Discover Applied Sciences

Research

Advancing food sustainability: a case study on improving rice yield prediction in Sri Lanka using weather-based, feature-engineered machine learning models

Aminda Amarasinghe[1] · Ishini Sangarasekara[1] · Nuwan De Silva[2] · Mojith Ariyaratne[2] · Ruwanga Amarasinghe[3] · Jinendra Bogahawatte[4] · Janaka Alawatugoda[5,6] · Damayanthi Herath[1]

Received: 30 July 2024 / Accepted: 18 October 2024
Published online: 11 November 2024

Abstract

Food sustainability is crucial aspect in achieving several United Nations (UN) Sustainable Development Goals (SDGs). By integrating advanced technologies for reliable and accurate decision-making, we can advance food sustainability and, consequently, make significant advances toward achieving the UN SDGs. Rice, a staple crop in many Asian and some African nations, is crucial to Sri Lanka as well. Serving as the primary food for most Sri Lankans, it plays a vital role in sustaining the livelihoods of over 1.8 million farmers. In Sri Lanka, rice is grown during two distinct seasons of the year (Yala and Maha). This study focuses on ML with feature engineering for rice yield prediction using weather data: Rainfall, Maximum temperature, Minimum temperature, and Radiation. The data from two districts in Yala and Maha seasons collected from 1982 to 2019 were used for evaluating two sets of models respectively. Data were pre-processed to handle the outliers and missing values and scaled using normalization. The machine learning models considered are Linear Regression (LR), Support Vector Machine (SVM), *k*-Nearest Neighbour (KNN), and Random Forest (RF). The performance of these models was evaluated using metrics: Root Mean Squared Error (RMSE), Relative Root Mean Squared Error (RRMSE), and Mean Absolute Error (MAE). The results demonstrate that Random Forest Regression with less number of features can yield comparable results compared to the original set of features.

Article Highlights

- Rice is a staple food and vital to the livelihoods of millions of people worldwide. Therefore, accurate and timely prediction of rice yield is essential for global food security.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42452-024-06300-7.

✉ Janaka Alawatugoda, jalawatugoda@ra.ac.ae; ✉ Damayanthi Herath, damayanthiherath@eng.pdn.ac.lk | [1]Department of Computer Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya 20400, Central Province, Sri Lanka. [2]Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya 20400, Central Province, Sri Lanka. [3]School of Technology, Sri Lanka Technological Campus, Padukka 10500, Western Province, Sri Lanka. [4]Department of Computing, Sri Lanka Institute of Information Technology, Malabe 10115, Western Province, Sri Lanka. [5]Research & Innovation Centers Division, Rabdan Academy, P.O. Box 114646 Abu Dhabi, UAE. [6]Institute for Integrated and Intelligent Systems, Griffith University, Nathan, QLD 4111, Australia.

- The study integrates machine learning techniques and feature engineering on weather data (including rainfall and temperature) to improve rice yield predictions, thus contributing to food sustainability and progress toward the UN Sustainable Development Goals (SDGs).
- Among the machine learning models evaluated (Linear Regression, Support Vector Machine, k-Nearest Neighbour, and Random Forest), Random Forest Regression demonstrated that fewer features could produce results comparable to those using a full set of features, highlighting its efficiency in rice yield prediction.

Keywords Food sustainability · Machine-learning · Feature engineering · Rice-yield prediction · Weather data

1 Introduction

Achieving the United Nations (UN) Sustainable Development Goals (SDGs) necessitates a global commitment to food sustainability [1]. It refers to ensuring present generations have access to adequate nutrition without compromising future generations' ability to meet their own needs. It aligns with several SDGs, including "Zero Hunger" (SDG 2), "Responsible Consumption and Production" (SDG 12), and "Climate Action" (SDG 13) [1]. Sustainable food systems promote environmentally friendly agricultural practices, reduce food waste, and guarantee equitable access to nutritious food [2].

Prioritizing sustainable food practices fosters a healthier planet, fuels inclusive economic growth, and advances progress towards the UN's vision for a more sustainable future [1, 2]. These goals can be achieved by supporting farmers, adopting agro-ecological methods, and integrating technological solutions for reliable decision-making.

Rice is a staple food in many Asian and African nations, with significant consumption in South and North America as well [3]. However, for Sri Lanka, rice holds a dual significance. It serves as the primary source of nutrition for most Sri Lankans, nourishing families as well as fueling the economy. Despite its importance, Sri Lanka's average rice yield remains below 4.3 tons per hectare (t/ha) and faces a potential % reduction due to climate change's impact on temperature and rainfall patterns [4]. This highlights the need to improve rice productivity while adapting to a changing climate to ensure food security and safeguard the agricultural sector's economic contribution [4].

Sri Lanka's climate can be broadly categorized into three zones based on rainfall patterns: the Wet Zone, Dry Zone, and Intermediate Zone [5]. The Wet Zone receives the most rain, exceeding 2,500 mm annually, while the Dry Zone receives less than 1,750 mm [5]. The Intermediate Zone experiences rainfall between these two extremes [5]. Two monsoons, the Southwest Monsoon (SWM) and Northeast Monsoon (NEM), and conventional rain patterns influence Sri Lanka's agricultural seasons [5]. Additionally, two inter-monsoon periods known as the First Inter Monsoon (FIM) and Second Inter Monsoon (SIM) contribute to rainfall [5].

These contrasting weather patterns result in four distinct rainfall seasons and two main rice cultivation seasons in Sri Lanka: Yala and Maha [5]. The Yala season encompasses the FIM and SWM periods, while the Maha season begins with SIM rains and continues until the NEM concludes [5]. The Yala season, typically lasting from March to September, is considered a minor growing season [6]. In contrast, the Maha season, spanning October to February of the following year, is the major growing season [6]. In 2022, the extent of rice cultivation during Maha and Yala seasons was 766,148 hectares (ha) and 481,289 ha, respectively, resulting in a total rice production of 3,392,905 tons [6]. However, the average yield during these seasons is still relatively low, with Maha and Yala seasons averaging around 2,853 kg/ha and 3,207 kg/ha, respectively [6].

1.1 Necessity for accurate rice yield prediction

Uncertainties in weather, production levels, policies, and prices pose significant challenges for the agricultural sector [7]. A sudden decline in rice production reduces marketable surplus, leading to financial losses for farmers and a price increase for consumers [7]. Conversely, a significant production surplus can cause a sharp decrease in rice prices, negatively impacting farmers' income [7]. Given rice's role as an essential commodity, its price fluctuations significantly impact inflation rates, wages, salaries, and various economic policies [8].

Therefore, accurate rice yield prediction is crucial for mitigating these challenges and ensuring Sri Lanka's food security. A reliable forecast allows for better surplus and deficit management. Knowing the anticipated yield enables informed decisions regarding imports and exports, stabilizing prices and guaranteeing fairer profits for farmers [9, 8, 2]. Additionally, accurate predictions empower policymakers to develop more effective strategies addressing potential shortfalls and ensuring a consistent supply of rice for the nation [5]. In Sri Lanka, Rice yield data are originally collected by the Department of Agriculture for each season from their crop cut survey. During the crop cut survey a team from the Department of Agriculture obtains representative samples from several locations (systematically) from farmer fields of each district and the average yield for each district is calculated. This is a very tedious and time and labor intensive activity demanding more robust, accurate and adaptoble approaches for rice yield prediction.

1.2 Harnessing machine learning for rice yield prediction

Big data and machine learning (ML) are transforming the rice sector, enabling data-driven crop monitoring, yield prediction, and disease detection [7]. Researchers are exploring various ML approaches, including hybrid models that combine techniques for improved accuracy [10]. Combining spatial and temporal data is also crucial for effective yield forecasting [11]. ML models depend on reliable data, and technologies like blockchain offer solutions for ensuring data integrity [12].

Furthermore, ML applications in the rice sector extend beyond yield prediction. Sensors and ML can determine moisture content (key for storage) and even predict disease outbreaks like rice blast, allowing for timely interventions [13, 14]. Deep learning, specifically with transfer learning, offers accurate tools for rice disease classification [15, 16]. ML combined with crop models and hyperspectral data enhances precision agriculture practices for rice [16, 17].

ML has been considered in addressing food safety concerns like arsenic contamination in rice [18].Studies have compared ML methods, underscoring how model choice depends on the complexity of the problem [19]. ML models like Random Forests and SVM, when used with time-series data, demonstrate effective yield prediction capabilities [20]. These methods have been shown to empower precision agriculture practices, such as the accurate detection of rice growth stages from UAV images [21]. In the context of rice yield prediction in Sri Lanka multiple ML methods using climatic data have been studied [2, 22]; However, to the best of our knowledge feature engineering selection coupled with machine learning for rice yield prediction with minimal number of features has not been studied before.

1.3 Climate change, sustainability, and the need for adaptation

ML enables more informed decision-making about crop selection and yield optimization, considering diverse environmental factors [23]. However, socioeconomic factors are equally crucial for translating these advancements into real-world improvements for farmers [24]. Sustainable practices, like cover crop integration, benefit soil health and crop productivity [25]. Rice yields are particularly sensitive to climate change, with studies demonstrating the negative impact of temperature and rainfall variation [26, 27, 22]. Understanding these climate-driven trends is vital for developing reliable ML-based yield prediction models and adaptation strategies [28]. Selecting the optimal model can be complex, and comparative studies offer valuable insights for specific contexts [29].

While significant advancements have been made in applying machine learning for rice yield prediction and understanding climate change's impact, these models often lack regional specificity. This is particularly critical for Sri Lanka, with its unique climate and agricultural practices. To address this gap and empower Sri Lankan farmers, this study explores the development of regionally-specific machine learning models for rice yield prediction.

1.4 Our contribution

By incorporating historical weather records and corresponding rice yield data, the study aims to develop and evaluate predictive models capable of capturing the intricate relationships between weather variables and rice yield.

The outcomes of this research are expected to contribute to the development of data-driven decision-support systems for farmers and researchers. By accurately predicting rice yields based on a minimum number of weather data, it can assist farmers in making decisions on their cultivation.

While our study is based on Sri Lankan data, its applicability extends to any geographical region and any commercial crop, provided that dependable weather data is accessible. Consequently, this model building process emerges as a valuable tool not only in fostering food sustainability, but also in making substantial progress towards achieving the United

Nations Sustainable Development Goals (UN SDGs).The source code of the experiments of this research can be accessed through https://github.com/AmindaUdayanga/Feature-Engineered-Rice-yield-Prediction-based-on-Weather-Data.

1.5 Organization of the paper

This paper begins with an Introduction that establishes the importance of rice cultivation, the challenges posed by climate change, and the potential of machine learning for yield prediction. A literature review within the Introduction examines existing applications of machine learning in this domain. The Materials and Methods section describes the dataset, data preprocessing, feature selection, model development, and evaluation metrics. The Results present the performance of different models and highlight key variables influencing yield. The Discussion analyzes the findings in the context of existing research, addresses limitations, and outlines practical implications and future directions. The Conclusion summarizes the study's outcomes and their significance for food security and climate change adaptation.

2 Materials and methods

In this section, we discuss about the materials and methods that have been used in this research. Particularly, we discuss in detail about the data, data analysis method, data pre-processing methods, feature selection methods, model training and evaluation that have been used in this research.

2.1 Data

Daily weather data which includes radiation, maximum and minimum temperature, and rainfall of two geographical areas (Kurunegala and Anuradhapura) were considered. Then yearly average climatic data for the seasons 'Yala' (May-August) and 'Maha' (September-March) were calculated using daily records. Total Rice Yield amount of two seasons (Yala and Maha) have been collected for both districts Kurunegala and Anuradhapura from the Department of Census and Statistics. These climatic data and rice yield data have been categorized into four distinct groups considering two rice cultivation seasons: Anuradhapura Maha, Anuradhapura Yala, Kurunegala Maha, Kurunegala Yala and created four distinct sub-datasets. Daily weather data (from 1980 to 2019) which includes solar radiation (mj/m2), maximum and minimum temperatures (°C) and, rainfall (mm) of Kurunegala and Anuradhapura were obtained from natural Resource Management Center (NRMC), Department of Agriculture, Sri Lanka. Weather data have been collected by NRMC from their weather stations situated at Mahailluppallama (Anuradhapura) and Bathlagoda (Kurunegala). Rice yield data (kg/ha) were for both areas for both seasons were collected from the Department of Cencus and Statistics Sri Lanka which is freely available at their website. Rice yield data available at Department of Census and statistics are originally collected by the Department of Agriculture each season from their crop cut survey (as mentioned in sect. 1.1).

A separate feature-engineered dataset was developed using temperature (considering the cardinal temperatures of rice) and rainfall data and the mentioned set of machine-learning models evaluated on the accuracy in rice yield

Table 1 Column names and meanings of feature-engineered datasets

Column name	Meaning
X1	Total rainfall (mm)
X2	Average of daily maximum temperatures between 29 °C and 33 °C
X3	Average of daily maximum temperatures $\geq$ 33 °C
X4	Average of daily maximum temperatures $\leq$ 29 °C
X5	Number of days with maximum temperatures between 29 °C and 33 °C
X6	Number of days with maximum temperatures $\geq$ 33 °C
X7	Number of days with maximum temperatures $\leq$ 29 °C
X8	Number of days with maximum temperatures $\leq$ 24 °C
X9	Number of days with maximum temperatures $>$ 24 °C
X10	Average of daily minimum temperatures $>$ 24 °C
X11	Average of daily minimum temperatures $\leq$ 24 °C

prediction for each group mentioned. Some columns of feature-engineered data sets have been named in the following short forms as in Table 1. In the rest of the article, those short forms are used.

Weather data until 2019 were used for analysis, as more recent data were not available at the time of the analysis. This can be identified as a limitation of our research.

2.2 Data analysis

An exploratory data analysis was conducted on the data to explore the relationship between climatic factors and rice yield and their correlations. Statistical analysis of regions for Maha and Yala Seasons is shown in Table 2. The highest correlation was noticed between the rice yield and the year. The amount of rice yield produced each year has increased over the years (Fig. 1). That may be due to various reasons like technological advancements, knowledge and expertise improvement, and the development of infrastructure.

The maximum temperature and average solar radiation also have a high correlation with each other. When solar radiation increases, it leads to higher temperatures as more energy is absorbed by the Earth's surface and it may be the reason for high correlation.

The average minimum and maximum temperature during the Yala season in both Anuradhapura and Kurunegala districs is higher than them in the Maha season. Yala season coincides with the period when the sun is directly overhead in Sri Lanka, resulting in higher solar radiation. The increased solar radiation leads to higher temperatures during the day in the Yala season compared to the Maha season. In the Maha season, the sun's angle is lower, resulting in less intense solar radiation and relatively lower temperatures. Figure 2 shows the variation of average maximum and minimum temperatures with the year.

2.3 Data pre-processing

In this study, data pre-processing was performed using Python, specifically with the Pandas and NumPy libraries. Pandas was utilized for data manipulation and analysis, and NumPy was used for numerical operations.

Under data pre-processing, one of the initial steps was the removal of rows that contain zero values for certain attributes. Then outliers were handled using the Interquartile Range (IQR) method. It is a statistical measure that provides information about the spread or dispersion of a dataset. It is calculated by taking the difference between the third quartile (Q3) and the first quartile (Q1) of the data(Eq. 1).

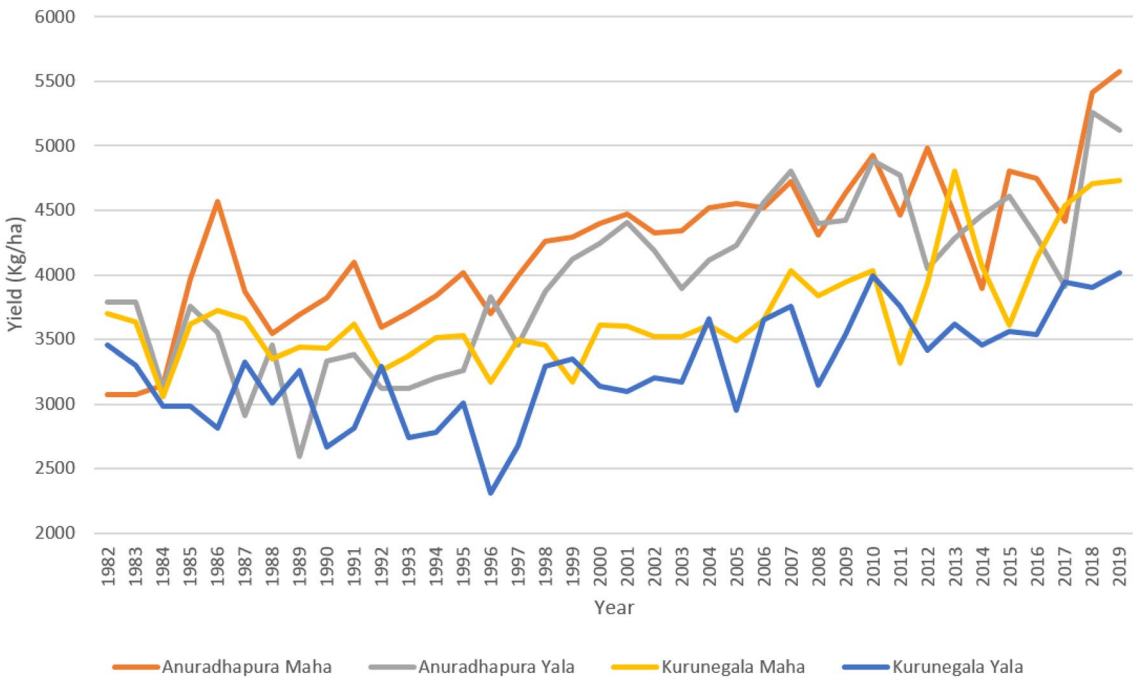

Fig. 1 Variation of rice yield with year

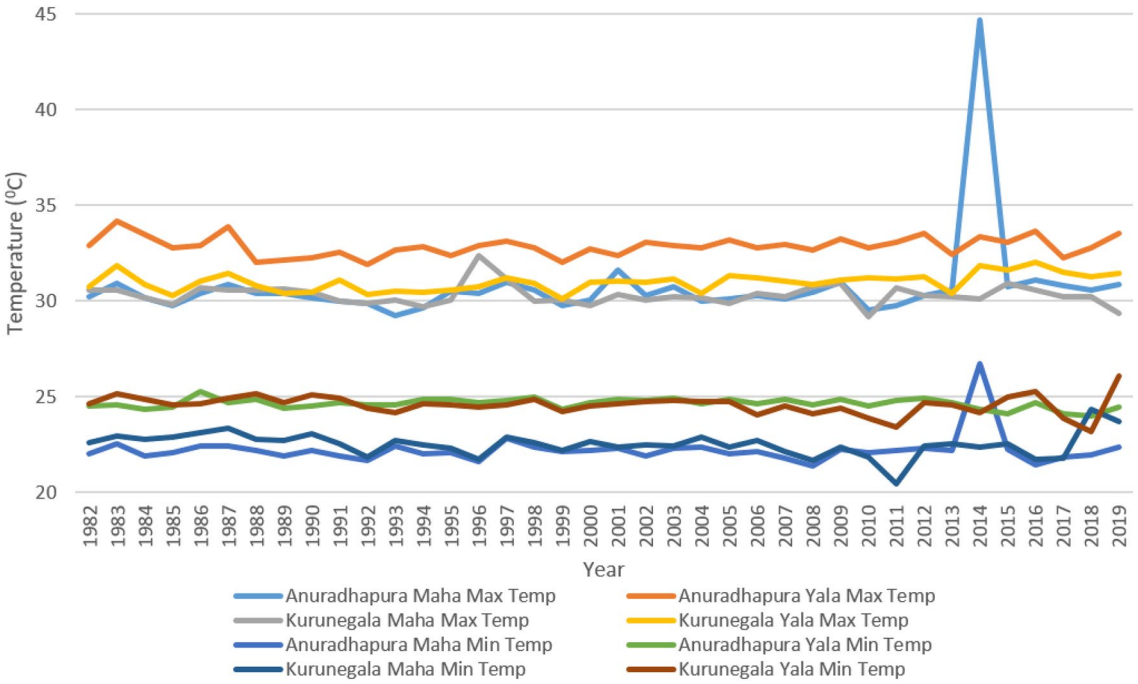

Fig. 2 Variation of average maximum and minimum temperatures with the year

$$\text{IQR} = Q_3 - Q_1 \tag{1}$$

$$\text{Outlier below } Q_1 = Q_1 - (\text{threshold} \times \text{IQR}) \tag{2}$$

$$\text{Outlier above } Q_3 = Q_3 + (\text{threshold} \times \text{IQR}) \tag{3}$$

Equations (2) and (3) show how outliers below Q_1 and above Q_3 are determined using the IQR method. These equations help in identifying data points that are significantly different from the rest of the dataset.

Lastly, datasets were split into train and test datasets by taking 70% of the total entries into the training dataset and the other 30% into the test dataset.

2.4 Feature selection

Feature selection was done considering feature importance and the best features were identified for each dataset separately using RandomForestRegressor. Then from 1 up to the total number of features, the most important features were mapped considering the cumulative importance i.e. when the number of features is equal to *k*, the most important *k* features were considered during the feature selection. Next, Random forest models were developed by changing the number of features considered and finally, the features were selected by considering the lowest RMSE value as in Figs. 3 and 4. Table 3 summarizes the selected features for each district and each season, and Table 4 summarizes the selected features for initial datasets. Further, Residual plots were generated and the distribution of residuals was observed to evaluate the assumption of normality. In the analysis of results, Diebold-Mariano Test was performed on the test data considering forecasts of mulitple models to better compare their results for robustness.

2.5 Model development

This research study focused on comparing the performance of different machine-learning models: Linear Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbour (KNN), and Random Forest (RF) for the datasets with climate

Table 2 Statistical analysis of regions for Maha and Yala Seasons

Metric	Season	Anuradhapura		Kurunegala		Total
		Maha	Yala	Maha	Yala	
Mean	X1	806.4	151.03	880.62	441.71	569.94
	X2	30.68	31.88	30.60	30.80	30.99
	X3	31.22	33.95	23.00	32.69	30.22
	X4	27.53	20.15	27.55	26.57	25.45
	X5	83.18	62.45	96.97	103.24	86.46
	X6	9.05	57.74	4.05	11.87	20.68
	X7	30.76	2.76	21.95	7.89	15.84
	X8	110.36	21.21	109.58	38.18	69.84
	X9	12.63	101.79	13.42	84.82	53.16
	X10	22.63	25.04	23.81	25.11	24.15
	X11	21.87	23.38	22.14	23.25	22.66
	Prod	4230.05	3963.20	3708.34	3278.43	3795.01
Standard deviation	X1	249.14	75.34	292.52	168.79	362.23
	X2	0.29	0.25	0.30	0.28	0.59
	X3	9.30	0.30	15.85	5.46	10.42
	X4	0.60	13.04	0.63	6.39	7.84
	X5	13.17	19.53	19.08	8.86	22.16
	X6	7.33	20.23	5.07	8.27	24.60
	X7	13.27	4.80	19.70	6.03	16.65
	X8	11.58	7.54	11.45	18.23	42.64
	X9	11.58	7.54	11.45	18.23	42.64
	X10	6.85	0.22	4.01	0.43	4.07
	X11	0.34	0.20	0.56	0.37	0.77
	Prod	575.13	638.05	425.13	408.90	624.38
Range	X1	1046.30	304.50	1074.70	636.30	1413.00
	X2	1.36	1.00	1.61	1.15	2.74
	X3	38.20	1.61	36.70	34.68	38.20
	X4	3.12	29.00	3.38	28.68	29.00
	X5	56.00	81.00	90.00	36.00	103.00
	X6	27.00	84.00	18.00	35.00	109.00
	X7	57.00	27.00	92.00	26.00	97.00
	X8	47.00	29.00	51.00	91.00	115.00
	X9	47.00	29.00	51.00	91.00	115.00
	X10	32.50	1.03	26.15	2.69	32.51
	X11	1.66	1.00	2.45	1.61	3.39
	Prod	2503.00	2668.00	1751.00	1706.00	3265.00

variables. When selecting a regression model, various factors come into play depending on the nature of the problem at hand.

Linear regression is a popular choice when the relationship between the independent and dependent variables is assumed to be linear. It provides a simple yet interpretable model, allowing for easy identification of the impact of predictors on the outcome.

Since Support vector regression is robust for outliers, it is particularly useful when dealing with datasets that have outliers. Accordingly, it was used in this research to test the performance of predicting rice yield.

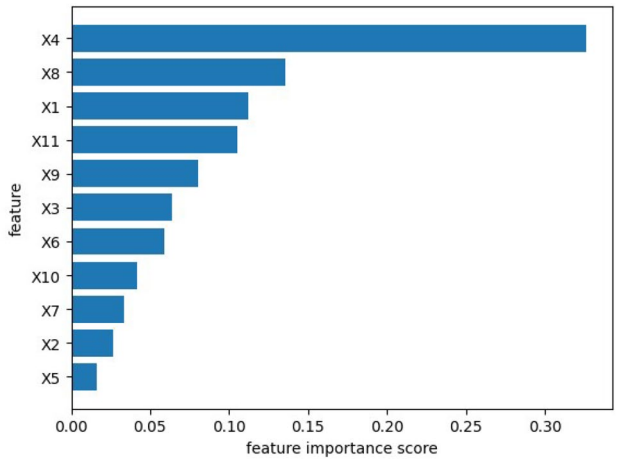

Fig. 3 Feature vs their corresponding feature importance

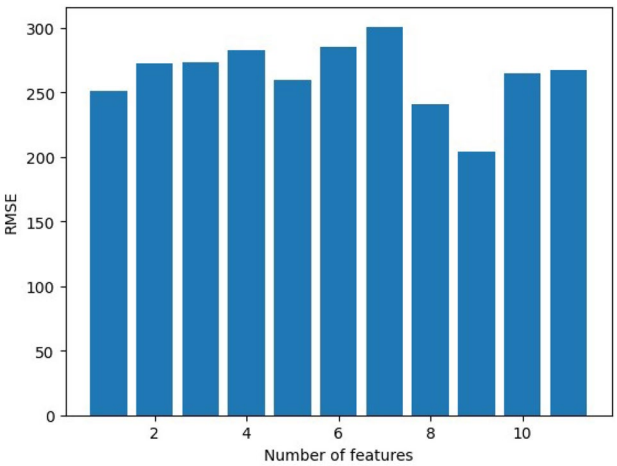

Fig. 4 Most-important features count vs RMSE

Table 3 Selected features for feature-engineered datasets

District and season	# of features selected	Selected features
Anuradhapura Maha	3	X2,X3,X10
Anuradhapura Yala	2	X1,X2
Kurunegala Maha	2	X1,X11
Kurunegala Yala	1	X4
Kurunegala total	9	X1, X3, X4, X6, X7, X8, X9, X10, X11
Anuradhapura total	9	X1, X2, X3, X4, X6, X7, X8, X9, X10
Total	1	X2

Table 4 Selected features for initial datasets

District and season	# of features selected	Selected features
Anuradhapura Maha	1	Radiation
Anuradhapura Yala	2	Rainfall, Radiation
Kurunegala Maha	3	Rainfall, Radiation, Minimum temperature
Kurunegala Yala	1	Rainfall
Kurunegala total	1	Radiation
Anuradhapura total	4	Minimum temperature, Rainfall, Maximum temperature, Radiation
Total	1	Radiation

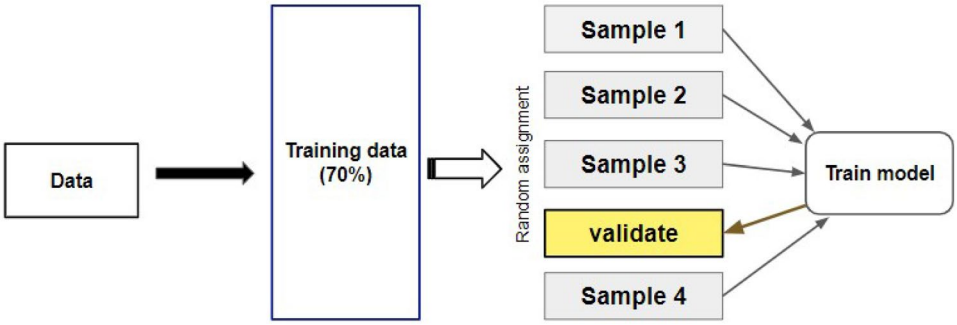

Fig. 5 Machine-learning workflow

KNN regression was used in this research to test the performance of predicting the rice yield correctly, considering the advantages it provides. KNN regression does not require any correlation between features and the target variable like in linear regression. In KNN regression *k* is an important parameter and we defined the value of *k* by selecting the *k* value which resulted in the lowest RMSE value when a KNN regression model was developed using 70% of data as train data and 30% of data as test data.

Random Forest is effective as it can capture non-linear relationships, provide feature importance measures and it can provide better predictions by combining multiple decision trees. Random Forest model has been shown to be more accurate in multiple studies [5].

The MLP (multi-layer perceptrons) regressor was considered due to its ability to handle complex nonlinear relationships. It uses multiple layers: input layer, one or more hidden layers, and a output layer for its operation. The machine-learning workflow involving any mentioned model is illustrated in Fig. 5.

For all the above model developments 5-fold cross validation was used after keeping the last 30% of the data as test data and splitting the rest of the data into 5 folds so that each fold is approximately in equal size. One fold was selected as the validation set and it was set to each fold in five instances while other folds were used for training. Then among those five models, the best model was selected by considering the lowest RMSE value. Further average RMSE and average RRMSE was tabulated for each method.

Next, selected models from each method were evaluated with the test sets and RMSE, RRMSE values were tabulated for each season and each region. That results and actual verses predicted plots were used to select the best performing model.

Python's scikit-learn library was used for training data, building machine learning models, and evaluating their performance.

2.6 Evaluation of models

The performance of each model was assessed in terms of the Root Mean Squared Error (RMSE), and Relative Root Mean Squared Error (RRMSE) by comparing the predicted and actual yield in validation sets. Root Mean Square Error (RMSE) is a measure of how well the predictions of a model align with the true values.

RMSE (Eq. 4) calculates the square root of the average of the squared differences between the predicted values and the true values. Given, n is the total number of data points, whereas y_{pr} and y_{tr} represent the predicted values and the true values respectively, RMSE is computed using the following formula,

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{pr,i} - y_{tr,i}\right)^2} \tag{4}$$

RMSE is preferred over other error metrics like Mean Absolute Error (MAE) because it penalizes larger errors more heavily. By squaring the differences between predicted and true values, it amplifies the impact of larger errors on the overall metric. Taking the square root at the end brings the RMSE back to the original scale of the data, making it easily interpretable. A lower RMSE value indicates better performance, as it signifies that the model's predictions are closer to the true values. However, it is important to consider the context of the problem and the range of the target variable when interpreting the RMSE value.

Relative Root Mean Square Error (RRMSE), also known as Normalized Root Mean Square Error (NRMSE), is a variant of the Root Mean Square Error (RMSE) that provides a relative measure of the prediction accuracy in regression tasks. RRMSE is useful for comparing the performance of different models or evaluating a model's performance across different datasets. RRMSE is calculated by dividing the RMSE by the range of the target variable(Eq. 5) or by the mean absolute value of the target variable(Eq. 6). Given, RMSE is the Root Mean Square Error, Range is the difference between the maximum and minimum values of the target variable and Mean Absolute Value is the mean of the absolute values of the target variable, the formula for RRMSE can be expressed as,

$$\text{R RMSE} = \frac{\text{RMSE}}{\text{Range}} \tag{5}$$

or

$$\text{RRMSE} = \frac{\text{RMSE}}{\text{Mean Absolute Value}} \tag{6}$$

A model was considered very good if RRMSE < 10%, good if RRMSE is in between 10% and 20%, fair if RRMSE is in between 20% and 30% and poor if RRMSE > 30%. When evaluating models to select an acceptable model which provides better predictions than other models, predicted vs actual plots and residual plots were also considered in addition to above metrics.

3 Results and discussion

In this section, we present the results of our study and discuss them in detail. The study was based on data from 1982 to 2019 in Anuradhapura and Kurunegala areas, where rice harvest is common in Sri Lanka.

3.1 Results

The main growing season, called Maha, always produced more rice than the off-season, called Yala. On average, Maha produced 770 kg more per hectare in Anuradhapura and 550 kg more per hectare in Kurunegala (see Fig. 1). Overall, Anuradhapura also had better yields than Kurunegala, about 425 kg more per hectare on average.

Both areas saw their harvests get bigger over time, but there were large swings from year to year, especially during Yala in Kurunegala. This suggests things like weather and pests play a big role. Looking at temperature data, we found that Yala was consistently hotter than Maha in both areas (Fig. 2), with a slight overall warming of about 0.03 °C per year.

The RF model did a better job of predicting harvests than other models considered (like LR, SVR, MLP, and KNN) across the board - in both areas, in both seasons, and with all our data (Tables 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17).

Checking how well the RF model fit the data, we found it was a good fit - it wasn't over or under predicting in a systematic way. We also looked at plots of the actual and predicted yields (Figs. 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17), and it did a pretty good job overall. However, it wasn't perfect, especially in some years and situations in the Yala season. This tells us that while the model is useful, there's still room for improvement, especially for predicting Yala harvests.

3.2 Discussion

In this study, we used machine learning to predict rice yield in Sri Lanka's dry and inter-central regions. Those are the major paddy producing areas of Sri Lanka. In this case we used less number of parameters. In particular, we take into account the factors of temperature, rainfall and solar radiation. The main purpose of this study was to investigate the practical use of the models we use.

The insights we collect from these predictive models helps policymakers make farming decisions. Knowing how weather impacts rice growth lets policymakers help farmers, ensure food supply, and promote eco-friendly farming. We used machine learning to make effective predictions about rice yield, better than the traditional, so we can forecast rice

Table 5 Train and validation errors for each machine-learning model for *Maha* season in Anuradhapura district

Metric	Dataset	Machine-learning Model									
		LR		SVR		MLP		KNN		RF	
		Train	Validation	Train	Validation	Train	Validation	Train	Validation	Train	Validation
RMSE (kg/ha)	Initial	302.96	442.59	470.07	539.34	364.88	393.50	386.68	416.25	155.34	431.48
	Feature-Engineered	426.27	495.46	464.29	479.07	414.98	508.29	213.05	487.33	177.15	464.71
RRMSE (%)	Initial	7.54	10.90	11.70	13.45	9.08	9.80	9.63	10.28	3.87	10.88
	Feature-Engineered	10.43	12.35	11.35	11.89	10.15	12.65	5.21	12.08	4.33	11.56
MAE (kg/ha)	Initial	243.19	400.82	374.70	471.04	285.33	323.15	315.92	357.72	130.33	368.87
	Feature-Engineered	352.67	416.50	366.31	418.81	339.23	422.02	156.50	418.50	137.77	390.02
R2 Score	Initial	0.56	− 2.43	− 0.058	− 2.91	0.36	− 0.76	0.28	− 1.90	0.88	− 0.56
	Feature-Engineered	0.09	−1.05	− 0.09	− 1.15	0.14	− 1.28	0.77	− 1.40	0.84	− 0.80
BIAS	Initial	− 1.47e−13	75.83	−9.35e+01	−89.42	8.07e−03	8.35	7.72e+01	85.75	− 1.39e+01	33.75
	Feature-Engineered	1.5e−13	−5.28	−0.01	−104.22	0.01	−2.07	9.49	−40.25	1.94	− 38.44

Table 6 Train and validation errors for each machine-learning model for *Yala* season in Anuradhapura district

Metric	Dataset	Machine-learning Model									
		LR		SVR		MLP		KNN		RF	
		Train	Validation	Train	Validation	Train	Validation	Train	Validation	Train	Validation
RMSE (kg/ha)	Initial	434.30	513.03	533.36	605.00	426.15	499.20	409.78	531.61	198.69	516.89
	Feature-Engineered	430.29	461.49	545.25	526.07	431.33	467.59	448.48	468.60	201.09	448.43
RRMSE (%)	Initial	11.67	13.78	14.32	16.11	11.45	13.43	11.01	14.29	5.35	13.91
	Feature-Engineered	11.21	12.17	14.20	13.86	11.24	12.33	11.69	12.38	5.24	12.92
MAE(kg/ha)	Initial	356.77	438.15	444.64	525.05	349.69	425.73	330.00	456.01	159.20	418.32
	Feature-Engineered	346.14	387.12	437.21	440.41	344.98	393.03	356.71	399.90	164.48	421.28
R2 Score	Initial	0.33	−1.29	−0.003	−3.44	0.35	−1.14	0.40	−1.71	0.86	−1.41
	Feature–Engineered	0.37	− 0.14	− 0.00	− 0.28	0.37	− 0.18	0.31	− 0.11	0.86	− 0.17
BIAS	Initial	5.31e-15	3.86	− 1.52e+01	− 9.99	− 3.026e-01	− 2.77	− 3.46e+01	− 31.59	− 7.36e+00	− 0.48
	Feature-Engineered	− 1.0e-13	6.06	− 19.32	− 17.67	0.84	7.94	25.46	43.89	− 10.99	− 46.64

Table 7 Train and validation errors for each machine-learning model for *Maha* season in Kurunegala district

Metric	Dataset	Machine-learning Model									
		LR		SVR		MLP		KNN		RF	
		Train	Validation	Train	Validation	Train	Validation	Train	Validation	Train	Validation
RMSE(kg/ha)	Initial	194.44	255.66	203.12	208.98	319.48	373.80	206.81	231.90	89.83	261.83
	Feature-Engineered	144.98	154.35	182.14	195.16	145.93	160.55	165.86	193.92	63.87	168.51
RRMSE (%)	Initial	5.52	7.23	5.77	5.94	9.07	10.54	5.87	6.58	2.55	7.43
	Feature-Engineered	4.12	4.37	5.18	5.59	4.14	4.55	4.71	5.54	1.81	4.79
MAE(kg/ha)	Initial	150.78	198.15	151.06	161.17	236.51	304.66	161.65	178.99	64.48	194.48
	Feature-Engineered	100.17	115.39	139.07	166.84	99.76	121.35	128.87	158.6	50.83	137.44
R2 Score	Initial	0.081	− 1.38	− 0.006	− 0.55	− 2.26	− 4.99	− 0.04	− 0.85	0.80	− 1.80
	Feature-Engineered	0.35	− 0.76	− 0.02	− 1.44	0.34	− 0.86	0.16	− 1.64	0.87	− 1.15
BIAS	Initial	− 1.62e−13	7.38	− 7.88e+00	− 8.09	8.50e+00	138.37	− 1.34e+01	− 13.81	− 2.85e+00	− 14.33
	Feature-Engineered	− 5.1e−15	− 11.99	− 22.71	− 31.22	0.41	− 9.38	− 3.85e1	− 58.67	− 7.67	− 43.69

Table 8 Train and validation errors for each machine-learning model for *Yala* season in Kurunegala district

Metric	Dataset	Machine-learning Model									
		LR		SVR		MLP		KNN		RF	
		Train	Validation	Train	Validation	Train	Validation	Train	Validation	Train	Validation
RMSE(kg/ha)	Initial	323.24	354.60	326.81	339.60	323.24	354.60	328.16	344.32	140.94	402.33
	Feature-Engineered	249.02	256.21	314.87	300.72	248.85	256.88	236.05	257.52	126.51	301.65
RRMSE(%)	Initial	10.41	11.43	10.53	10.95	10.40	11.43	10.57	11.11	4.54	13.01
	Feature-Engineered	8.09	8.32	10.24	9.79	8.09	8.34	7.67	8.36	4.11	9.83
MAE(kg/ha)	Initial	258.31	282.84	258.19	271.88	258.32	282.84	262.63	270.84	122.84	360.85
	Feature-Engineered	179.13	192.37	246.95	263.03	178.95	193.14	181.79	206.21	95.01	249.74
R2 Score	Initial	0.020	− 0.329	− 0.001	− 0.23	0.020	− 0.329	− 0.10	− 0.26	0.81	− 0.86
	Feature-Engineered	0.38	− 1.15	0.01	− 0.41	0.38	− 1.16	0.44	− 1.14	0.84	− 2.16
BIAS	Initial	1.098e−13	− 2.62	− 1.20e+01	− 10.999	2.50e-03	− 2.62	− 1.50e+01	− 10.40	1.15e+00	− 40.86
	Feature-Engineered	9.1e−15	− 6.77	− 7.37	− 7.27	0.03	− 6.05	2.07e1	27.96	8.71	22.62

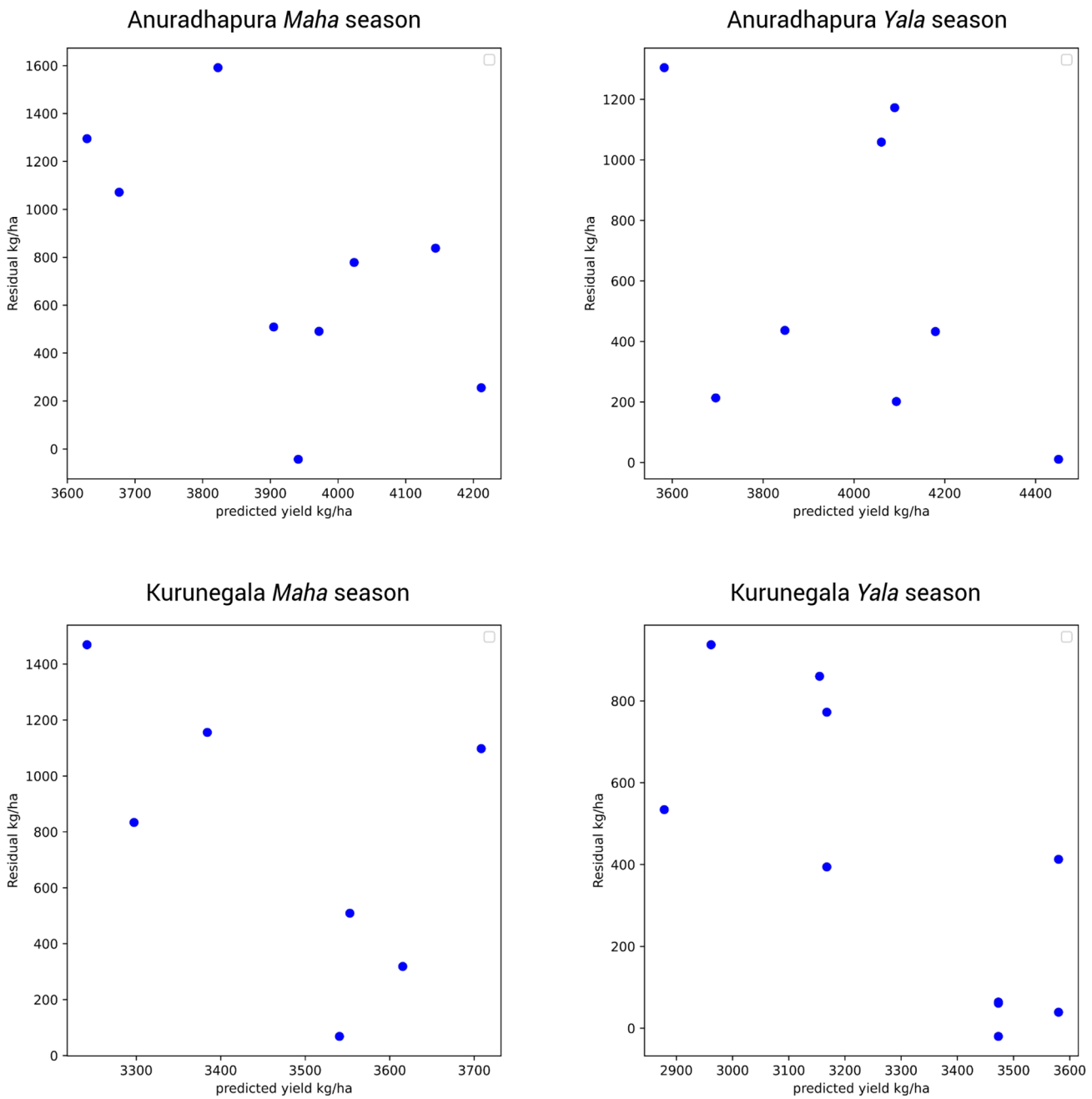

Fig. 6 Residual plots for test set of feature-engineered datasets

yield more accurately. With climate change affecting farming globaly, predicting crop yields well is key for being ready and adapting. This study stresses how crucial it is to use climate data in our predictions, so we can act smartly to protect rice production from changing weather [26, 27, 22].

By focusing on simple weather data as inputs, we can implement this approach cost effectively, which is especially helpful in places where resources are limited. The insights we get from these predictive models can help policymakers make decisions based on evidence. When policymakers understand how weather affects rice yield, they can make plans to help farmers, make sure we have enough food, and encourage farming practices that do not harm the environment. The Random Forest (RF) model consistently did better than other models in capturing how weather affects rice yield, which matches with what other studies have found [19, 20].

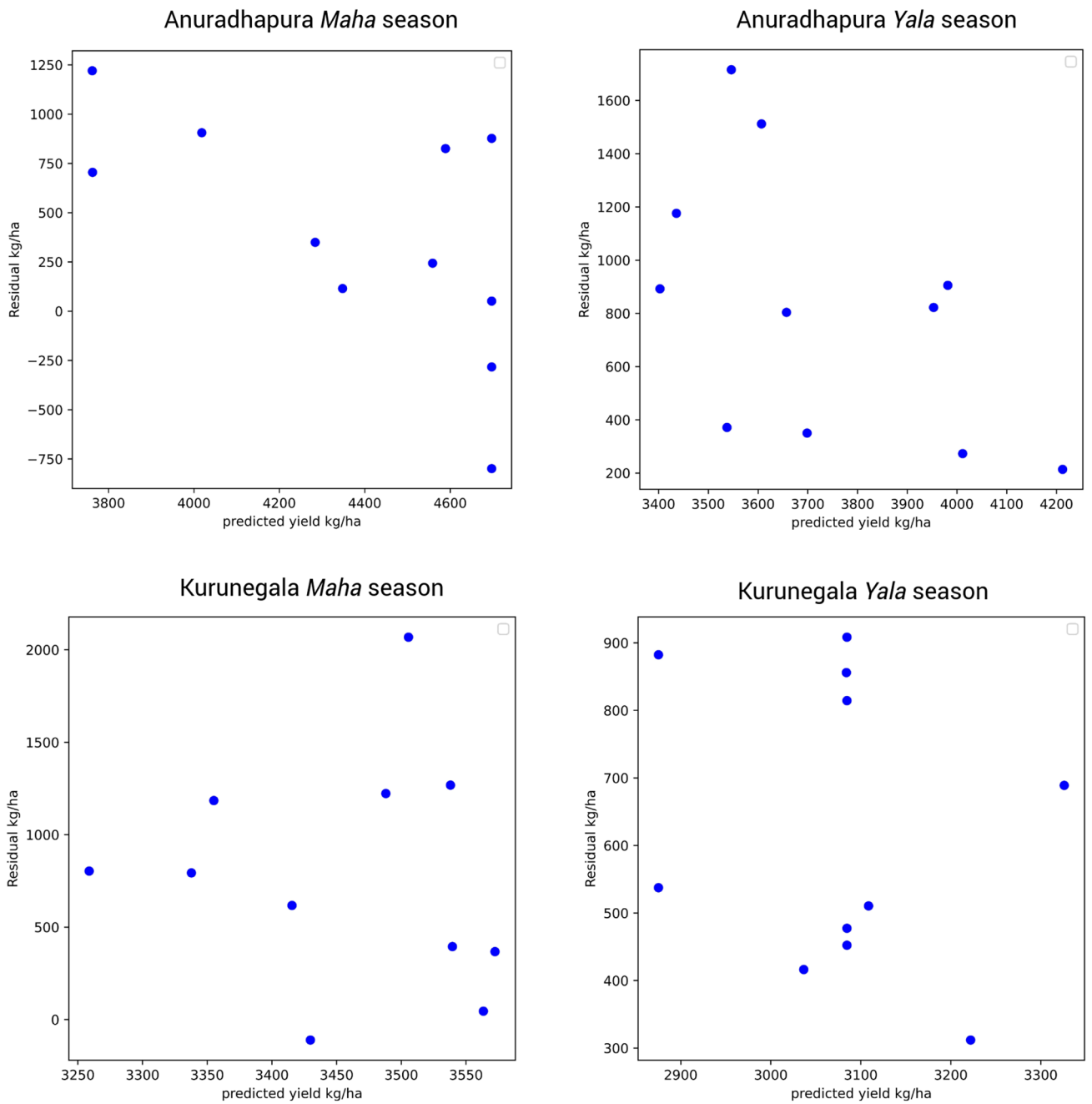

Fig. 7 Residual plots for test set of initial datasets

Table 9 RMSE and RRMSE values for the test set of Anuradhapura *Maha* season

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	685.597	725.076	685.603	792.289	671.918
	Feature Engineered	670.09	643.36	616.49	870.20	896.463
RRMSE (%)	Initial	14.41	15.24	14.41	16.66	14.13
	Feature Engineered	14.32	13.75	13.18	18.60	19.16

Table 10 RMSE and RRMSE values for the test set of Anuradhapura *Yala* season

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	946.079	894.663	950.269	1031.733	1015.833
	Feature Engineered	763.38	865.922	764.744	812.059	1010.568
RRMSE (%)	Initial	20.78	19.65	20.87	22.66	22.31
	Feature Engineered	16.58	18.81	16.61	17.64	21.95

Table 11 RMSE and RRMSE values for the test set of Kurunegala *Maha* season

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE (kg/ha)	Initial	941.848	937.642	940.969	923.589	988.437
	Feature Engineered	967.849	840.888	978.917	871.25	907.14
RRMSE (%)	Initial	22.20	22.11	22.18	21.77	23.30
	Feature Engineered	22.74	19.76	23	20.47	21.31

Table 12 RMSE and RRMSE values for the test set of Kurunegala *Yala* season

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	573.988	596.298	573.948	655.564	711.980
	Feature Engineered	636.791	718.392	636.782	625.76	532.589
RRMSE (%)	Initial	15.50	16.11	15.50	17.71	19.23
	Feature Engineered	17.23	19.43	17.23	16.93	14.41

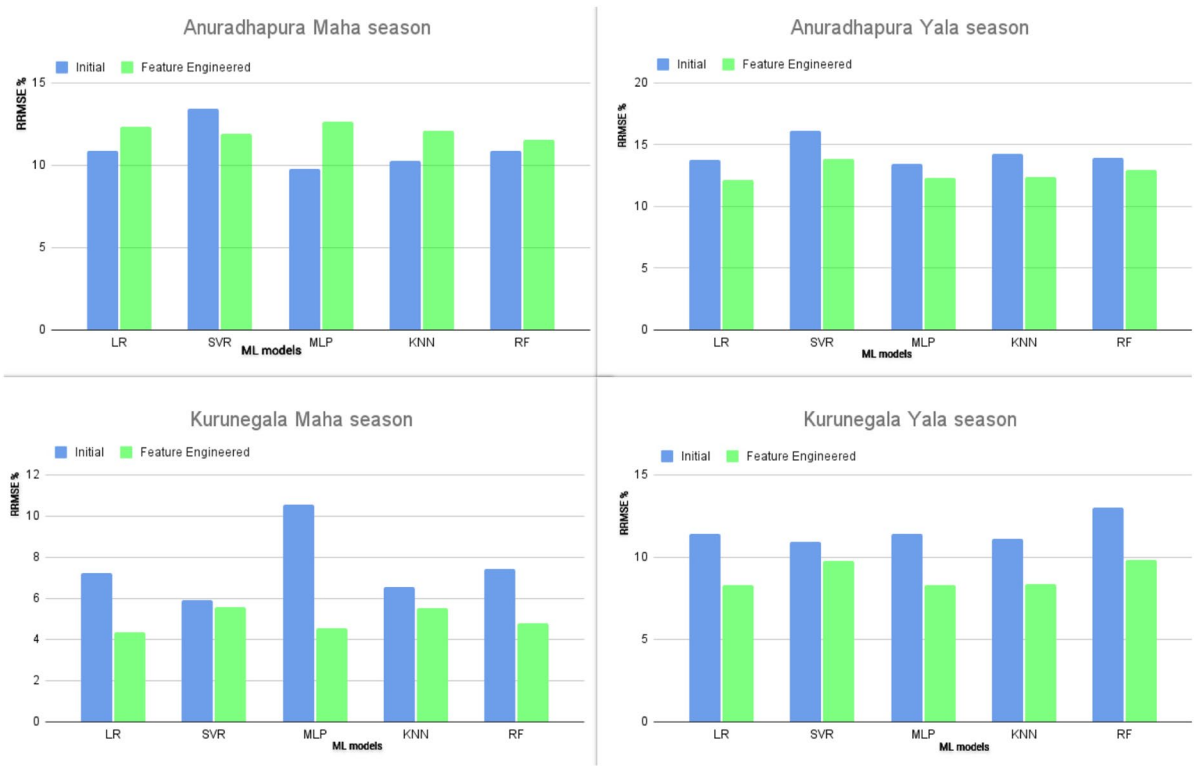

Fig. 8 Average RRMSE for validation datasets for two seasons of Anuradhapura and Kurunegala

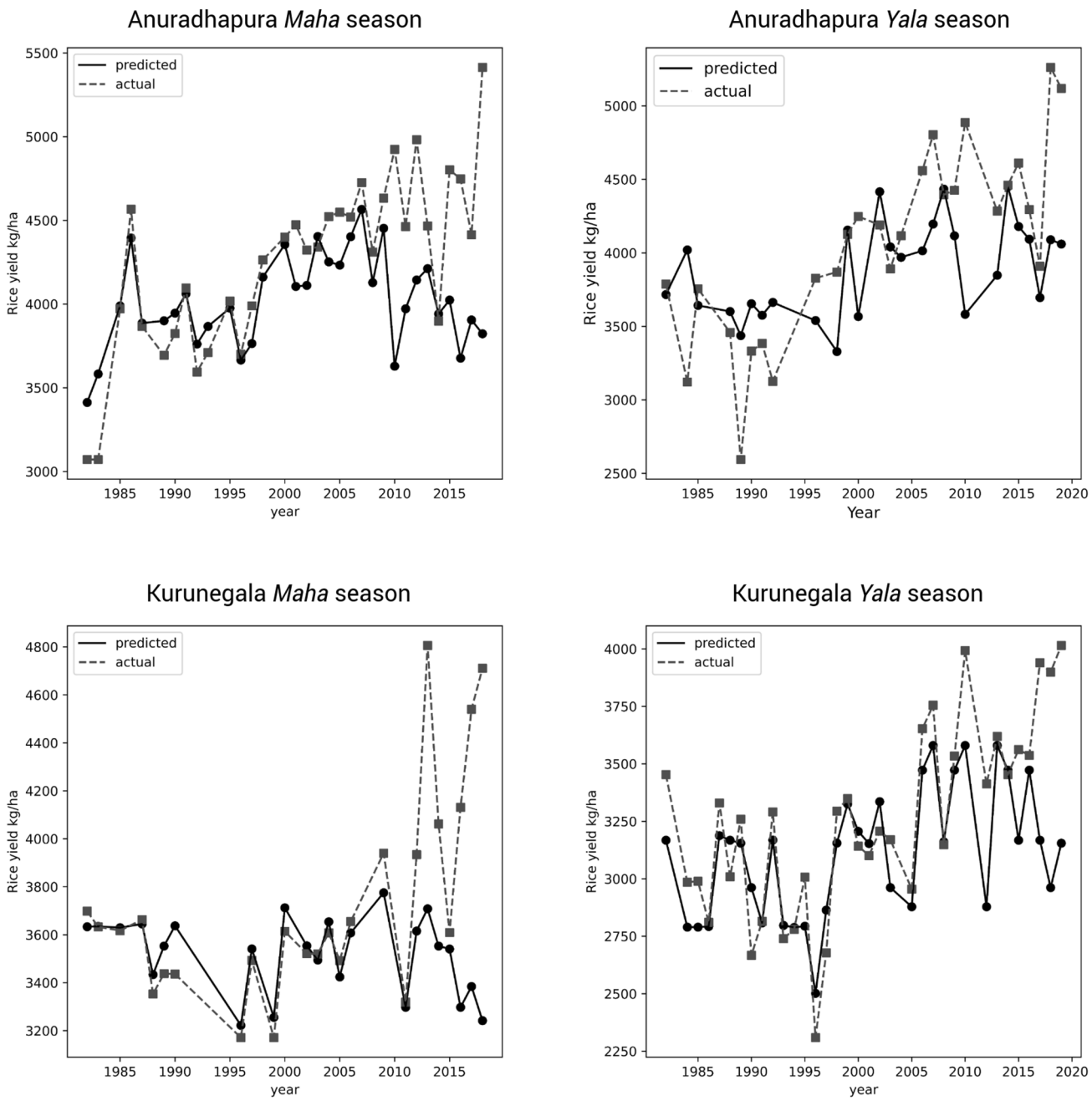

Fig. 9 Actual vs predicted plots of feature-engineered datasets

The model's ability to handle complex relationships among different factors likely enhances its performance compared to other models like LR, SVR, MLP, and KNN. When considering the factors that most affect rice yield, temperature-related variables, especially cooler temperatures, appear to be the most significant. This emphasizes the importance of maintaining ideal temperatures during key growth periods to ensure optimal rice production.

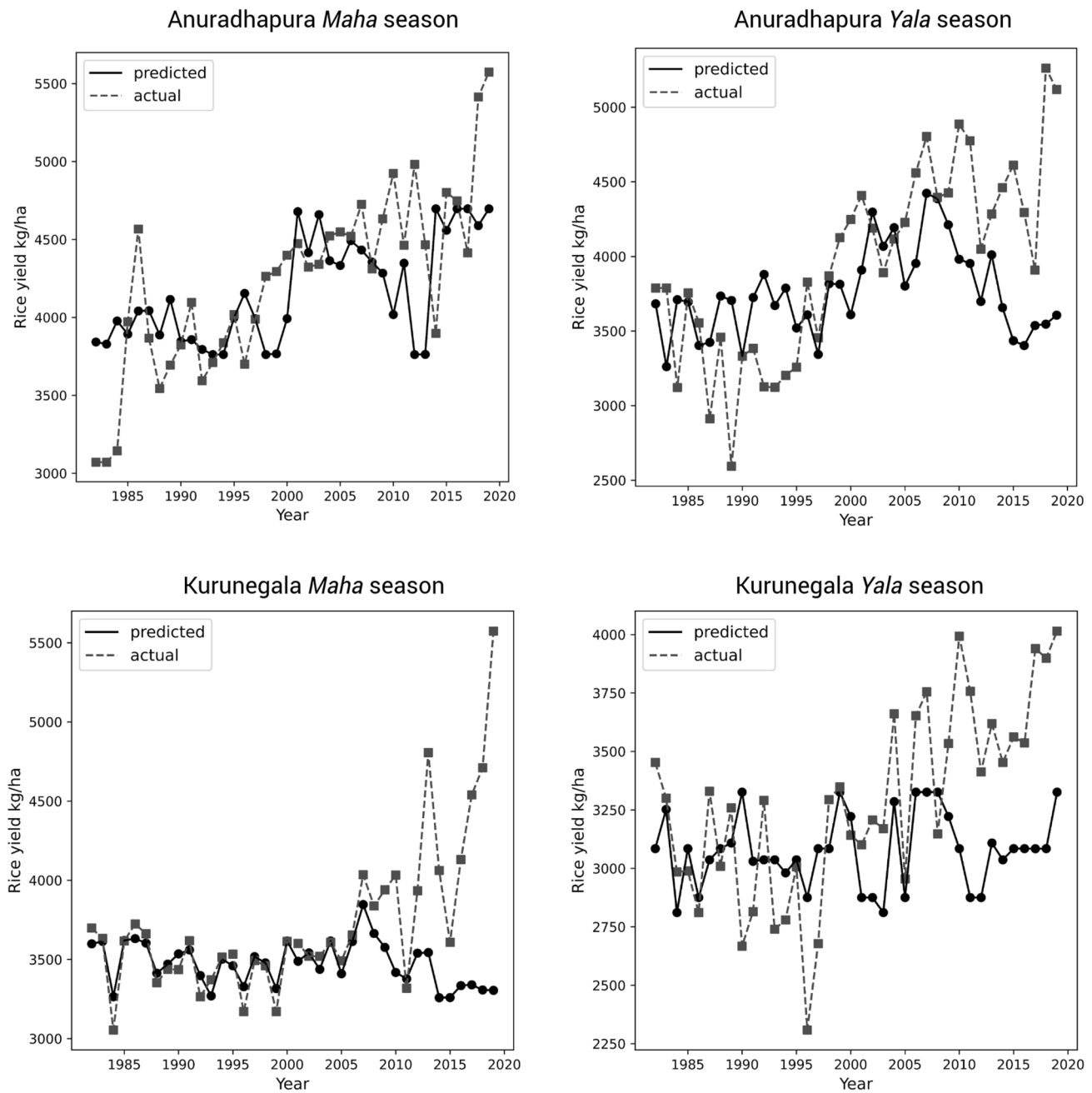

Fig. 10 Actual vs predicted plots of initial datasets

The results also demonstrate that feature engineering can improve the model and reduce the need for lots of attribute data. It also shows that picking features in machine learning depends on the problem. While feature engineering makes the RF model more accurate, especially in the Kurunegala Yala season, it made some other models less accurate in certain situations. This tells us that the best features for predicting rice yield might change depending on where, when, and what model we are using, like previous studies have noted [29].

We need more research to understand why these differences occur and to develop better strategies for the model to work well in different situations. By checking the residuals and comparing actual yields to predicted ones, we can see

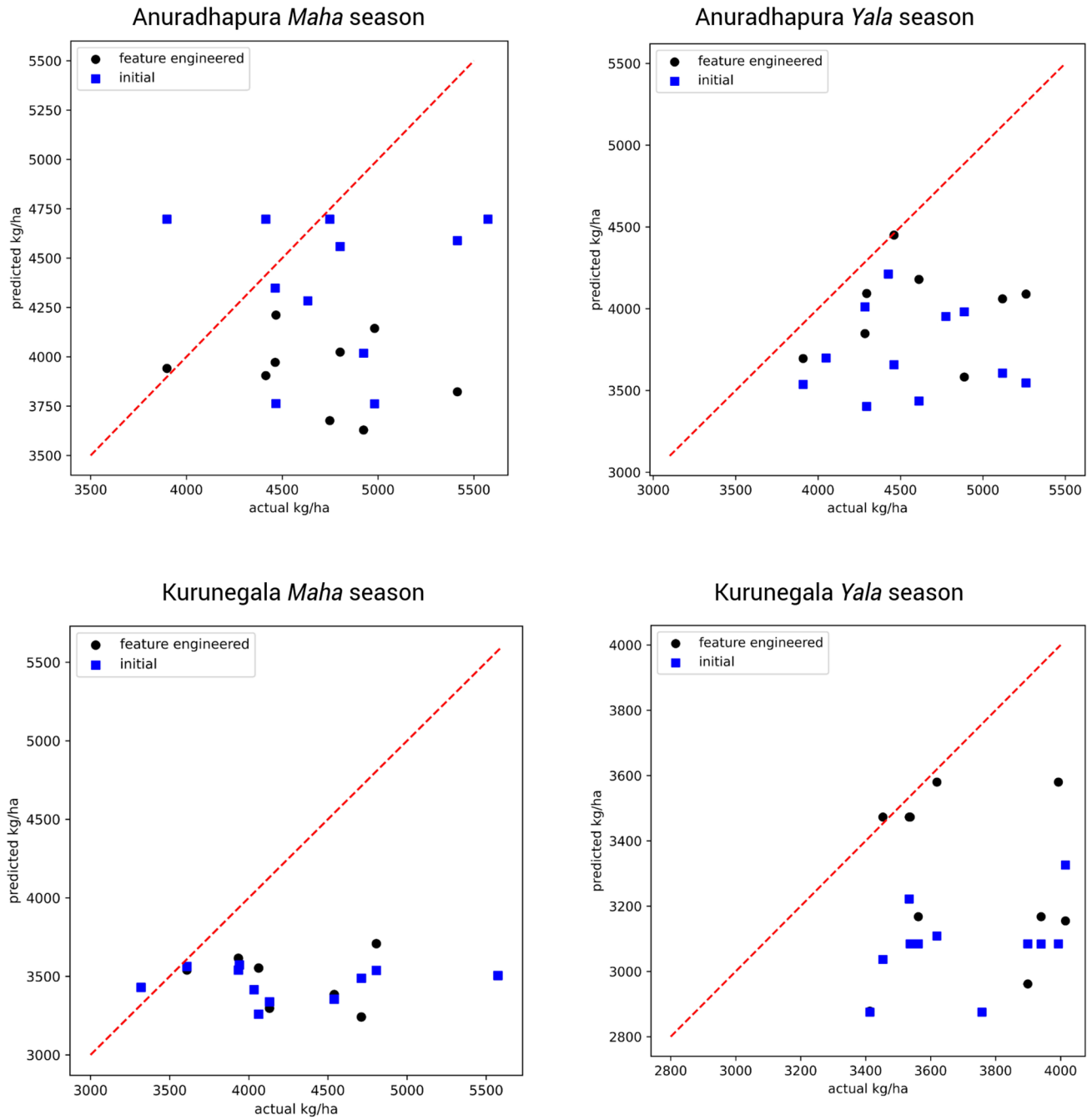

Fig. 11 Actual vs predicted plots for validation sets

Table 13 Average RMSE and RRMSE values for validation sets of Anuradhapura total dataset

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	571.81	538.89	568.00	515.80	524.80
	Feature-Engineered	722.06	539.18	617.39	507.92	539.72
RRMSE(%)	Initial	14.66	13.82	14.56	13.27	13.53
	Feature-Engineered	18.71	13.83	15.74	13.05	13.88

Table 14 Average RMSE and RRMSE values for validation sets of Kurunegala total dataset

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE (kg/ha)	Initial	245.33	237.30	243.06	241.51	277.56
	Feature-Engineered	289.29	236.50	583.83	229.20	225.23
RRMSE(%)	Initial	7.408	7.16	7.34	7.29	8.35
	Feature-Engineered	8.76	7.13	17.56	6.92	6.82

Table 15 RMSE and RRMSE values for the test set of Anuradhapura total dataset

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	929.30	990.90	931.12	865.61	900.34
	Feature Engineered	635.33	835.45	1092.07	783.63	846.50
RRMSE (%)	Initial	19.97	21.29	20.00	18.60	19.34
	Feature Engineered	13.65	17.95	23.46	16.84	18.19

Table 16 RMSE and RRMSE values for the test set of Kurunegala total dataset

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	764.55	760.86	769.50	745.66	852.74
	Feature Engineered	642.81	686.66	952.45	734.68	681.79
RRMSE (%)	Initial	19.25	19.16	19.37	18.77	21.47
	Feature Engineered	16.34	17.46	24.21	18.68	17.33

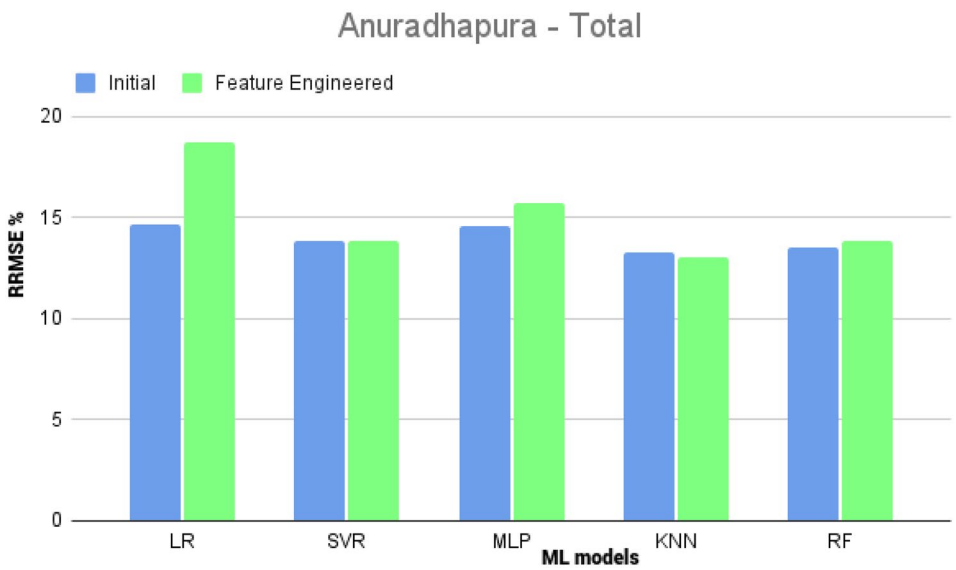

Fig. 12 Average RRMSE for validation sets of Anuradhapura total dataset

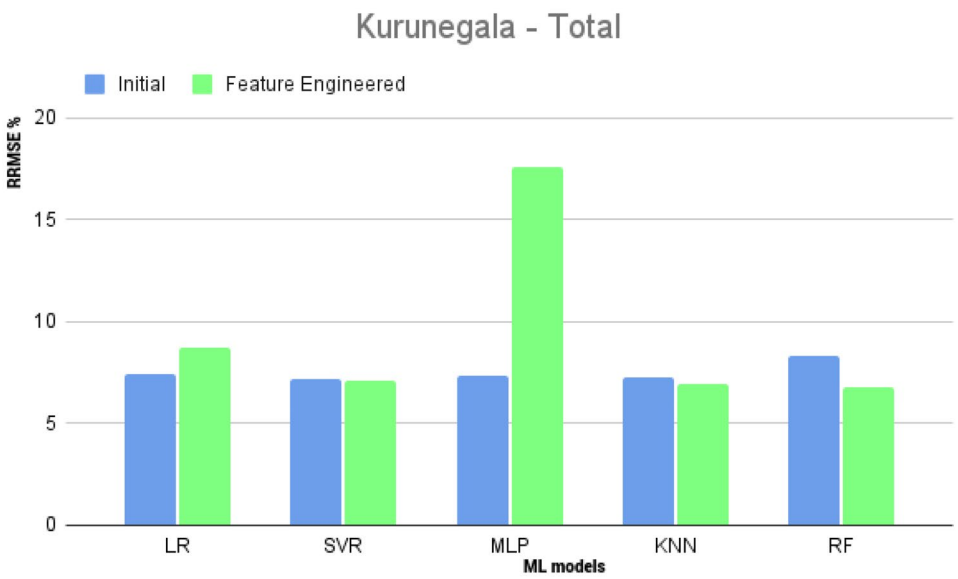

Fig. 13 Average RRMSE for validation sets of Kurunegala total dataset

how well the model is doing. The randomly scattered residuals and absence of systematic bias indicate that the RF model fits the data well and generally provides accurate yield predictions.

Generally, Random Forest showed lower RMSE and MAE values compared to other models, indicating better prediction accuracy for both test and validation datasets. For example, in the Anuradhapura Maha dataset, Random Forest achieved an RMSE of 568.087 on the test set, which is significantly lower than that of the SVR (663.644) and KNN (691.728).Linear Regression, although simpler, performed reasonably well but was often outperformed by more sophisticated models like Random Forest and SVR in terms of RMSE and R2 score.

As per the residual distribution plots, there were no specific patterns observed in the distributions that violate the normality 7,6. The Diebold-Mariano Test value corresponding to multiple models calculated on the test data is provided in Supplementary Material. The p-values indicate the significance of these differences. A low p-value (typically less than 0.05 or 0.1) suggests a significant difference in the predictive accuracy of the two models being compared.

The feature-engineered datasets showed an overall improvement in model performance. This is evidenced by reduced RMSE and RRMSE values across most models and datasets. For instance, in the Anuradhapura Yala dataset, the RMSE of Linear Regression decreased from 946.079 in the initial dataset to 763.380 after feature engineering. Accordingly, this work demonstrates how appropriate feature engineering coupled with Machine Learning benefit in rice yield prediction based on less number of climatic variables.

This study fits closely with a two of United Nations Sustainable Development Goals (SDGs): Zero Hunger (SDG 2) and Climate Action (SDG 13).

4 Conclusion

It was notable that Random Forest outperformed other models. This allowed us to confirm that random forest performs well even with low parameters which has not been demonstrated in previous work on the rice yield prediction in Sri Lanka. This is a situation that has not really been used in research before. Being able to show high performance using such low weather data was a specialty of our research. The results demonstrate that comparable performance can be achieved with less number of features, through feature engineering. This result is important especially when it is difficult to gather a lot of weather data in practice. However, the varying impact of feature engineering across different models and data sets underscores the need for tailored feature engineering strategies. Although the model shows limitations in predicting yields under specific conditions, especially for the early season and initial data set, this research still provides a valuable tool for data-driven decision making in rice cultivation, contributing to improved agricultural practices, improved food security, and climate change adaptation.

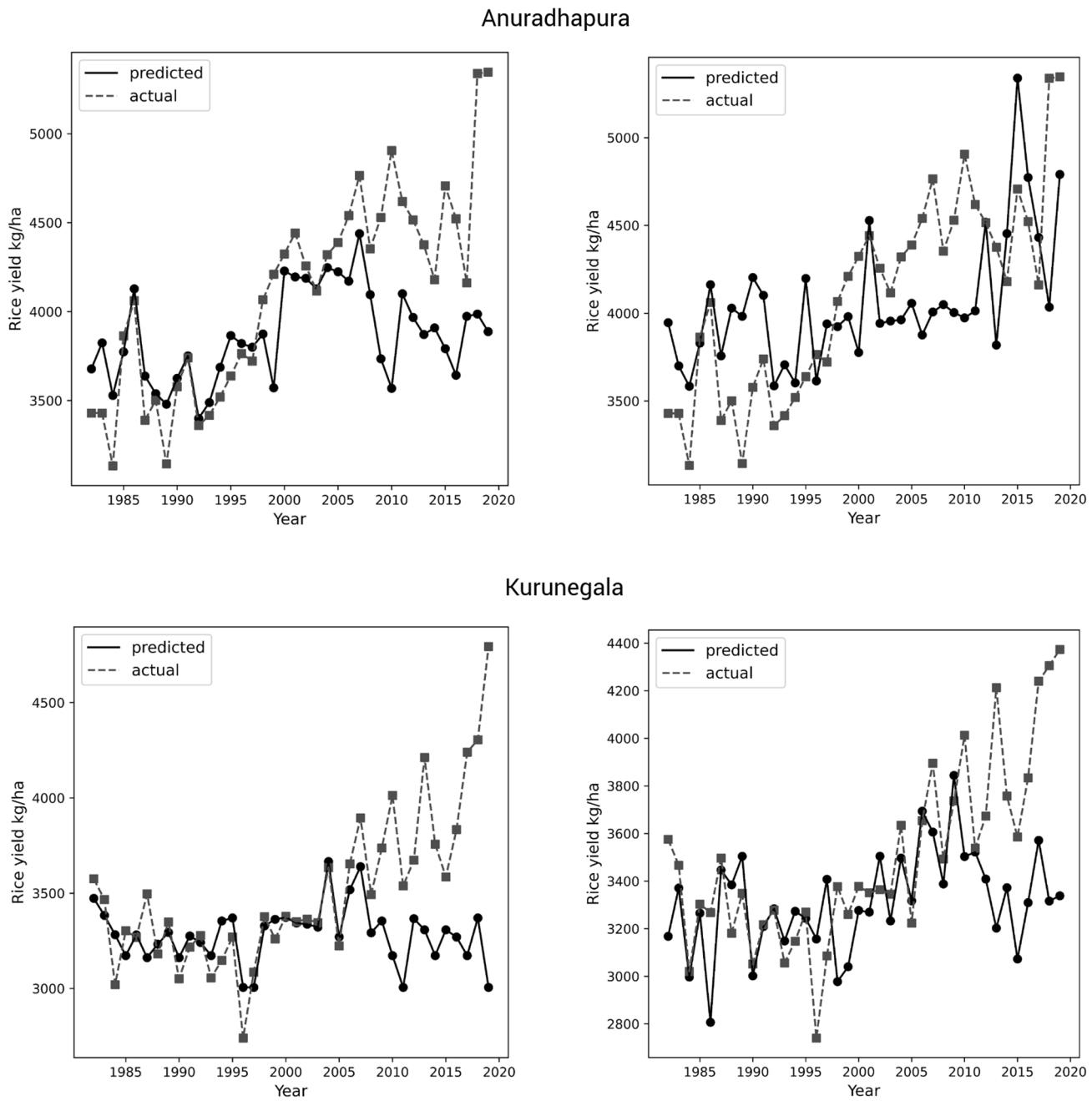

Fig. 14 Actual vs predicted plots for Total datasets of Anuradhapura and Kurunegala initial dataset (left) and feature engineered dataset (right)

Table 17 Average RMSE and RRMSE values for test sets of total dataset

Metric	Dataset	Machine-learning model				
		LR	SVR	MLP	KNN	RF
RMSE(kg/ha)	Initial	724.00	728.66	723.97	709.51	704.37
	Feature-Engineered	804.74	767.14	805.38	808.64	778.26
RRMSE(%)	Initial	16.79	16.89	16.78	16.45	16.33
	Feature-Engineered	18.74	17.87	18.76	18.83	18.12

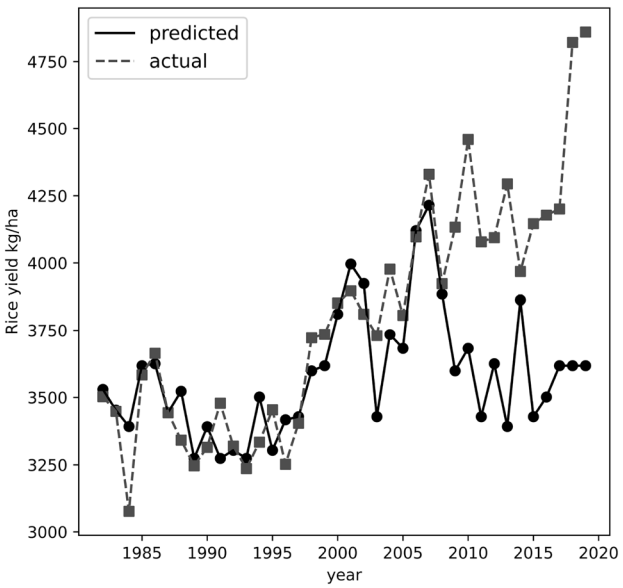

Fig. 15 Actual vs predicted plot for feature-engineered total dataset

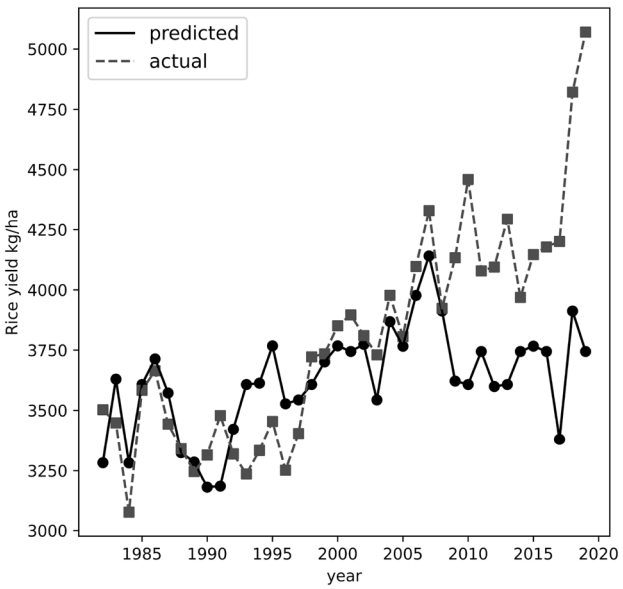

Fig. 16 Actual vs predicted plot for initial total dataset

The focus of our work was to derive a method for rice yield prediction with minimal number of features, considering the difficulties in measurements of climatic variables in developing countries similar to Sri Lanka. Compared to traditional methods, machine learning is able to extract complex patterns and enables developing adoptable solutions. We considered machine learning methods in their simplest and basic forms for rice yield prediction and observed decent performances. Exploration on improvements to the machine learning models used including regularisation, incorporating more interpretable methods is proposed as a future directive. The insights gained from extended

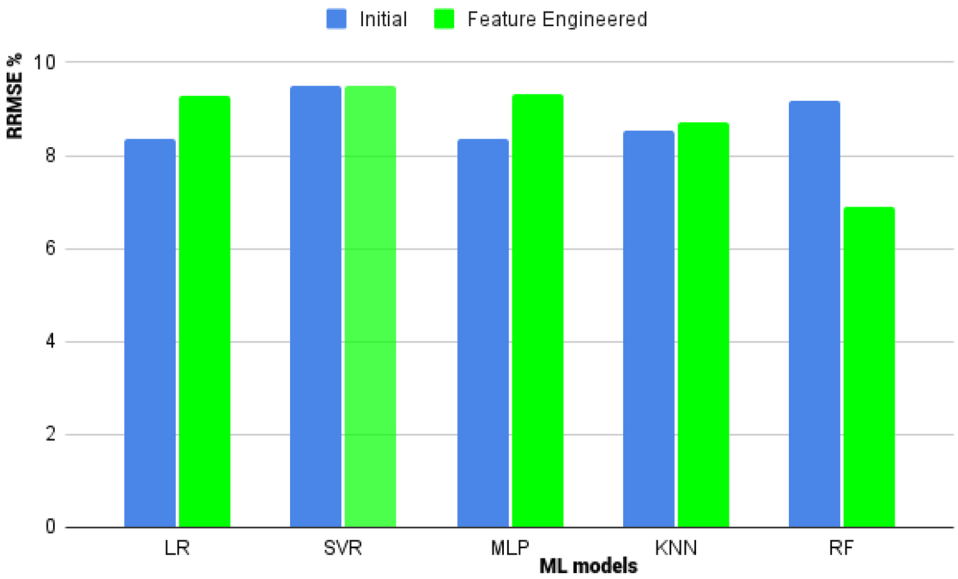

Fig. 17 Average RRMSE values for validation sets of total dataset

work can further inform climate-smart agricultural policies and contribute to the achievement of the United Nations Sustainable Development Goals.

Acknowledgements We acknowledge Research and Innovation Centers Division of Rabdan Academy, Abu Dhabi, United Arab Emirates for publication support and Natural Resources Management Centre (NRMC), Department of Agriculture, Peradeniya, Sri Lanka for weather data.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Nuwan De Silva, Janaka Alawatugoda, Damayanthi Herath, Mojith Ariyaratne and Ruwanga Amarasinghe. The first draft of the manuscript was written by Aminda Amarasinghe, Ishini Sangarasekara and Jinendra Bogahawatte, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work is supported by Rabdan Academy Research Funding.

Data availability Data sets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that there is no Competing interests.

References

1. United Nations Department of Economic and Social Affairs Sustainable Development, Available at https://sdgs.un.org/goals, Accessed 22 February 2024.
2. Wickramasinghe L, Weliwatta R, Ekanayake P, Jayasinghe J. Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques. J Math. 2021. https://doi.org/10.1155/2021/6646126.

3. Dias M.P.N.M., Navaratne C.M., Weerasinghe K.D.N., Hettiarachchi R.H.A.N. Application of DSSAT crop simulation model to identify the changes of rice growth and yield in nilwala river basin for mid-centuries under changing climatic conditions. Procedia Food Sci. 2016;6:159–63. https://doi.org/10.1016/j.profoo.2016.02.039.
4. Rezapour S, Jooyandeh E, Ramezanzade M, Mostafaeipour A, Jahangiri M, Issakhov A, Chowdhury S, Techato K. Forecasting rainfed agricultural production in arid and semi-arid lands using learning machine methods: a case study. Sustainability. 2021;13:4607. https://doi.org/10.3390/su13094607.
5. Ekanayake P, Rankothge W, Weliwatta R, Jayasinghe JW. Machine learning modelling of the relationship between weather and paddy yield in Sri Lanka. J Math. 2021. https://doi.org/10.1155/2021/9941899.
6. Paddy Statistics (2022), Department of Census and Statistics, Sri Lanka, Available at http://www.statistics.gov.lk/Agriculture/StaticalInformation/PaddyStatistics/MetricUnits/IncludingMahaweli/2021-2022Maha.pdf, Accessed 2 January 2023.
7. Alfred R, Obit JH, Chin CP-Y, Haviluddin H, Lim Y. Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. IEEE Access. 2021;9:50358–80. https://doi.org/10.1109/ACCESS.2021.3069449.
8. Hathurusingha C, Abdelhamid N, Airehrour D. Forecasting models based on data analytics for predicting rice price volatility: a case study of the Sri Lankan rice market. J Inf Knowl Manag. 2019;18(01):1950006. https://doi.org/10.1142/S0219649219500060.
9. Zhao S, Zheng H, Chi M, Chai X, Liu Y. Rapid yield prediction in paddy fields based on 2D image modelling of rice panicles. Comput Electron Agric. 2019;162:759–66. https://doi.org/10.1016/j.compag.2019.05.020.
10. Lingwal S, Bhatia KK, Singh M. A novel machine learning approach for rice yield estimation. J Exp Theor Artif Intell. 2024;36(3):337–56. https://doi.org/10.1080/0952813X.2022.2062458.
11. Chu Z, Yu J. An end-to-end model for rice yield prediction using deep learning fusion. Comput Electron Agric. 2020;174: 105471. https://doi.org/10.1016/j.compag.2020.105471.
12. Nesarani A, Ramar R, Pandian S. An efficient approach for rice prediction from authenticated Block chain node using machine learning technique. Environ Technol Innov. 2020;20: 101064. https://doi.org/10.1016/j.eti.2020.101064.
13. Azmi N, Kamarudin LM, Zakaria A, Ndzi DL, Rahiman MHF, Zakaria SMMS, Mohamed L. RF-based moisture content determination in rice using machine learning techniques. Sensors. 2021;21(5):1875. https://doi.org/10.3390/s21051875.
14. Liu L-W, Hsieh S-H, Lin S-J, Wang Y-M, Lin W-S. Rice Blast (Magnaporthe oryzae) Occurrence prediction and the key factor sensitivity analysis by machine learning. Agronomy. 2021;11(4):771. https://doi.org/10.3390/agronomy11040771.
15. Krishnamoorthy N, Prasad LN, Kumar CP, Subedi B, Abraha HB, Sathishkumar VE. Rice leaf diseases prediction using deep neural networks with transfer learning. Environ Res. 2021;198: 111275. https://doi.org/10.1016/j.envres.2021.111275.
16. Jeong S, Ko J, Yeom J-M. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. Sci Total Environ. 2022;802: 149726. https://doi.org/10.1016/j.scitotenv.2021.149726.
17. An G, Xing M, He B, Liao C, Huang X, Shang J, Kang H. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. Remote Sens. 2020;12(18):3104. https://doi.org/10.3390/rs12183104.
18. Sengupta S, Bhattacharyya K, Mandal J, Bhattacharya P, Halder S, Pari A. Deficit irrigation and organic amendments can reduce dietary arsenic risk from rice: introducing machine learning-based prediction models from field data. Agric Ecosyst Environ. 2021;319: 107516. https://doi.org/10.1016/j.agee.2021.107516.
19. Grinberg NF, Orhobor OI, King RD. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat. Mach Learn. 2020;109(2):251–77. https://doi.org/10.1007/s10994-019-05848-5.
20. Son NT, Chen CF, Chen CR, Guo HY, Cheng YS, Chen SL, et al. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan, Int J Remote Sens. 2020;41(20):7868–88. https://doi.org/10.1080/01431161.2020.1766148.
21. Tan S, Liu J, Lu H, Lan M, Yu J, Liao G, Wang Y, Li Z, Qi L, Ma X. Machine learning approaches for rice seedling growth stages detection. Front Plant Sci. 2022;13: 914771. https://doi.org/10.3389/fpls.2022.914771.
22. Sarwary M, Samiappan S, Khan GD, Moahid M. Climate change and cereal crops productivity in Afghanistan: evidence based on panel regression model. Sustainability. 2023;15(14):10963. https://doi.org/10.3390/su151410963.
23. Iniyan S, Varma VA, Teja Naidu CT. Crop yield prediction using machine learning techniques. Adv Eng Softw. 2023;175: 103326. https://doi.org/10.1016/j.advengsoft.2022.103326.
24. Islam MA, Rahman MC, Sarkar MAR, Siddique MAB. Assessing impact of BRRI released modern rice varieties adoption on farmers' welfare in bangladesh: application of panel treatment effect model. Bangladesh Rice J. 2020;23(1):1–11. https://doi.org/10.3329/brj.v23i1.46076.
25. Fan F, van der Werf W, Makowski D, Ram LJ, Huang W, Li C, Zhang C, Cong W-F, Zhang F. Cover crops promote primary crop yield in China: a meta-regression of factors affecting yield gain. Field Crops Res. 2021;271: 108237. https://doi.org/10.1016/j.fcr.2021.108237.
26. Manik MMH, Alam MT, Hossain MS. Climate change and aman rice yield nexus in the North-Western part of Bangladesh: using quantile regression. J Contemp Issues Thought. 2020;10:27–35. https://doi.org/10.37134/jcit.vol10.3.2020.
27. Joseph M, Moonsammy S, Davis H, Warner D, Adams A, Timothy OTD. Modelling climate variabilities and global rice production: a panel regression and time series analysis. Heliyon. 2023;9(4): e15480. https://doi.org/10.1016/j.heliyon.2023.e15480.
28. Wangkheimayum N, Paliwal HB. Development of rice yield forecasting model using linear regression for imphal west district, Manipur, India. Int J Environ Clim Change. 2023;13(9):485–90. https://doi.org/10.9734/ijecc/2023/v13i92258.
29. Satpathi A, Setiya P, Das B, Nain AS, Jha PK, Singh S, Singh S. Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India. Sustainability. 2023;15(3):2786. https://doi.org/10.3390/su15032786.