 3. Impact of monthly temperature on potato yield

Table 2. Correlations and their statistical significance between monthly rainfall and potato yield

Months	Rainfall (mm)	Pearson correlation	p-value	Spearman correlation
rain_1m	34.62	-0.179	6.86E-17	-0.1715
rain_2m	9.01	-0.1597	1.02E-13	-0.0954
rain_3m	7.37	-0.0675	1.78E-03	-0.0047
rain_4m	19.87	-0.1634	2.67E-14	-0.1559

Table 3. Correlations and their statistical significance between monthly temperature and potato yield

Months	Temperature (°C)	Pearson correlation	p-value	Spearman correlation
temp_1m	23.58	0.0594	5.95E-03	0.0705
temp _2m	19.63	0.2391	3.02E-29	0.2032
temp _3m	18.36	-0.0781	2.92E-04	-0.0286
temp _4m	20.81	0.1135	1.38E-07	0.108

3.3. Monthly weather requirements for optimal potato yield

Figure 4 shows the optimum monthly rainfall for potato cultivation in Bangladesh. In October (Figure 4(a)), rainfall should ideally be below 21 mm, while the average is 34.62 mm. In such a situation, farmers can consider implementing drainage improvements. In November (Figure 4(b)), 5 mm (optimal) rainfall is related to a yield of 10-12 tons per hectare. In December (Figure 4(c)), 1-5 mm of rainfall is optimum to achieve the best yield. In January (Figure 4(d)), 5-15 mm of rainfall is associated with a yield of 12 tons per acre. This experiment indicates that farmers can effectively employ proactive strategies to regulate soil moisture levels to an optimal range throughout the growing season. By implementing this approach, they can establish optimal circumstances to grow potatoes, which can increase crop productivity.

The optimal temperature for the cultivation of potatoes in Bangladesh in October (Figure 5(a)) is 22.8-23.7 °C, with an average of 23.58 °C. The optimal temperature range for November (Figure 5(b)) is between 18.8 and 19.5 °C, with an average of 19.63 °C. The optimal temperature range for December (Figure 5(c)) is 17.6-18.5 °C, with an average of 18.36 °C. The optimal temperature range for January

(Figure 5(d)) is 20.2-21.1 °C, with an average of 20.81 °C. Bangladesh temperatures typically remain within ideal levels, leading to excellent crop yields.

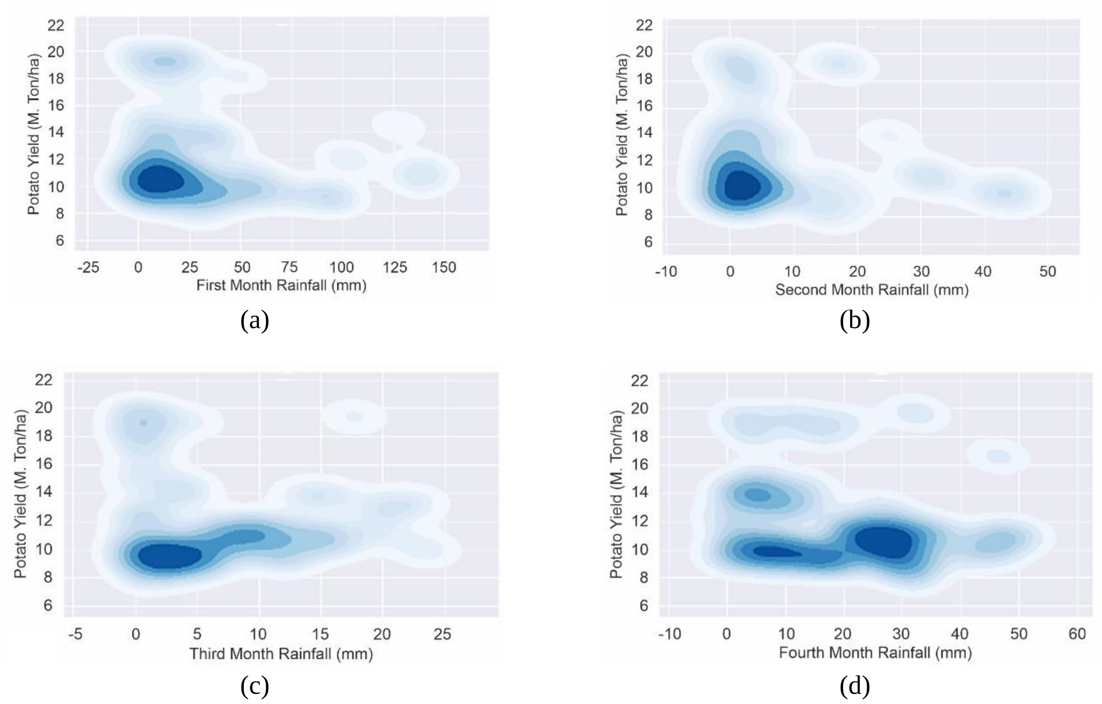


Figure 4. Correlation between rainfall and potato production: (a) first month rainfall vs. potato yield, (b) second month rainfall vs. potato yield, (c) third month rainfall vs. potato yield, and (d) fourth month rainfall vs. potato yield

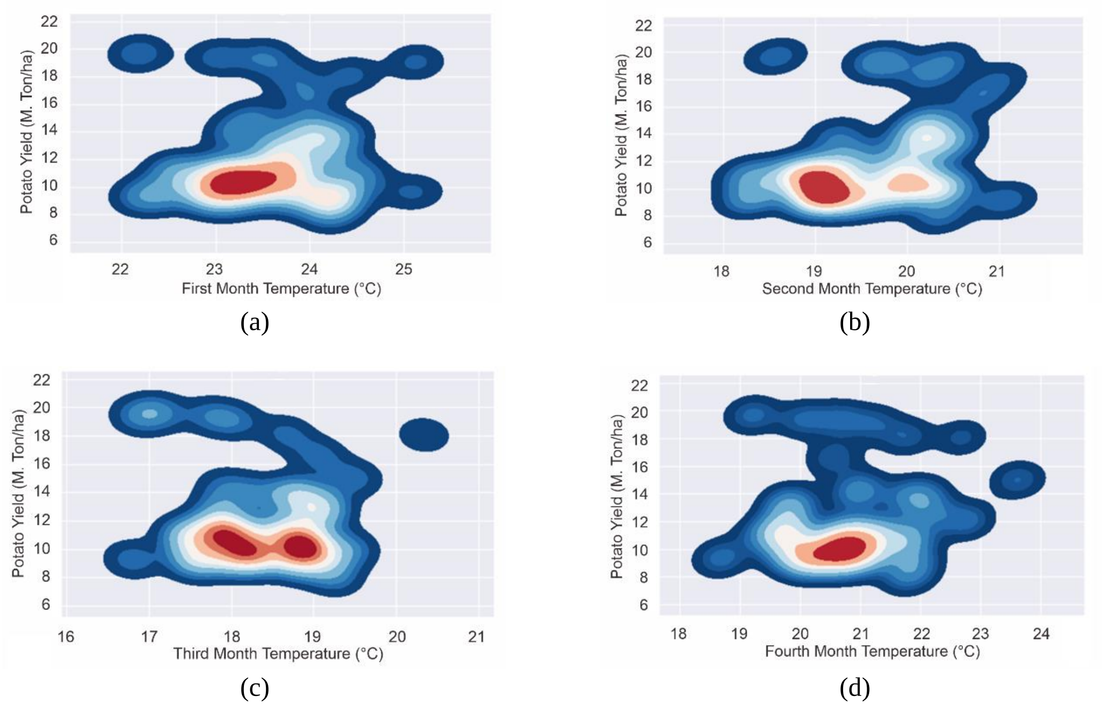


Figure 5. Correlation between temperature and potato production: (a) first-month temperature vs. potato yield, (b) second-month temperature vs. potato yield, (c) third-month temperature vs. potato yield, and (d) fourth-month temperature vs. potato yield

3.4. Machine learning-based prediction of potato yield
3.4.1. Model evaluation
The efficacy of machine learning models in predicting potato yield was evaluated using four commonly used assessment metrics: MAE, RMSE, MAPE, and coefficient of determination (R^2). MAE is the average difference between the predicted and actual yield, with lower values indicating better performance. RMSE assesses the magnitude of errors, with lower values indicating better performance. MAPE is a percentage of the actual yield, and lower values indicate better accuracy. R^2 is the proportion of variation explained by the model, with a value of 1 indicating a perfect match. These criteria help evaluate the performance of the model and select the most accurate model to predict potato production. The mathematical formula for MAE, RMSE, MAPE, and R^2 is expressed in (9) to (12) respectively. Here, y_{act} is the actual crop yields; y_{pred} denotes the predicted crop yields; y_{avg} represents the mean of crop yields and N is the total number of observations.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{pred} - y_{act}| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{pred} - y_{act})^2} \tag{10}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_{pred}-y_{act}|}{|y_{act}|} \tag{11}$$

$$R^2 = 1 - \frac{\Sigma(y_{pred}-y_{act})^2}{\Sigma(y_{pred}-y_{avg})^2} \tag{12}$$

3.4.2. Training and validating machine learning models for prediction
Cross-validation of K-folds is utilized to determine the optimal model setup and mitigate overfitting. GridsearchCV automates hyperparameter optimization for machine learning models. Training and testing Split the data into two sets: 80% for training and 20% for testing.

K selection: the value of k=5 was selected to balance bias and computational expense. Optimized models: K-neighbors regressor with 7 neighbors, SVR with a kernel of the radial basis function, a C value of 10.0, and an epsilon of 0.01. In addition, the RF regressor is set with a maximum depth of 10, 150 estimators, and a random state of 42.

The results offer optimized models for the precise prediction of potato yield. A comparison of model performance during 5-fold cross-validation, focusing on various metrics, is presented in Table 4. Lower MAE and RMSE values indicate more accuracy, and the RF model performs exceptionally well. MAPE measures the percentage error, while R^2 indicates the explained variance. A larger value is better for both metrics, and RF models are generally superior. Overall, the RF regressor model is optimal for forecasting potato yield due to its consistently excellent accuracy, precision, and stability. Adaptively depicts the intricate correlation between meteorological conditions and potato yield.

Table 4. Model performance through 5-fold cross-validation

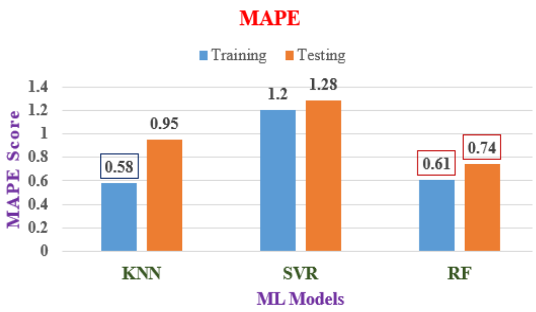
Fold	K-nearest neighbors		SVR		RF regressor	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.1282	0.3683	0.4944	1.1339	0.0848	0.1069
2	0.1155	0.1996	0.3211	0.5396	0.0816	0.1025
3	0.128	0.3457	0.3748	0.8793	0.0767	0.0981
4	0.1163	0.2975	0.4207	0.9456	0.0779	0.0983
5	0.1176	0.2556	0.3269	0.7091	0.0807	0.1011
Mean	0.1211	0.2933	0.3876	0.8415	0.0803	0.1014
Standard deviation	0.0057	0.061	0.0644	0.2032	0.0029	0.0032

3.4.3. Performance comparison of the models
Table 5 shows the performance of the models on both the training and the test data regarding MAE and RMSE. RF consistently performs in training and testing with the lowest MAE and RMSE. KNN exhibits strong training performance with low MAE but performs poorly in testing showing the sign of an overfitted model. SVR has lower accuracy than KNN and RF in training and testing. In summary, the RF model is highly reliable for forecasting potato yield due to its consistent performance in the training and testing stages.

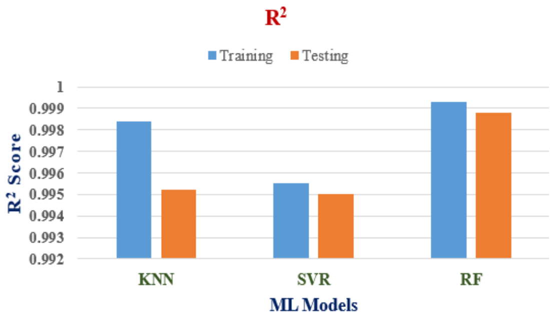
Table 5. Performance of the models in data gathering

	MAE		RMSE	
	Training	Testing	Training	Testing
KNN	0.0663	0.1107	0.1301	0.2115
SVR	0.1383	0.149	0.2165	0.2148
RF	0.0699	0.0829	0.088	0.1047

Figure 6 shows the comparative performance evaluation of three optimized machine learning models employed in this study, focusing on MAPE (Figure 6(a)) and R^2 (Figure 6(b)). The RF model consistently demonstrates low MAPE and high R^2 values in both training and testing datasets, indicating good precision. KNN has good performance, but it has a more significant discrepancy between training and testing data, making it less resilient. SVR has a lower accuracy than KNN and RF in training and testing. In summary, RF is the most reliable and precise model because of its consistently superior performance in all metrics (MAE, RMSE, MAPE, and R^2) during the training and testing stages.



(a)



(b)

Figure 6. Performance comparison of machine learning models for (a) MAPE and (b) R^2

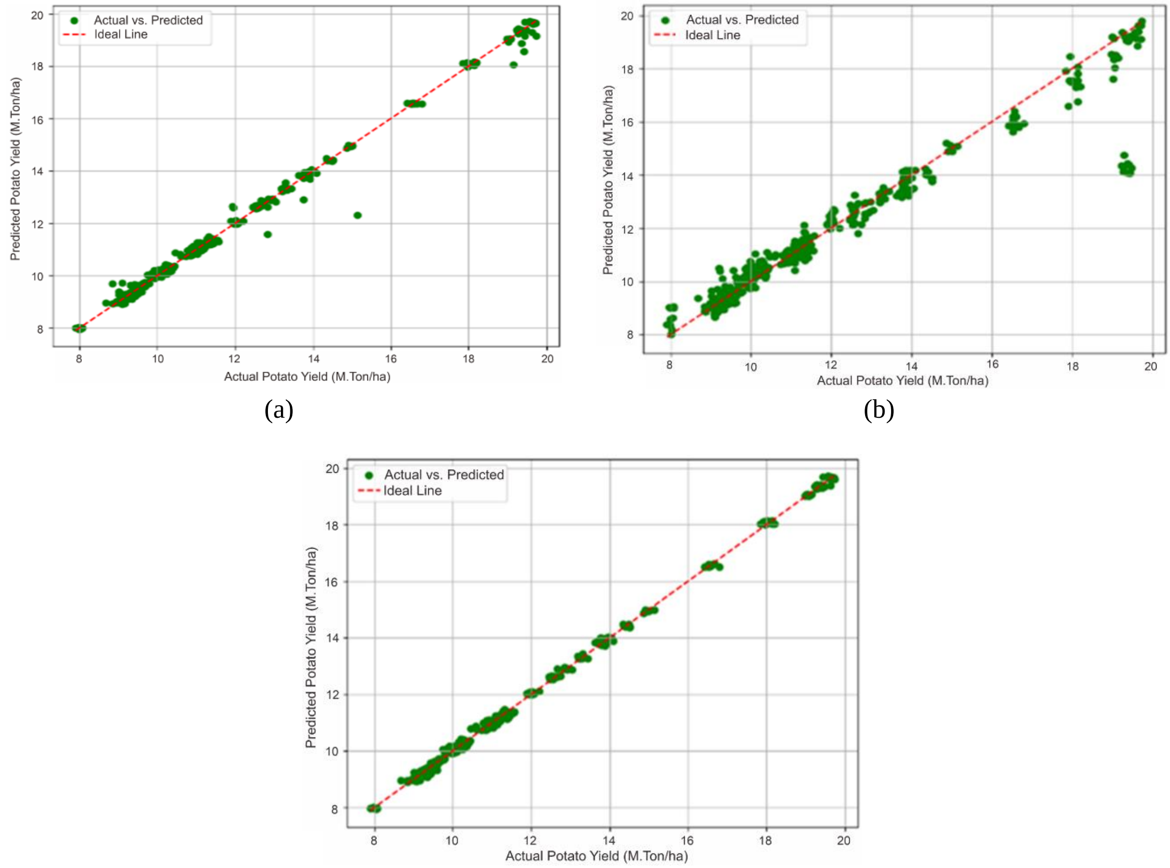
Table 6 compares the state-of-the-art approaches and suggested methods regarding the MAE, RMSE, and R^2 assessment metrics. The AtLSTM model [31] shows promise with its reported RMSE of 0.41. However, the lack of other metrics prevents a conclusive conclusion. The hybrid random forest-reinforcement learning (RF-RL) [32] exhibits lower MAE and RMSE, but its lower R^2 value of 0.87 indicates potential downsides due to higher complexity. Although online sequential extreme learning machines coupled with ant colony optimization (ACO-OSELM) [33] and data transfer rate (DTR) [34] demonstrate good R^2 values (0.990 and 0.997, respectively), their significantly larger MAE and RMSE in comparison to our RF model (for example, ACO-OSELM has an MAE of 42.37) indicate a possible issue of overfitting. Moreover, the absence of comprehensive measurements for DTR poses a barrier to conducting a thorough comparison. Improved cat swarm optimization based recurrent neural network (ICSO-RNN) [35] demonstrates a high R^2 value comparable to our proposed model. However, its significantly higher MAE and RMSE values suggest a less accurate overall fit. On the other hand, the proposed RF model exhibits numerous advantages. It attains the most minimal MAE and RMSE values that have been published, indicating a higher level of accuracy in fitting and improved aptitude for generalization. Moreover, the exceptionally high R^2 indicates a robust ability to explain the variations in potato yield. Significantly, the RF model attains this level of accuracy using a more straightforward methodology in contrast to approaches prone to overfitting.

Table 6. Comparative analysis of state-of-the-art methods and the proposed RF method

Authors	Methods	MAE	RMSE	$\mathbf{R^2}$
Liu *et al*. [31]	AtLSTM	N/A	0.41	0.810
Elavarasan and Vincent [32]	Hybrid RF-RL	0.17	0.23	0.870
Ali *et al*. [33]	ACO-OSELM	42.37	67.12	0.990
Ahmed *et al*. [34]	DTR	N/A	1.03	0.997
Reddy and Madapuri [35]	ICSO-RNN	0.77	0.82	0.990
Our proposed method	RF	0.08	0.10	0.999

3.4.4. Correlation between the predicted and actual potato yield

Figure 7 shows the relationship between the projected potato yield and the actual potato yield. The KNN (Figure 7(a)) model is generally precise with sporadic outliers, suggesting challenges within certain yield intervals. SVR models (Figure 7(b)) generally provide accurate forecasts, although there is some variability, indicating difficulties in correctly projecting specific yields. The data points are closely grouped around the diagonal line in the RF model (Figure 7(c)), indicating exact and reliable yield forecasts. In conclusion, RF is the optimal model for forecasting potato production because it provides accurate predictions that closely match actual yields throughout the range.



(a) (b) (c)

Figure 7. Correlation between predicted and actual potato yield (a) KNN, (b) SVR, and (c) RF

4. DISCUSSION

Controllable elements such as fertilizers and crop varieties impact productivity, while uncontrolled factors such as weather play a significant role, but are challenging to forecast [36]. This study aims to predict potato productivity in Bangladesh using easily accessible monthly rainfall and temperature data. The study examines the impact of monthly rainfall and temperature on potato production considering the varying requirements of the crop at different growth stages, offering significant insight into the intricate connection between meteorological conditions and crop yield.

4.1. The growing stages of the potato plants and their climate requirements

Water requirements at different growth stages have a substantial impact on potato yield. This study identifies the ideal rainfall and temperature levels required for each growth stage, sprout development, vegetative growth, tuber start, bulking, and maturity, to obtain an optimal yield of 10,000-12,000 tons per hectare. Figure 8 illustrates the necessary conditions for rainfall and temperature to achieve the highest possible potato production. Significant discoveries include: during the first month of sprout development, ideal conditions are 1-21 mm (Figure 8(a)) of rainfall and temperatures between 22.8-23.7 °C (Figure 8(b)). In the second month, which includes vegetative growth and tuber initiation, optimal parameters are 0-5 mm

(Figure 8(a)) rainfall and temperatures ranging from 18.8-19.5 °C (Figure 8(b)). For the third month, during tuber bulking, favorable conditions consist of 1-5 mm (Figure 8(a)) rainfall and temperatures between 17.6-18.5 °C (Figure 8(b)). Finally, during the fourth month of maturity, the preferred conditions are 5-15 mm (Figure 8(a)) of rainfall and temperatures ranging from 20.2-21.1 °C (Figure 8(b)). The results provide significant information on improving potato production by implementing efficient weather control techniques.

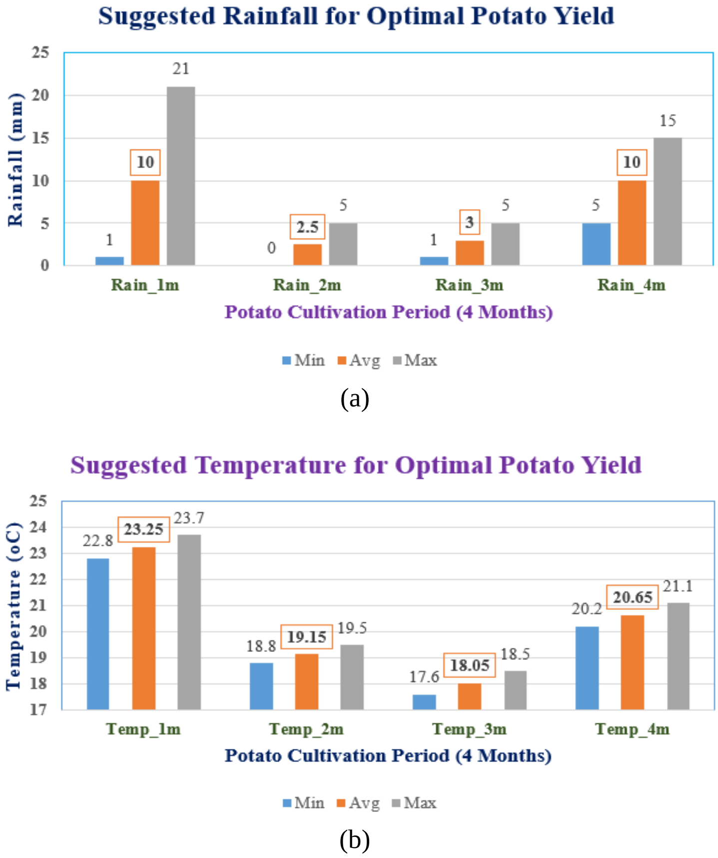


Figure 8. Optimal climate requirement for optimal potato yield (a) rainfall and (b) temperature

4.2. Recommendation to farmers

This study examined how meteorological conditions affect potato production in Bangladesh and provided helpful suggestions to farmers. Farmers can mitigate excessive rainfall and waterlogging by improving drainage systems, choosing moisture-tolerant crop varieties, implementing climate-resilient techniques, using weather monitoring devices, and adopting precision agriculture for water and nutrient control, particularly during the planting and harvest period of October to January. Continuous monitoring of temperature during the growing season is essential. Planting efforts should be adjusted according to optimal temperature conditions, especially in November (2nd month). Farmers should consider possible adverse impacts in December (3rd month) and use temperature-sensitive methods and robust cultivars to reduce crop losses. Bangladeshi farmers must adopt a comprehensive strategy that regulates temperature and rainfall, resilient practices and cultivars, and continuous monitoring, knowledge acquisition, and collaboration within the agricultural community to effectively address climate challenges, increase potato yields, and guarantee long-term sustainability and food security.

4.3. Prediction and the performance of models

The study investigated the application of machine learning to estimate potato yield in Bangladesh. The RF model is the most effective among the three models based on all evaluation metrics (MAE, RMSE, MAPE, and R^2). This emphasizes machine learning's potential for early yield prediction, mainly due to the easily accessible climatic data (precipitation and temperature) and its substantial impact on potato harvests.

However, restrictions on data accessibility for additional pertinent elements, such as soil moisture and solar radiation, impeded further enhancement. Future research will use the internet of things (IoT) to collect data on other characteristics. The goal is to incorporate these data into the RF model to improve the accuracy and completeness of potato yield forecasts in Bangladesh. This method shows potential to improve farming techniques and guarantee food security in the area.

5. CONCLUSION

This study examines the prediction of potato yield in Bangladesh, specifically focusing on monthly rainfall and temperature data. Anticipating crop yields in advance can help prepare for and reduce food scarcity. The study determines that the RF model is the most effective predictor among three machine learning models (KNN, RF, and SVR) when evaluated using government data. The work demonstrates the statistical importance of temperature and precipitation in yield using accurate data and focusing on data integrity. Examining the importance of features indicates that December (3^{rd} month) rainfall has a significant impact, while temperature plays a subtle but important role during the growing season. Weather-yield correlations confirm these findings, providing practical insights presented visually for farmers. Recommendations include improving drainage systems, using climate-resilient techniques, and adopting precision agriculture methods customized to meteorological conditions. Continuous monitoring, adaptive strategies, and teamwork are essential to address climate concerns. The study models, especially RF, are promising, but future research seeks to improve forecasts by including elements such as solar radiation, wind speed, and soil moisture. This will require combining the RF model with the IoT technologies to enable ongoing data collection and precise yield forecasting. This comprehensive strategy represents the evolving direction of agricultural research, combining evidence-based knowledge and technological progress to encourage sustainable and eco-friendly farming methods.

REFERENCES

[1] M. S. A. Salan, M. M. Hossain, I. H. Sumon, M. M. Rahman, M. A. Kabir, and A. K. Majumder, "Measuring the impact of climate change on potato production in Bangladesh using bayesian hierarchical spatial-temporal modeling," *PLOS ONE*, vol. 17, no. 11, Nov. 2022, doi: 10.1371/journal.pone.0277933.

[2] A. Devaux, J.-P. Goffart, P. Kromann, J. Andrade-Piedra, V. Polar, and G. Hareau, "The potato of the future: opportunities and challenges in sustainable agri-food systems," *Potato Research*, vol. 64, no. 4, pp. 681–720, Dec. 2021, doi: 10.1007/s11540-021-09501-4.

[3] R. Abbasi, P. Martinez, and R. Ahmad, "The digitization of agricultural industry–a systematic literature review on agriculture 4.0," *Smart Agricultural Technology*, vol. 2, Dec. 2022, doi: 10.1016/j.atech.2022.100042.

[4] U. Ahmad and L. Sharma, "A review of best management practices for potato crop using precision agricultural technologies," *Smart Agricultural Technology*, vol. 4, Aug. 2023, doi: 10.1016/j.atech.2023.100220.

[5] Y. Lin *et al.*, "Methodological evolution of potato yield prediction: a comprehensive review," *Frontiers in Plant Science*, vol. 14, Jul. 2023, doi: 10.3389/fpls.2023.1214006.

[6] C. Maraveas, P. G. Asteris, K. G. Arvanitis, T. Bartzanas, and D. Loukatos, "Application of bio and nature-inspired algorithms in agricultural engineering," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1979–2012, Apr. 2023, doi: 10.1007/s11831-022-09857-x.

[7] M. H. Mobarak *et al.*, "Scope of machine learning in materials research—a review," *Applied Surface Science Advances*, vol. 18, Dec. 2023, doi: 10.1016/j.apsadv.2023.100523.

[8] M. V. Subbarao, K. Venkatarao, S. Chittineni, and S. Kompella, "Utilizing deep learning, feature ranking, and selection strategies to classify diverse information technology ticketing data effectively," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 4, pp. 1985-1994, Dec. 2023, doi: 10.11591/ijai.v12.i4.pp1985-1994.

[9] A. Jannat, Y. Ishikawa-Ishiwata, and J. Furuya, "Assessing the impacts of climate variations on the potato production in Bangladesh: a supply and demand model approach," *Sustainability*, vol. 13, no. 9, Apr. 2021, doi: 10.3390/su13095011.

[10] E. Yeleliere, P. Antwi-Agyei, and L. Guodaar, "Farmers response to climate variability and change in rainfed farming systems: Insight from lived experiences of farmers," *Heliyon*, vol. 9, no. 9, Sep. 2023, doi: 10.1016/j.heliyon.2023.e19656.

[11] J. Liang, H. Li, N. Li, Q. Yang, and L. Li, "Analysis and prediction of the impact of socio-economic and meteorological factors on rapeseed yield based on machine learning," *Agronomy*, vol. 13, no. 7, Jul. 2023, doi: 10.3390/agronomy13071867.

[12] B. Panigrahi, K. C. R. Kathala, and M. Sujatha, "A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models," *Procedia Computer Science*, vol. 218, pp. 2684–2693, 2023, doi: 10.1016/j.procs.2023.01.241.

[13] T. Adekanmbi *et al.*, "Assessing future climate change impacts on potato yields—a case study for Prince Edward Island, Canada," *Foods*, vol. 12, no. 6, Mar. 2023, doi: 10.3390/foods12061176.

[14] M. Heino *et al.*, "Increased probability of hot and dry weather extremes during the growing season threatens global crop yields," *Scientific Reports*, vol. 13, no. 1, Mar. 2023, doi: 10.1038/s41598-023-29378-2.

[15] U. Sekar and K. Selvarajan, "Machine learning approach for predicting heart and diabetes diseases using data-driven analysis," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 4, pp. 1687-1694, Dec. 2023, doi: 10.11591/ijai.v12.i4.pp1687-1694.

[16] B. Dey, J. Ferdous, and R. Ahmed, "Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables," *Heliyon*, vol. 10, no. 3, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25112.

[17] E. Elbasi *et al.*, "Crop prediction model using machine learning algorithms," *Applied Sciences*, vol. 13, no. 16, Aug. 2023, doi: 10.3390/app13169288.

[18] P. Ekanayake, W. Rankothge, R. Weliwatta, and J. W. Jayasinghe, "Machine learning modelling of the relationship between weather and paddy yield in Sri Lanka," *Journal of Mathematics*, vol. 2021, pp. 1–14, May 2021, doi: 10.1155/2021/9941899.

[19] A. Liu, Z. Zhang, M. J. C. Crabbe, and L. C. Das, "Multifractal features and dynamical thresholds of temperature extremes in Bangladesh," *Fractal and Fractional*, vol. 7, no. 7, Jul. 2023, doi: 10.3390/fractalfract7070540.

[20] M. J. Hoque, M. S. Islam, and M. Khaliluzzaman, "A fuzzy logic-and internet of things-based smart irrigation system," in *The 10th International Electronic Conference on Sensors and Applications*, Nov. 2023, doi: 10.3390/ecsa-10-16243.

[21] L. Chen *et al.*, "Artificial intelligence-based solutions for climate change: a review," *Environmental Chemistry Letters*, vol. 21, no. 5, pp. 2525–2557, Oct. 2023, doi: 10.1007/s10311-023-01617-y.

[22] S. F. Ahmed *et al.*, "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023, doi: 10.1007/s10462-023-10466-8.

[23] I. A. Chincinska, M. Miklaszewska, and D. Sołtys-Kalina, "Recent advances and challenges in potato improvement using CRISPR/Cas genome editing," *Planta*, vol. 257, no. 1, Jan. 2023, doi: 10.1007/s00425-022-04054-3.

[24] M. J. Hoque *et al.*, "Incorporating meteorological data and pesticide information to forecast crop yields using machine learning," *IEEE Access*, vol. 12, pp. 47768–47786, 2024, doi: 10.1109/ACCESS.2024.3383309.

[25] X. Yang, Z. Hua, L. Li, X. Huo, and Z. Zhao, "Multi-source information fusion-driven corn yield prediction using the random forest from the perspective of agricultural and forestry economic management," *Scientific Reports*, vol. 14, no. 1, Feb. 2024, doi: 10.1038/s41598-024-54354-9.

[26] I. K. Nti, A. Zaman, O. Nyarko-Boateng, A. F. Adekoya, and F. Keyeremeh, "A predictive analytics model for crop suitability and productivity with tree-based ensemble learning," *Decision Analytics Journal*, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100311.

[27] J. Hoque, S. Islam, Khaliluzzaman, A. Al Muntasir, and M. A. Mohsin, "Revolutionizing malaria diagnosis: deep learning-powered detection of parasite-infected red blood cells," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 4, pp. 4518–4530, 2024, doi: 10.11591/ijece.v14i4.pp4518-4530.

[28] H.-N. Huang *et al.*, "Employing feature engineering strategies to improve the performance of machine learning algorithms on echocardiogram dataset," *DIGITAL HEALTH*, vol. 9, Jan. 2023, doi: 10.1177/20552076231207589.

[29] R. M. Elzaki, M. M. M. Elrasheed, and N. A. Elmulthum, "Optimal crop combination under soaring oil and energy prices in the kingdom of Saudi Arabia," *Socio-Economic Planning Sciences*, vol. 83, Oct. 2022, doi: 10.1016/j.seps.2022.101367.

[30] S. K. S. Durai and M. D. Shamili, "Smart farming using machine learning and deep learning techniques," *Decision Analytics Journal*, vol. 3, Jun. 2022, doi: 10.1016/j.dajour.2022.100041.

[31] Y. Liu, S. Wang, J. Chen, B. Chen, X. Wang, D. Hao, and L. Sun, "Rice yield prediction and model interpretation based on satellite and climatic indicators using a transformer method," *Remote Sensing*, vol. 14, no. 19, 2022, doi: 10.3390/rs14195045.

[32] D. Elavarasan and P. M. D. R. Vincent, "A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 11, pp. 10009–10022, Nov. 2021, doi: 10.1007/s12652-020-02752-y.

[33] M. Ali *et al.*, "Coupled online sequential extreme learning machine model with ant colony optimization algorithm for wheat yield prediction," *Scientific Reports*, vol. 12, no. 1, Mar. 2022, doi: 10.1038/s41598-022-09482-5.

[34] Z. U. Ahmed, M. G. Mortuza, M. J. Uddin, M. H. Kabir, M. Mahiuddin, and M. J. Hoque, "Internet of things based patient health monitoring system using wearable biomedical device," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*, Dec. 2018, pp. 1–5, doi: 10.1109/CIET.2018.8660846.

[35] D. J. Reddy and R. K. Madapuri, "Agriculture crop yield prediction using inertia based cat swarm optimization," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 2, Apr. 2024, doi: 10.11591/ijece.v14i2.pp1700-1710.

[36] L. T. Mbah, E. L. Molua, E. Bomdzele, and B. M. J. Egwu, "Farmers' response to maize production risks in Cameroon: an application of the criticality risk matrix model," *Heliyon*, vol. 9, no. 4, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15124.

BIOGRAPHIES OF AUTHORS

Md. Jiabul Hoque a prominent faculty member under the Faculty of Science & Engineering at International Islamic University Chittagong, is a distinguished Ph.D. research fellow at Chittagong University of Engineering and Technology (CUET). He holds an M.Sc. in CSE from the University of Greenwich, London, UK, and a B.Sc. in CSE from CUET. With a strong focus on machine learning and IoT, he has authored over 30 research articles in prestigious international journals and conferences. He actively participates in workshops, seminars, and symposiums in these fields, contributing significantly to academic and scientific progress. His dedication and expertise are driving advancements in these domains, making him a recognized figure in the global research community. He can be contacted at email: jia99cse@gmail.com or jiabul.hoque@iiuc.ac.bd.

Dr. **Md. Saiful Islam** is an Associate Professor of Department of Electronics and Telecommunication Engineering at Chittagong University of Engineering and Technology, Chittagong, Bangladesh. He received a B.Sc. degree in electrical and electronic engineering from Chittagong University of Engineering and Technology in 2010. He obtained M.S. and Ph.D. degree from the University of Ulsan, Korea. His research interests include artificial intelligence, signal processing, image and video processing, fault detection, and diagnosis. The above research lines have produced more than 50 publications in international journals, presentations at international conferences, and book chapters. He also acts as a referee in several highly reputed journals and international conferences. He can be contacted at email: saiful05eee@cuet.ac.bd.

Abdullah Al Noman is a graduate student with a Bachelor of Science degree in computer and communication engineering from the International Islamic University Chittagong (IIUC). His research interests are primarily in deep learning-based network security, and he is currently involved in four research projects in this field. He is dedicated to advancing network technology through machine learning and has developed a strong understanding of these areas during his academic journey. He is committed to contributing to the progression of machine learning, artificial intelligence, network security, and threat analysis. He is determined to pursue further education and research in these subjects. He can be contacted at email: noman.cce@gmail.com.

Md. Abrarul Hoque holds a Bachelor of Science degree in computer and communication engineering from the International Islamic University Chittagong. His research interests include information technology, cybersecurity, artificial intelligence, machine learning, internet of things (IoT), and automation technology. Currently, he is working on a project titled "IoT and ML-driven precision farming for optimal water management". He aims to pursue postgraduate studies to further advance his expertise. He can be contacted at email: abra375585@gmail.com.

Irfan A. Chowdhury holds a Bachelor of Science in computer and communication engineering. He plans to pursue a master's degree in machine learning, focusing on artificial intelligence and data science. With a strong aptitude for problem-solving and a passion for leveraging technology to address real-world challenges, he aims to drive advancements in technology and improve lives globally. He can be contacted at email: irfan01cce@gmail.com.

Dr. Mohammed Saifuddin is an Assistant Professor of Chemistry and Environmental Science at the International Islamic University Chittagong, Bangladesh. He obtained his B.Sc. and M.Sc. degree in chemistry from the University of Chittagong, Bangladesh. He received another M.Sc. degree in biotechnology and a Ph.D. in soil & root mechanics from the University of Malaya, Malaysia. Then he joined as a postdoctoral fellow at the University of Malaya. He has over forty research publications in reputed peer-reviewed international journals and presented his research at more than six national and international conferences at home and abroad. In other tasks, he has been a guest reviewer for different international journals. For his research efforts, he was awarded the IIUC Research Achievement Award for ‘outstanding research’ as well. He can be contacted at email: saifuddin@iiuc.ac.bd.