ASSESSING CLIMATIC AND VEGETATIVE INFLUENCES ON SEASONAL RICE YIELDS IN RAJSHAHI: A COMPARATIVE ANALYSIS USING MACHINE LEARNING AND REMOTE SENSING

Chinmoy Debnath Abir*[1], Sadia Afroze[2], Musfique Hossain[3] and Md. Rimon Hossain[4]

[1] *Undergraduate Student, Rajshahi University of Engineering & Technology, Bangladesh, e-mail: chinmoydebnath990@gmail.com*
[2] *Undergraduate Student, Rajshahi University of Engineering & Technology, Bangladesh, e-mail: 2007037.sadia@gmail.com*
[3] *Undergraduate Student, Rajshahi University of Engineering & Technology, Bangladesh, e-mail: 2007030.musfique@gmail.com*
[4] *Undergraduate Student, Rajshahi University of Engineering & Technology, Bangladesh, e-mail: 2007014.rimon@gmail.com*

***Corresponding Author**

ABSTRACT

The climate and vegetation of Bangladesh substantially determine its agricultural yields, particularly as seen in the rain-fed areas like Rajshahi. The understanding of those factors is essential for climate-resilient agricultural planning. This study investigates the performance of different machine learning models, utilizing specific and influential climatic and vegetative remote sensing data, to predict seasonal rice yields in this region. Additionally, it incorporates SHAP interpretability, marking one of the first efforts of its kind in this area. The method applies two machine learning algorithms, XGBoost and Random Forest, trained on daily temperature, rainfall, and NDVI data between 2000 and 2022 under two input scenarios: full-feature models with the use of temperature, rainfall, and NDVI as well as only-NDVI models to emulate conditions of data scarcity. To evaluate the reliability of the models and enhance output interpretability at regional decision-making, LOOCV (Leave-One-Out Cross-Validation) and SHAP (SHapley Additive exPlanations) analysis have been used. Results demonstrate that full-feature models consistently outperformed NDVI-only versions. The Random Forest Models performed best across all three rice seasons, with Aman rice achieving the highest accuracy (RMSE = 0.311, R^2 = 0.300). SHAP analysis provides information that for Boro and Aman yields, temperature and NDVI were most influential, while precipitation was more important for Aus yield prediction. The findings have been set within a context of line plots resembling temperature, rainfall, and NDVI over time, as well as actual vs predicted scatter plots for rice yields, along with bar, beeswarm, and waterfall visualizations for SHAP analysis. This study thereby establishes that models incorporating climate and vegetative parameters generally outperform NDVI-only models in predicting rice yields, with due consideration to seasonal variation in the predictability of yields from different model types. This study also creates a basis for building intervention and adaptation strategies directed toward specific impacts of climate change on rice production through identification of the most relevant factors across seasons. Such results can be taken into the making of Rajshahi's localized climate-smart agricultural policies along with necessary IoT integration and adaptation strategies under bodies such as the Department of Agricultural Extension (DAE) and the Barind Multipurpose Development Authority (BMDA).

Keywords: *Rice yield prediction, Machine learning, Remote sensing, SHAP analysis, Rajshahi*

1. INTRODUCTION

In Bangladesh, agriculture has been one of the main pillars of the economy as it generates almost 12.7 percent of the national GDP and 40 percent of the working population (Bangladesh Bureau of Statistics [BBS], 2023). Rice is one of the main crops and occupies nearly three-quarters of the total crop area and is the main source of calories eaten by the population (Islam et al., 2021). Nonetheless, climate variability is increasingly threatening the productivity of rice cultivation through variation in rainfall levels, rise in temperature, and the changing dynamics of vegetation (Rahman et al., 2022). Such changes are especially intense in Rajshahi, a semi-arid rain-fed farmland in the northwest of Bangladesh, where the lack of water and unreasonable heat have become even stronger over recent decades (Hossain et al., 2020).

Knowledge of how climatic and vegetative conditions affect crop yields has become the CSA strategies pillar (Food and Agriculture Organization [FAO], 2017). Climate-smart agriculture focuses on climate, agricultural productivity, and sustainability through the unification of climate and vegetation data in agricultural planning. Nonlinear and multivariate interactions of climatic factors and crop performance are not however always well reflected using traditional methods of statistics (Chen et al., 2023). It is now possible to extract meaningful patterns in complex datasets, which advances remote sensing and machine learning (ML) to better forecast yields and make environmental decisions (Yadav et al., 2022).

Random Forest (RF) and Extreme Gradient Boosting (XGBoost) are some of the most popular machine learning models that are highly predictive and can be trained on heterogeneous data to predict crop yields (Kamble et al., 2021; Sun et al., 2020). These models are better than traditional regression-based models in that they can model nonlinear relationships between temperature, rainfall and vegetation indices. Moreover, explainable AI models like the SHapley Additive explanations (SHAP) have also added another layer of explainability to the field of agricultural modelling providing a numerical measure of the contribution of each feature to model outputs (Lundberg and Lee, 2017). Despite these developments, few studies had been done on seasonal comparative modelling of rice yields in mountain areas such as Rajshahi that have high climatic extremes between Boro, Aus, and Aman seasons.

Vegetation indices, especially the Normalized Difference Vegetation Index (NDVI), can be used to track the health of crops and predict their yield with the help of remote sensing (Zhang et al., 2021). But NDVI by itself can be a good way to filter minimal information, which is not related to climatic variables. A combination of NDVI and other climate-related parameters (temperature, rainfall, etc.) can help gain a more in-depth insight into the impact of environmental stressors on crop production (Dutta et al., 2022). However, the joint use of machine learning, NDVI, and multi-seasonal rice yield prediction by using SHAP-based interpretability in Bangladesh has not been thoroughly studied.

Considering this gap, the current study will compare the joint effect of climatic and vegetative variables on seasonal rice yields in Rajshahi District using XGBoost and Random Forest. The research additionally makes the comparison of full-feature (Climate with NDVI) models to the NDVI-only models to simulate the performance under data-scarcity conditions. As well, SHAP analysis is used to give interpretable information on the feature contributions of each rice season. In this way, the study attempts to add to the creation of region-sensitive, climate-intelligent agricultural policies, which are aligned with the local programs, including the Department of Agricultural Extension (DAE) and the Barind Multipurpose Development Authority (BMDA).

2. METHODOLOGY

2.1 Study Area and Context

The research was done on Rajshahi District, located in the northwest region of Bangladesh (24.37°N, 88.60°E). The climate of the area is tropical monsoon with separate Boro (January-May), Aus (March-July), and Aman (July-November) rice seasons. The climate is mainly rain-fed with little irrigation cover in the non-dry Boro period, thus making Rajshahi a handy research subject in terms of

studying the relation of climate and vegetation to yield. The groundwater-based irrigation systems are also a problem affecting the area due to the activities of the Barind Multipurpose Development Authority (BMDA).

2.2 Data Collection and Processing

Both climate datasets and remote sensing datasets have been used in the study, which covers the years 2000-2022. Climatic data on daily temperature (°C) and rainfall (mm) were collected at the Bangladesh Meteorological Department (BMD) and compared with the NASA POWER database to assure the consistency in time. For vegetation data, MODIS MOD13Q1 (250 m, 16-day composite) satellite data was used to obtain the normalized difference vegetation index (NDVI). The values of the NDVI were reduced to mean values monthly and averaged spatially across the borders of Rajshahi District through the Google Earth Engine (GEE). Rice yield data on the seasonal average rice yield (t/ha) in Boro, Aus, and Aman were taken after the Bangladesh Bureau of Statistics (BBS). These data were compared to similar climatic and NDVI data of the years. The datasets were then combined into one large time-series panel where the row of data reflected the mean seasonal statistics of temperature, rainfall, NDVI, and yield of a particular year. Interquartile range (IQR) filtering was used to identify outliers and correct them when the situation arose by use of linear interpolation.

2.3 Feature Selection and Seasonal Segregation

The model inputs were taken to be variables, which included temperature (°C) - mean temperature (daily) seasonally; Rainfall (mm) - cumulative precipitation (daily) of the season; NDVI (unitless) means vegetative health index. The data were seasonally separated according to the months of standard growing Boro: January-May, Aus: March-July, Aman: July-November. A dataset was considered as a season to ensure that there were seasonal climatic variability and variations in crop management practices.

2.4 Model Development

Rice yield was predicted using two ensemble learning algorithms to predict based on the chosen features: Extreme Gradient Boosting (XGBoost) - it is a decision tree-based boosting algorithm that can address more complicated nonlinear relationships and interactions between features and Random Forest (RF) - an ensemble model that relies on bagging to decrease overfitting and offers interpretable importance of features. Both models were coded in Python (v3.12.7) in scikit-learn and xgboost libraries. The normalization of inputs was done through a z-score that is used to assure similar scales among predictors. There were two modeling scenarios that were tested: Full-Feature Model: NDVI, temperature, and rainfall. NDVI-Only Model: NDVI alone as the predictor, which is ideal in environments with limited data in most developing regions.

2.5 Model Validation

To obtain reliable performance evaluation, Leave-One-Out Cross-Validation (LOOCV) was used. In this method, data of a single year were sequentially eliminated in the test, and the rest of the years were utilized in training. This approach has less overfitting and can produce stable performance values even when working with small datasets. Accuracy of the models was determined by: Root Mean Square Error (RMSE) - quantifies the size of errors in prediction. Coefficient of Determination (R2) - measures goodness-of-fit of observed and predicted yields.

2.6 Interpretation and Explainability of Models

SHAP (SHapley Additive exPlanations) was applied to calculate the contribution of each input feature to the predicted yields of rice to interpret the model predictions. Three SHAP graphs were created: Bar plots of the importance of features in the globe, Plots of beeswarm interaction Beeswarm feature interaction. Waterfall diagrams describing local (year-to-year) forecasts. These explainable methods

of AI assisted in determining the types of environmental variables that had the strongest impact on the yield in the various rice seasons.

2.7 Climatic and Vegetative Trend Analysis and Comparative Model Performance

To examine the long-term climatic and vegetative trends in Rajshahi and compare the predictive performance of machine learning models under different input configurations, seasonal trends in temperature, rainfall, and NDVI over the period 2000–2022 were analyzed to contextualize interannual variability in rice yields. In parallel, comparative model performance is evaluated using RMSE, R^2, and SHAP-based interpretations to assess the added value of integrating climatic variables alongside NDVI. This combined analysis provides insight into how environmental trends influence seasonal yield predictability and model behavior.

2.8 Policy Coherence and Regionalism

Lastly, findings were also correlated with the presence of climate-smart agricultural programs in Rajshahi, such as the Department of Agricultural Extension (DAE), the BMDA, and the Smart Agriculture Programs. The model outcomes are interpretable to aid in data-driven decision-making to optimize resources, irrigation scheduling, and adaptive management in relation to different climatic changes.

3. RESULTS

3.1 Model Performance Comparison Across the Seasons

The machine learning models were implemented in two configurations namely: (1) Full-feature models which combine temperature, rainfall, and NDVI; and (2) NDVI-only models which simulate the data-scarce conditions. A Leave-One-Out Cross-Validation (LOOCV) was used to perform a strong assessment of the small real annual samples (2000-2022). Table 1 summarizes the model performance measures.

Table 1: Performance comparison of XGBoost and Random Forest models for seasonal rice yield prediction in Rajshahi District using full-feature (temperature, rainfall, NDVI) and NDVI-only inputs under Leave-One-Out Cross-Validation (2000–2022).

Season	Model	Input Features	RMSE	R^2
Boro	XGBoost	Full (Temp, Rain, NDVI)	0.401	-0.014
Boro	Random Forest	Full (Temp, Rain, NDVI)	0.359	-0.008
Boro	XGBoost	NDVI Only	0.474	-0.069
Boro	Random Forest	NDVI Only	0.456	-0.056
Aus	XGBoost	Full (Temp, Rain, NDVI)	0.569	-0.085
Aus	Random Forest	Full (Temp, Rain, NDVI)	0.540	-0.073
Aus	XGBoost	NDVI Only	0.627	-0.106
Aus	Random Forest	NDVI Only	0.612	-0.099
Aman	XGBoost	Full (Temp, Rain, NDVI)	0.332	0.238
Aman	**Random Forest**	**Full (Temp, Rain, NDVI)**	**0.311**	**0.300**
Aman	XGBoost	NDVI Only	0.377	0.201
Aman	Random Forest	NDVI Only	0.355	0.243

Table 1 shows the performance variables of the XGBoost and random forest models using the full-feature (temperature, rainfall, and NDVI) and the NDVI-only input variable in the seasons of Boro, Aus and Aman rice. The findings show that full-feature models always had lower RMSE and higher values of R^2 than those of NDVI-only models and demonstrate the high importance of climatic variables in enhancing prediction accuracy. The best performance of the models in all the seasons was

observed in the Random Forest with full-feature inputs, especially Aman season (RMSE = 0.311, R^2 = 0.300). The Aus season had a relatively lower predictability as it had large variability of rainfall in the pre-monsoon season.

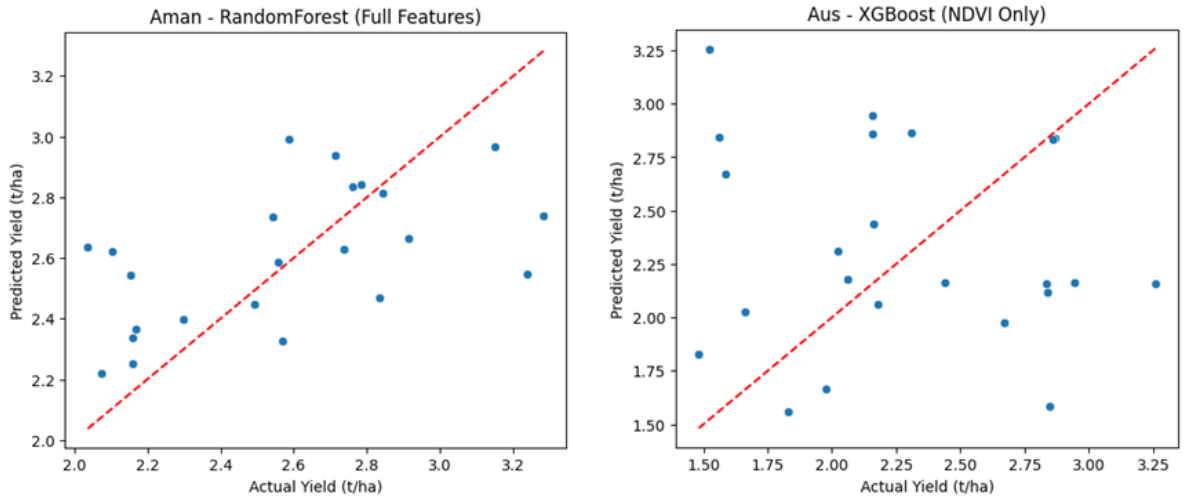


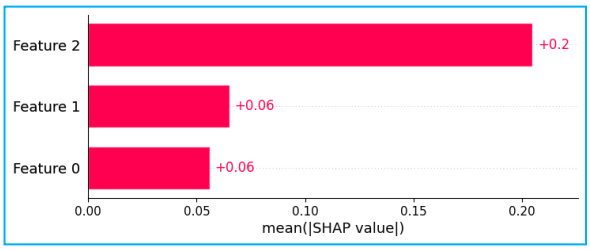
Figure 1: Actual versus predicted seasonal rice yields for the best-performing (Aman–Random Forest, full-feature) and lowest-performing (Aus–XGBoost, NDVI-only) models under Leave-One-Out Cross-Validation, illustrating differences in predictive reliability across seasons.

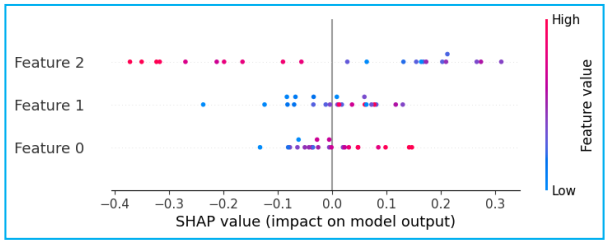
The representative scatter plots (Figure 1) represent the correlation between actual and predicted yields of the best and the worst cases (Aman-RF Full vs. Aus-XGBoost NDVI-only). The points that are relatively near the 1:1 line mean that the model prediction can be regarded as reliable, whereas the higher the discrepancy in the NDVI-only plots, the more it is possible to stress that vegetation indices are not so effective in predicting the climatic factors in the absence of their input.

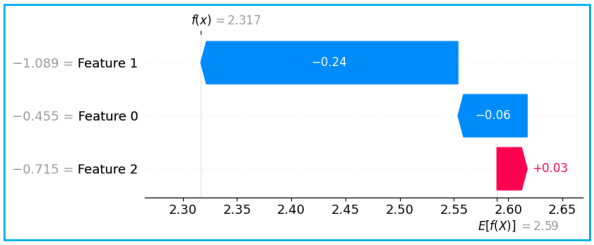
These findings indicate that incorporation of temperature and rainfall on top of NDVI data support more of the inter-annual variation in the yields produced by the models. The differences in the seasonal performances of models also suggest that the influence of climatic parameters on yield prediction is different in different levels based on the growth stage of the crop and exposure to weather. To explore these relationships in more detail, Section 3.2 introduces SHAP-based analyses, which break down the predictions of the models and show the relative significance and direction of the various features between the rice seasons.

3.2 SHAP-Based Interpretation

To analyze the influence of individual features (Feature 0 = Temperature, Feature 1 = Rainfall, Feature 2 = NDVI) on the yield predictions, SHAP (SHapley Additive exPlanations) analysis was conducted on both the XGBoost and the Random Forest models when the full-feature and the NDVI-only was considered. This method of interpretation will quantify the direction and strength of the influence of each variable on the yield of rice, which will make the process of decision-making of the models transparent. Figure 2, summarizes the feature importance using bar plot, beeswarm plot and waterfall plot on the best performing model (Random Forest - Aman - Full







Fratures).

Figure 2: SHAP bar, beeswarm, and waterfall plots for the Aman season using the Random Forest full-feature model, showing global feature importance, feature impact distribution, and local prediction contributions for a representative year.

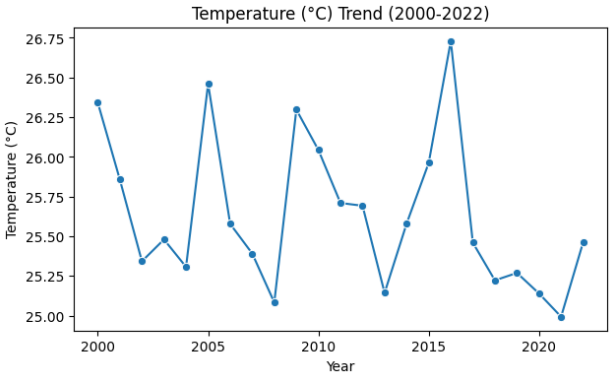
In all three seasons, SHAP bar and beeswarm plots showed that temperature and NDVI were always the most prominent predictors, whereas the effect of rainfall was more fluctuating. The SHAP values during the Boro season showed that increased temperatures and high vegetative vigor (high NDVI) tended to make a positive contribution to the yield projections, which was expected considering the reliance of the crop on the existence of stable winter conditions and sufficient irrigation. Conversely, the Aus season was more sensitive to rainfall changes with higher predicted yields generally being associated with heavy rainfall and therefore it can be implied that waterlogging or leaching of nutrients during early developmental stages.

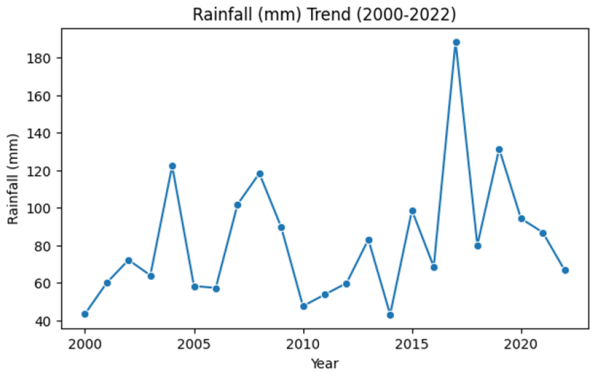
Temperature and NDVI also proved to be significant factors to the Aman season, however, the proportion of NDVI was especially large, which suggests that the late-season vegetation health may be a determining factor with regards to harvest. SHAP waterfall plots also supported that all conditions of moderate temperature and dense canopy had the tendency of encouraging higher yield forecasts in this season.

In the comparison between full-feature and NDVI-only models, the SHAP analyses focused on the idea that NDVI in isolation is able to represent some patterns of vegetation, but did not manage to explain climatic stressors that significantly influence inter-annual variability. Having temperature and rainfall in the full-feature models resulted in a more balanced and context-sensitive prediction model, which accounts for the improvement in performance in Table 1.

3.3 Climatic and Vegetative Trends Over the Years

The climatic and vegetative long-term trends (2000-2022) were used to provide the context of the model results. Figure 3 shows the climatic and vegetative trend of Rajshahi between 2000 and 2022 in the long term. The trend of the temperature shows a slow yet steady increase in all three seasons of rice with highest increase recorded in Boro and Aman. This warming trend is an indication of rising evapotranspiration and possible crop stress especially where the reproductive stage is sensitive to heat. There is a significant variation in the year-to-year movement of the rainfall pattern





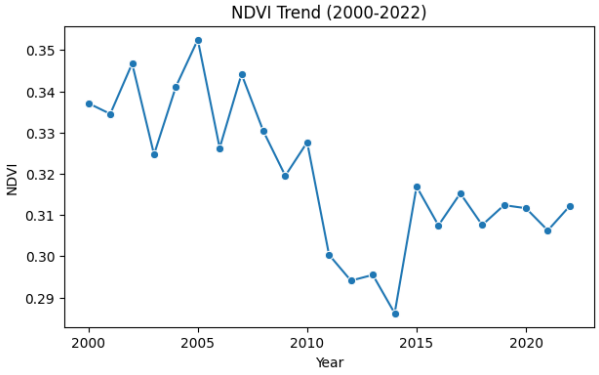


Figure 3: Long-term annual trends of mean temperature (°C), total rainfall (mm), and NDVI in Rajshahi District during 2000–2022, showing interannual variability and gradual warming that provide environmental context for the rice yield prediction models.

with the Aus season recording the greatest change. Such anomalous rains can also affect the uniformity of soil moisture and stability of the Aus yields. Conversely, irrigation and late-monsoon rainfall respectively support Boro and Aman to receive relatively steady levels of precipitation.
The trend of NDVI shows a slight positive growth which means a good increase in vegetative health. This may be due to enhanced irrigation control, increased adoption of high yielding forms and development of infrastructural projects through projects such as the Barind Multipurpose Development Authority (BMDA). It is important to note that the highest level of NDVI occurs in the Aman months, which coincides with peak rainfall and moderate temperatures, which indicate good soils.
Collectively, the trends highlight the changing climate-vegetation interactions upon which seasonal fluctuations in rice yields in Rajshahi are based. Such long-term environmental understanding integrated enhances the interpretive richness of model predictions and underscores the value of local and season specific adaptation planning.

3.4 Policy Relevance and Implications

The discussion reports the importance of combining both climatic and vegetative cues in rice production prediction within semi-arid agroecosystems such as Rajshahi. Since the local rainfall is becoming increasingly unpredictable and the temperature is gradually increasing, the adaptive measures of the IoT-based monitoring of the local microclimate, the real-time NDVI tracking, and the climate-sensitive irrigation schedule within the framework of Barind Multipurpose Development Authority (BMDA) might increase the resilience of yields. Likewise, the Department of Agricultural Extension (DAE) can use the insights into the importance of features as identified by SHAP to design various kinds of season-specific advisory services and advance the idea of precision farming.

4. CONCLUSIONS

This paper has investigated how the climatic and vegetative variables affect seasonal rice yield in the Rajshahi district of Bangladesh with XGBoost and Random Forest models. The study compared the predictive performance of full-feature (climate + NDVI) models to NDVI-only models by integrating temperature, rainfall and NDVI data between the years 2000 and 2022 to review the applicability of the models in both circumstances (equivalent data-rich and data-scarcity). It was evident in the result that the use of climatic parameters and NDVI greatly improves the accuracy of prediction of yield. Random Forest full-feature model was the most successful in all the seasons of rice and the highest accuracy was found in the Aman season (RMSE = 0.311, R2 = 0.300). Conversely, models using NDVI alone showed relatively low predictive ability which explains the need to combine various climatic variables to derive more accurate yield estimates. The SHAP interpretability analysis was also useful in giving insights to the contribution of every feature to the model predictions. Boro and Aman yield were identified to be most sensitive to temperature, NDVI, whereas Aus yield variability was most sensitive to rainfall. The results are in line with the agro-climatic attributes of Rajshahi where seasonal patterns of rainfall distribution and changes in temperature are important factors to the crop dynamic growth. In addition to model performance, the results highlight the promise of explainable machine learning to make data-driven decisions in agriculture. Transparent assessment of the model results is made possible with the integration of SHAP analysis, which gives the opportunity to policymakers and researchers to understand the sensitivities of the model to each season and implement priority to adaptation approaches. Overall, the research confirms that the predictability and understandability of rice yield is boosted by integrating climatic and vegetative variables. The results provide a foundation for future development of operational yield forecasting systems, subject to further validation and higher-resolution data. Despite the encouraging results, several limitations should be acknowledged. The yield predictions are based on annual, district-averaged data and do not

capture intra-seasonal variability or field-level heterogeneity. Additionally, the models were evaluated using historical datasets and have not yet been validated for real-time or operational forecasting. Therefore, the findings should be interpreted as indicative rather than deployable predictions, highlighting patterns and sensitivities rather than serving as direct decision-support tools. Future studies can build on this framework to include the IoT-based real-time sensing and spatially resolved data that can be used to enhance predictive accuracy and facilitate adaptive management under the shifting climate regimes.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the Department of Urban and Regional Planning, Rajshahi University of Engineering & Technology (RUET), for providing technical guidance and research facilities. Special appreciation is also extended to the Bangladesh Meteorological Department (BMD) for access to essential climatic datasets that made this research possible.

DECLARATION OF USE OF AI

The authors declare that artificial intelligence tools were used only for language correcting and clarity improvement during manuscript preparation. The research methodology, data analysis, modeling, interpretation, and conclusions were conducted entirely by the authors without the use of AI tools. The authors take full responsibility for the content of this work.

REFERENCES

Bangladesh Bureau of Statistics (BBS). (2023). *Yearbook of Agricultural Statistics 2023*. Dhaka: Government of the People's Republic of Bangladesh.

Chen, Y., Zhao, J., & Liu, F. (2023). Integrating climate variables and satellite data for crop yield prediction using machine learning. *Computers and Electronics in Agriculture*, 212, 108054.

Dutta, S., Alam, M., & Saha, S. (2022). Climate variability and rice yield prediction using multi-source data fusion in South Asia. *Environmental Modelling & Software*, 156, 105469.

Food and Agriculture Organization (FAO). (2017). *Climate-Smart Agriculture Sourcebook (2nd ed.)*. Rome: FAO.

Hossain, M. A., Rahman, M. M., & Sarker, M. A. (2020). Climatic trends and agricultural vulnerability in northwestern Bangladesh. *Theoretical and Applied Climatology*, 142(1–2), 105–119.

Islam, M. N., Hasan, M. A., & Rahman, M. S. (2021). Rice production systems and challenges of climate change in Bangladesh. *Agricultural Systems*, 187, 103027.

Kamble, A., Khedkar, S., & Bhosale, R. (2021). Evaluation of machine learning algorithms for rice yield prediction using satellite data. *Journal of Applied Remote Sensing*, 15(3), 034521.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

Rahman, M. S., Hossain, M. A., & Karim, M. N. (2022). Climate change impacts on crop yield and adaptive management strategies in Bangladesh. *Sustainability*, 14(9), 5567.

Sun, L., Di, L., & Fang, H. (2020). Using Random Forest and XGBoost to predict wheat yield based on remote sensing data. *Remote Sensing*, 12(9), 1456.

Yadav, V., Ghosh, S., & Sharma, R. (2022). Machine learning-based crop yield forecasting under climate change scenarios. *Agricultural Water Management*, 266, 107591.

Zhang, Y., Jiang, L., & Chen, J. (2021). Remote sensing-based vegetation indices for rice yield estimation: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178, 178–196.