IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 September 2023, accepted 23 September 2023, date of publication 29 September 2023,
date of current version 9 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3321020

RESEARCH ARTICLE

Tackling Food Insecurity Using Remote Sensing and Machine Learning-Based Crop Yield Prediction

UFERAH SHAFI[1], RAFIA MUMTAZ[1], (Senior Member, IEEE), ZAHID ANWAR[2], MUHAMMAD MUZYYAB AJMAL[1], MUHAMMAD AJMAL KHAN[1], ZAHID MAHMOOD[3], MAQSOOD QAMAR[3], AND HAFIZ MUHAMMAD JHANZAB[4]

[1]School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
[2]Department of Computer Science, The Sheila and Robert Challey Institute for Global Innovation and Growth, North Dakota State University (NDSU), Fargo, ND 58102, USA
[3]Wheat Programme, Crop Sciences Institute, National Agricultural Research Centre, Islamabad 44000, Pakistan
[4]Department of Agronomy, Faculty of Agriculture, The University of Agriculture, Dera Ismail Khan, Dera Ismail Khan 29111, Pakistan

Corresponding author: Rafia Mumtaz (rafia.mumtaz@seecs.edu.pk)

This work was supported in part by the IoT Laboratory, School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan; in part by The Sheila and Robert Challey Institute of Global Innovation and Growth, North Dakota State University (NDSU), USA, in collaboration with the National Agriculture Research Centre (NARC), Islamabad; and in part by the National Center for Artificial Intelligence (NCAI), Islamabad.

ABSTRACT Precise estimation of crop yield is crucial for ensuring food security, managing the supply chain, optimally utilizing resources, promoting economic growth, enhancing climate resilience, controlling losses, and mitigating risks in the agricultural industry. Accurate yield prediction depends upon several interactive factors, including crop genotype, climate conditions, soil fertility, sowing & irrigation plan, and crop management practices. For this purpose, remote sensing data and machine learning (ML) algorithms are emerging as indispensable tools that can significantly increase farm productivity while using minimal resources and reducing environmental impact. In this context, the study presents a framework for wheat grain yield prediction using three regression techniques including Random Forest, Xtreme Gradient Boosting (XGB) regression, and Least Absolute Shrinkage & Selection Operator (LASSO) regression. Various aspects of the three models are investigated and results are compared to explore the optimal technique. Drone-based multispectral sensors are employed to acquire data from three wheat experimental fields with three different sowing dates (SD1, SD2, SD3), and the effect of the seeding plan on crop yield is examined. The prediction performance of models is assessed at different growth stages of the crop using several evaluation metrics. The results show that LASSO achieved the highest performance in April with the coefficient of determination (R^2) of 0.93 and mean absolute error (MAE) of 21.72. The average annual predicted yield is 260.54 g/m^2, 201.64 g/m^2, and 47.29 g/m^2 in the wheat field with SD1, SD2, and SD3 respectively. This study can help farmers and agronomists to make informed decisions about crop management activities such as planting & harvest plans, and resource handling.

INDEX TERMS Regression, wheat yield, remote sensing, machine learning, food security, unmanned aerial vehicle (UAV), vegetation indices (VI's).

I. INTRODUCTION

According to the World Food Programme (WFP) the number of people facing high levels of food insecurity in 2023 more than doubled the number in 2020. The war in Ukraine, supply chain disruptions, the continued economic fallout of the COVID-19 pandemic, heat waves, heavy rainfall, and droughts due to global warming are all factors pushing food prices to all-time highs. Without appropriate solutions, falling crop yields will push many people into poverty. As an

The associate editor coordinating the review of this manuscript and approving it for publication was Yang Tang.

example, approximately 43 million people in Africa alone may fall below the poverty line by 2030. Accurate and timely prediction of crop yields of large farmlands using innovative technologies such as UAV monitoring, multi-spectral sensors, satellite imagery analysis, and use of machine learning tools is a promising area of research to tackle world food insecurity. Recognizing the importance of this, the upcoming USA farm bill - a massive piece of legislation that funds agricultural programs budgeted at more than USD $1 trillion is expected to direct billions of dollars to such solutions that help farmers conserve resources, fight climate change or cope with disasters.

Wheat is the most widely grown crop in the world, owing to its vital role in global food security and contribution to the national economy of a country. For 35% of the world's population, wheat-based foods serve as their primary source of nutrition crop [1] and contribute more calories & protein to the global diet than any other grain crop. There are various factors that significantly affect the global food supply chain, such as climate change, population growth, urbanization, market trends, pandemics, regional conflicts, plant diseases, availability & management of agricultural resources, etc [2]. In this perspective, timely yield prediction of wheat yield prior to harvesting can help farmers and other stakeholders to plan and implement necessary interventions for mitigating any adverse impact and ensuring food security. For this purpose, several techniques have been developed, including process-based simulation models and data analysis-based statistical algorithms employing multi-source data [3]. Among these techniques, Machine Learning (ML) is a powerful statistical technique that delivers promising results due to its ability to autonomously learn complex relationships and solve complicated real-world problems. Random Forests (RF), Linear Regression, Least Absolute Shrinkage and Selection Operator (LASSO), K-Nearest Neighbor (KNN), Ridge Regression, Support Vector Machine (SVM), Gradient Boosting algorithms, Light Gradient Boosting (LightGBoost), Convolutional Neural Network (CNN), and Deep Neural Network (DNN) are well-known ML techniques for yield prediction [4], [5].

For the prediction of crop yield using ML techniques, data acquisition is a critical preliminary phase that substantially impacts the quality and accuracy of the prediction. In this context, remote sensing platforms are commonly employed to acquire optical, multispectral, and hyperspectral data. Analysis of this data provides useful insights about crop phonology and forms the basis for estimating crop yield [6], [7]. Commonly used remote sensing platforms include satellites, specially equipped planes, and unmanned aerial vehicles (UAVs). Each platform collects data with its own specific spatial & temporal resolution and acquisition rate [8]. Typically, the satellites provide low spatial resolution data with a fixed temporal resolution, which limits their use for certain agricultural applications. Recently, UAVs and drones have become promising substitutes for remote sensing satellites as these can collect high-resolution data with flexible timings, making these more appropriate for crop yield prediction [9] Following data collection, data pre-processing is the next critical process where the collected data is reviewed, formatted, and prepared for further analysis. It includes noise removal, dealing with the inconsistent & missing values, data augmentation & aggregation, feature selection & creation, and discretization etc [10].

The remote sensing data, containing hyperspectral and multispectral information, is used to compute different vegetation indices (VIs) which help to capture several parameters related to crop phenology and growth. These VIs are derived from the measurement of reflected solar radiations across the electromagnetic spectrum that represent specific vegetation characteristics. The most common VIs are the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Leaf Area Index (LAI), Infrared Percentage Vegetation Index (IPVI), Normalized Difference Red Edge Vegetation Index (NDRE), Normalized Difference Water Index (NDWI), Atmospherically Resistant Vegetation Index (ARVI), Wide Dynamic Range Vegetation Index (WDRVI), Green Ratio Vegetation Index (GRVI), and Green Chlorophyll Vegetation Index (GCI) [8], [11], [12]. Subsequently, the computed VIs are utilized as input for the ML algorithm to perform a particular task of interest including yield prediction. Duan et al. [13] used UAV drone imagery to compute several VIs including NDRE, NDVI, GNDVI, EVI, etc, and then employed linear regression for the estimation of rice yield. The results show that NDVI and GNDVI are the most appropriate VIs for rice yield prediction with an estimation error of less than 10%. In another study [14], UAV multispectral data is used to compute WDRVI, NDVI, and GRVI for maize yield prediction. The results indicate that WDRVI is the most relevant VI to predict maize yield with the nitrogen application of 250-300 kg/ha.

In addition to the ML techniques, the usage of deep learning (DL) techniques is also becoming increasingly popular in the agriculture sector where deep CNN and long short-term memory networks (LSTM) are commonly employed architectures. Nevaruori et al. [15] used a deep CNN model with six layers to predict the wheat yield using UAV multispectral and optical data. The model was able to accurately predict yield with a mean absolute error of 484.3 kg/ha and a mean absolute percentage error of 8.8%. Similarly, Wang et al. [16], used LSTM to predict the wheat yield using LAI where the MSE was found to be 522.3 kg/ha with a coefficient of determination (R^2) as 0.87. Cao et al. [17] explored the usage of random forest, deep neural network, LSTM, and 1D CNN for wheat yield prediction and compared results obtained from the application of ML and DL techniques. The results reveal that all aforementioned models have the predictive capability to estimate the winter wheat yield with the $R^2 \geq 0.85$ and $RMSE \leq 768$ kg/ha.

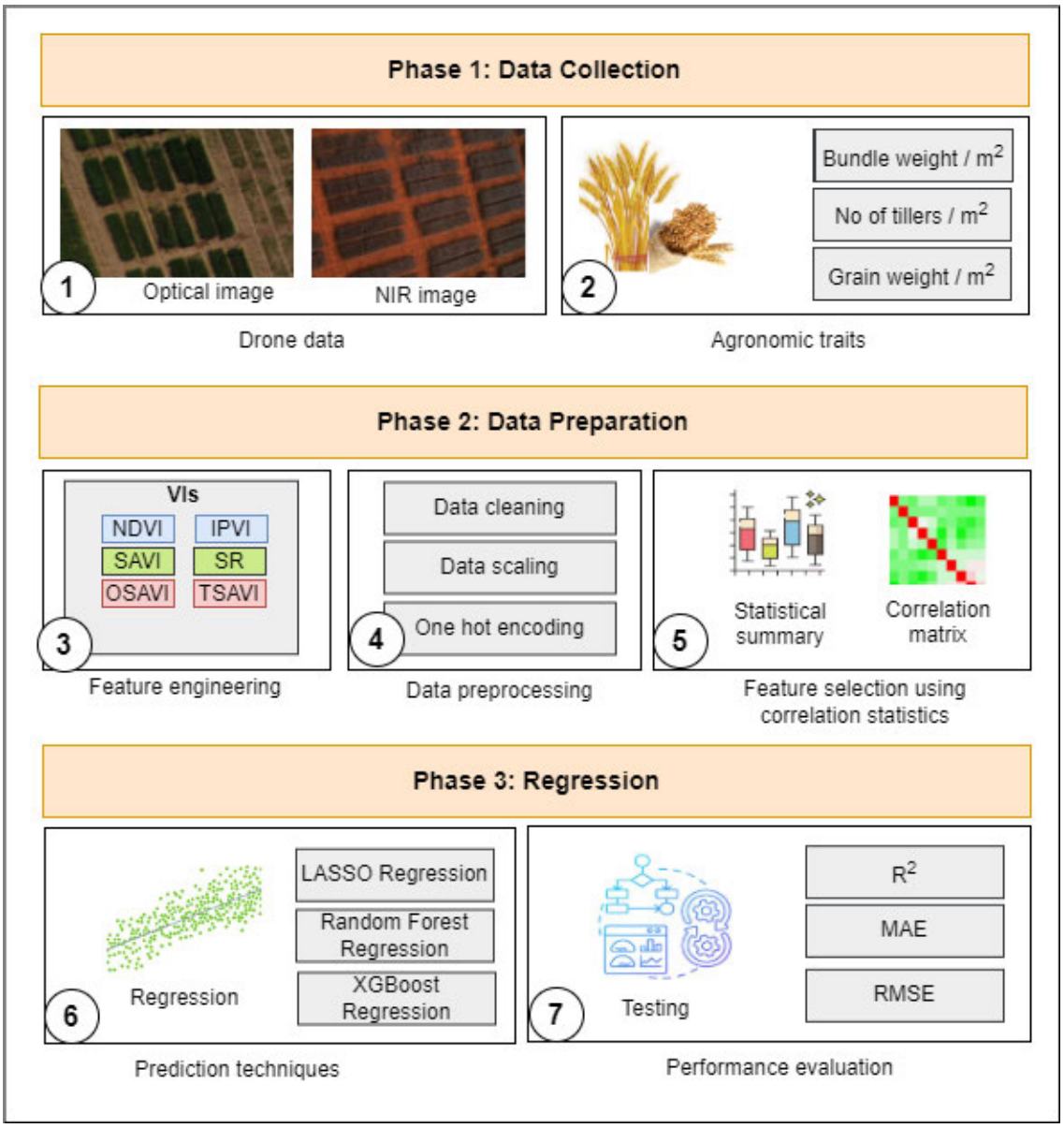


FIGURE 1. Wheat grain yield prediction workflow.

Recent studies highlight the increasing trend of utilizing data from multiple sources to enhance predictive performance. For this purpose, data from heterogeneous sources such as meteorological data, soil-related data, and remote sensing data are exploited. In [18], wheat yield is predicted using multi-source data including remote sensing data, climate data, and soil data. For this purpose, eight different ML techniques are applied to the collected data, where Gaussian process regression (GPR), SVM, and Random Forest (RF) achieved the highest performance with prediction error < 10%. Moreover, the predictive performance is evaluated in the four wheat growth stages to find the best time for predicting the crop yield. Similarly, in [19], a novel approach for crop yield prediction is presented that integrates data from several sensors (RGB, multi-spectral, and thermal infrared) installed on UAV platform which collects extensive data sets related to plant health, growth patterns, and environmental variables. Subsequently, different ML models are applied to combined data including Random Forests, NN, SVM, Cubist, and Ridge Regression for grain yield estimation. The key finding of the study is that multi-sensor data fusion-based yield prediction performed better than individual-sensor data. In [20], crop yield is predicted using MODIS data along with the twelve different climate variables using ML techniques including LASSO, SVM, NN, and FR regression. The results indicate that SVM outperformed with the R^2 of 0.79. Another study [21] presents a framework to predict crop yield using data collected from different sources including environmental data, Sentinel-2 data, and yield data. The results show that RF achieved the highest accuracy with the R^2 of 0.91.

It is observed from the literature that the utilization of data from diverse sources has become common. Moreover, the significant parameters related to crop yield are recorded during the entire growth cycle to perform yield prediction. However, a few studies have investigated the different growth stages to find the most appropriate stage for precise yield estimation. Furthermore, a notable gap exists in the literature pertaining to the influence of sowing dates on crop production and its implications for crop productivity enhancement. While various studies have examined predictive models based on comprehensive datasets, there is a limited focus on the temporal aspect of crop growth, specifically the effect of different sowing dates on subsequent yield production. This gap is particularly significant because the timing of crop sowing has a direct impact on crop development, phenology, and yield. In this context, the proposed study aims to bridge this gap by introducing a framework that incorporates heterogeneous data and predictive models, and also explicitly investigates the impact of different sowing dates on crop yield. For this purpose, multispectral images of the crop field are captured throughout the growth cycle and various VIs are computed to assess crop phenology. Additionally, field surveys are performed to record several agronomic parameters to analyze the behavior of crop growth. Subsequently, different prediction models are applied to the collected datasets and further evaluated for different growth stages to discover the time window that optimally captures the crop progression. The step-wise workflow of the proposed framework for wheat grain yield prediction is shown in Figure 1.

The objectives of this research are listed below:

1) Feature Selection for Prediction: To select the best set of predictors for enhancing the prediction performance
2) Optimal Time for Prediction: To identify the suitable time window for accurate wheat yield prediction
3) Best Regression Model Selection: To identify the most appropriate prediction model for wheat yield estimation.
4) Optimal Sowing Timing: To explore the effects of different sowing dates on crop yield and find the best time for crop sowing.

In light of these objectives, the research endeavors to provide valuable insights into the optimization of wheat yield prediction by leveraging different data sources, temporal considerations, and robust predictive modeling techniques.

II. MATERIALS AND METHODS

A. STUDY AREA AND EXPERIMENTAL DESIGN

This research is based on data collected from the wheat experimental field of the National Agricultural Research Centre (NARC), located in Islamabad, Pakistan (33.6692481° N, 73.1076928° E). The experimental field consists of three main plots where wheat is grown with three different sowing dates (SD) including (i) SD1: Nov 15, 2021, (ii) SD2: Dec 15, 2021, (iii) SD3: Jan 15, 2022. Each of these plots is further divided into three replications and each replicate contains plots of 15 different wheat varieties of area (1.5m X 6m). Wheat seeds of fifteen different varieties (V1, V2, ...V15) are planted at the rate of 112.5 g/plot in every replicate. Hence, there are 45 plots for each SD organized in three replicates and each replicate contains 15 plots corresponding to 15 varieties of wheat seed to minimize statistical error for the study. The experimental setup utilized a randomized complete block design (RCBD) as illustrated in Figure 2.

B. DATA PREPARATION

1) DATA COLLECTION

For the purpose of wheat yield estimation, data pertaining to multispectral bands and various agronomic traits are collected during the whole growth cycle of the crop. The multispectral data is captured by DJI Phantom 4 drone mounted with the Sentera multispectral imager that acquires red and near-infrared (NIR) bands. The drone is employed to collect aerial imagery of the field using a flight pattern that is fully automated and designed using customized Sentera 'FlightAgent' software. For data acquisition, multiple flights are carried out at the height of 80ft with more than 80% overlapping of ground coverage during days of clear skies and minimal wind speeds, between 10:00 am to 11:00 am local time. Drone data collection was initiated in February 2022, coinciding with the 'single shot stage' of the crop sown under SD3, and concluded during the 'ripening stage' of the crop sown under SD1. Subsequently, data was acquired via eight drone sessions on the following dates: (i) February 10, 2022, (ii) February 21, 2022, (iii) March 2, 2022, (iv) March 11, 2022, (v) March 17, 2022, (vi) March 31, 2022, (vii) April 8, 2022, (viii) April 15, 2022. After capturing multiple individual images covering the entire experimental field, the raw images are processed to generate a mosaic using WebODM which is an open-source software developed by OpenDroneMap [22]. This powerful tool is capable of generating point clouds, georeferenced models, elevation models, and 3D maps. It provides support for multiple processing engines, enhancing the efficiency of UAV and satellite image processing using Structure from Motion (SfM) and Multi-View Stereo (MVS) techniques. The software employs a web-based interface, simplifying the utilization of complex image processing algorithms. Its primary objective is to analyze extensive datasets and transform photographs into accurate georeferenced outputs. These outputs find applications in diverse fields like agriculture, urban planning, and environmental monitoring, among others.

After generating the ortho mosaic images by WebODM, the resultant images are further segmented into 135 polygonal shapes in order to extract valuable insights for the crop sown in each plot. Additionally, several ground surveys were performed during the month of March 2022 and April 2022; where the wheat crop undergoes different development stages with respect to their sowing dates. Subsequently, the parameters related to wheat yield are recorded including

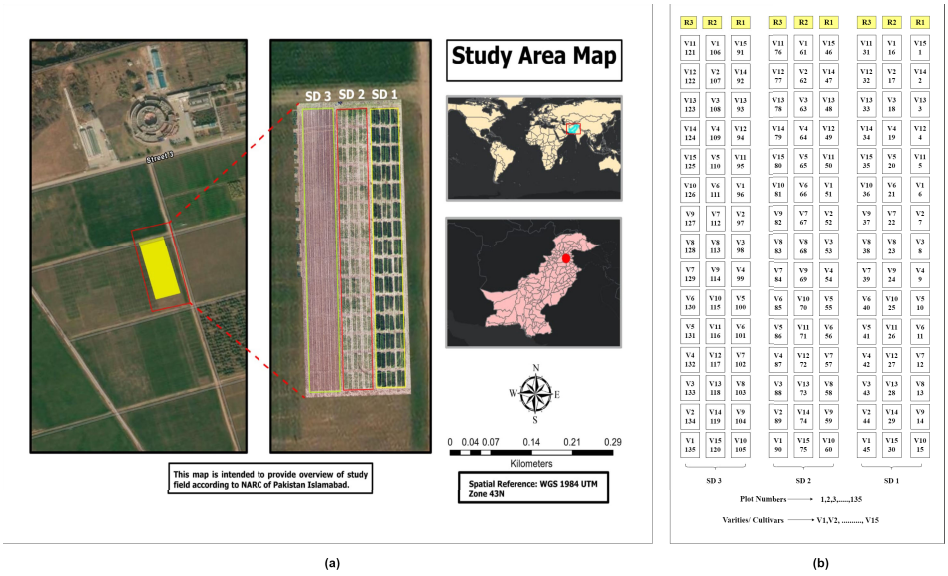


FIGURE 2. Experimental design (a) Study area, (b) Experimental field layout.

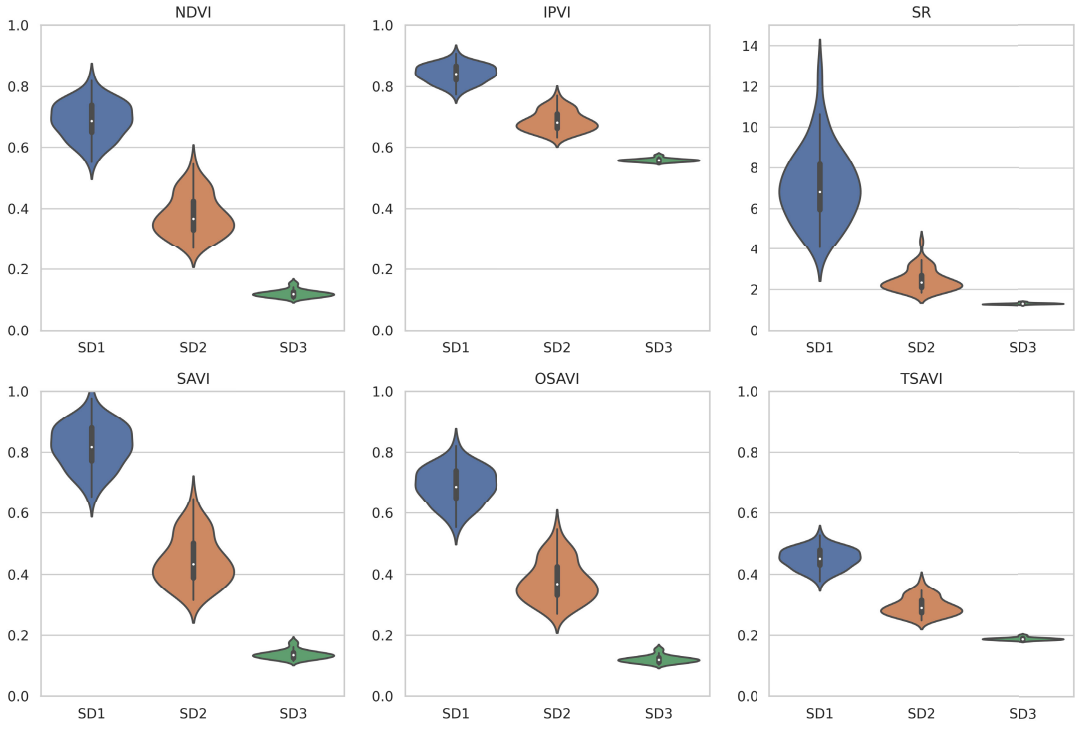


FIGURE 3. Statistical data distribution of all computed VIs corresponding to different sowing dates SD1, SD2, and SD3 in the month of February 2022.

the number of tillers/m^2, bundle weight/m^2, and grains weight/m^2, where all collected data is used to predict the wheat yield.

2) DATA PREPROCESSING

It is an essential phase that involves transforming the raw data into an appropriate format prior to the application of advanced data analysis techniques. It mainly focuses on data cleaning, feature engineering, data scaling, dealing with categorical features, data integration, and feature selection [23], [24]. In order to enhance the quality of the data for regression analysis, we performed the following preprocessing steps considering our dataset:

- Data Cleaning: It is a mandatory phase prior to applying the regression technique and leads to the high performance of the ML model. For this purpose, the

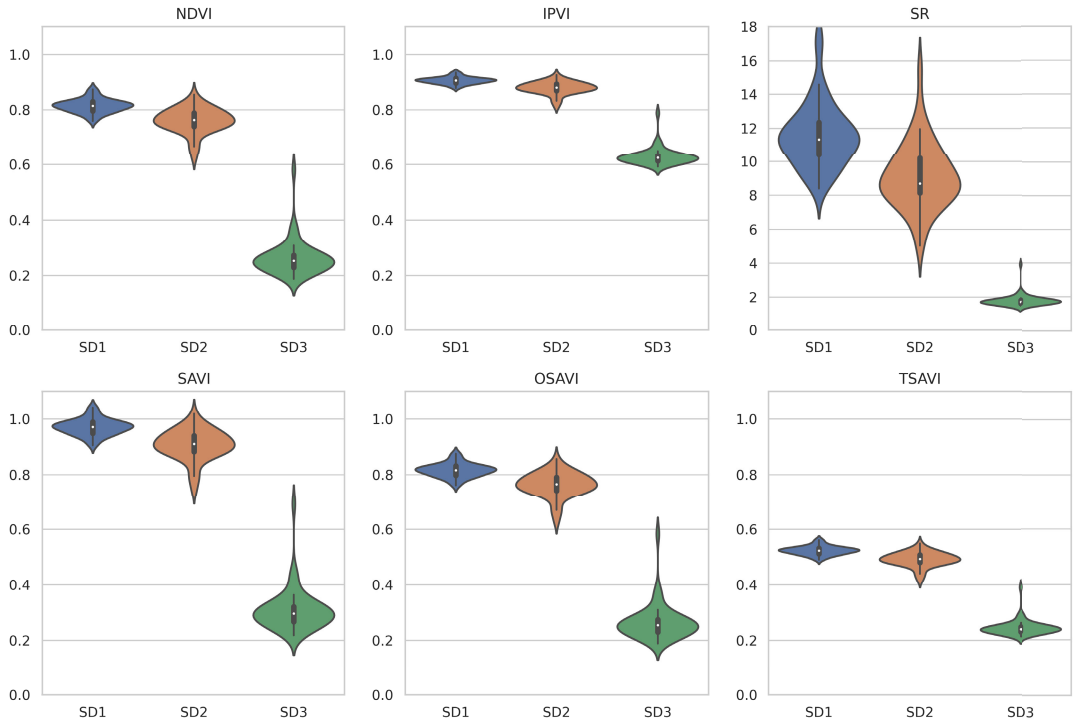


FIGURE 4. Statistical data distribution of all computed VIs corresponding to different sowing dates SD1, SD2, and SD3 in the month of March 2022.

collected data is deeply analyzed to check for outliers, missing values, noise, and inconsistencies in the data points. Observed anomalies in the dataset are removed prior to the application of ML.

- Feature Engineering: It involves the manipulation of data to extract underlying significant patterns using domain knowledge that substantially impacts the performance of ML algorithms. For this study, nine features are generated from the collected data including six VIs, crop growth stage, no. of tillers/m^2, and bundle weight/m^2. Subsequently, these features are fed into regression models to predict the grain yield.
- Data Scaling: It refers to transforming the data to fit inside a certain range to improve the effectiveness of the ML model. Later, data scaling is applied to the computed features to fit them into a specific scale of [0-1].
- One hot Encoding: It is a technique used to deal with the categorical features in the dataset as discussed in [25], [26]. In the collected feature set for wheat grain yield prediction, the growth stage is a categorical feature that has been transformed into a numerical feature using one-hot encoding.
- Feature selection: It is an important preprocessing phase that boosts the performance of the ML regression model and prevents overfitting. In this step, a subset of the most relevant features is selected from the large feature set to perform regression analysis. The feature set created for the wheat grain yield prediction contains nine features. The top 'k' most important features, out of the complete feature set, are selected by computing the correlation of each feature with the target variable [27].

C. VEGETATION INDICES (VIs)

Vegetation Indices (VIs) are derived from the measured values of surface reflectance at two or more wavelengths to emphasize a specific characteristic of vegetation. The multispectral data collected by the drone is used to compute several VIs including NDVI, SR, IPVI, SAVI, OSAVI, and TSAVI by using relationships and hyperparameters as given in Table 1.

Figure 3 shows various VIs, computed in the month of February, for the crop sown on different dates i.e., SD1, SD2, and SD3. It is evident from Figure 3 that all computed VIs have higher values in the wheat field with SD1 where the crop is in the 'stem elongation' stage. On the other hand, the crop with SD2 is in the 'tillering stage' and all VIs have slightly smaller values as compared to the early planted wheat crop with SD1. However, all VIs have very small values in the wheat field with SD3 which is still in the 'single shot' stage. It is also evident from these plots that the wheat crop with SD1 is in the '50% heading' stage, the wheat crop with SD2 is in the 'booting' stage, and the wheat crop with SD3 is in the 'stem elongation' stage. Subsequently, the variation in all computed VIs with respect to the crop growth can be visualized in Figure 4 which lists observed parameters in the month of March. Likewise, Figure 5 depicts the values of VIs

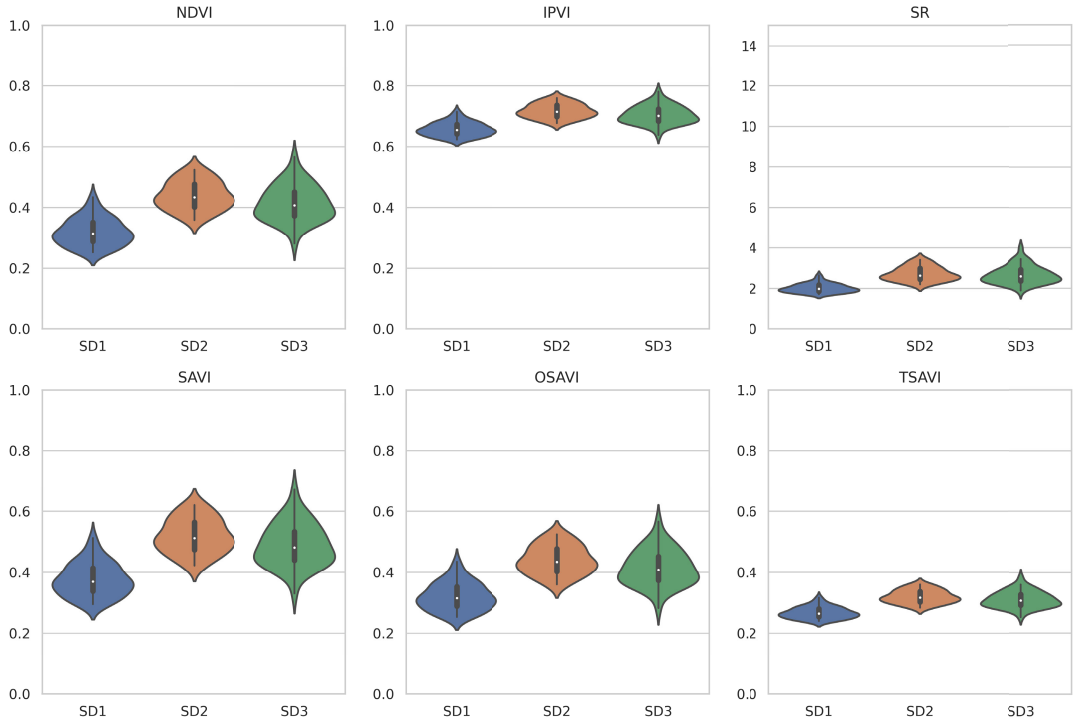


FIGURE 5. Statistical data distribution of all computed VIs corresponding to different sowing dates SD1, SD2, and SD3 in the month of April 2022.

in the month of April where the wheat crop with SD1 is in the '50% ripening' stage, the wheat crop with SD2 is in the 'milk development' stage, and wheat crop with SD3 is in the '100% heading' stage.

D. REGRESSION TECHNIQUES

After preprocessing the collected data, different regression techniques are employed to predict the grain yield of the wheat crop. These techniques are discussed in the following subsections:

- Least Absolute Shrinkage and Selection Operator (LASSO): It is a popular regression technique that uses a statistical approach to determine the linear relationship between features and the target variable [34]. To prevent overfitting and optimize feature selection, a penalty term is incorporated into the cost function that incentivizes the model to choose a subset of the most significant features. The goal of LASSO regression is to find the set of predictor variables that strongly influence the output while penalizing the magnitude of the regression coefficients to avoid overfitting. This is accomplished by employing a penalty term that represents the total of the absolute values of the regression coefficients. In this way, regression coefficients having the least relevance to output are effectively set to zero, removing the corresponding predictors from the model [35]. LASSO is particularly useful for high dimensional data having a large number of predictor variables and some of these may not be relevant for predicting the outcome variable. In this research study, the LASSO regression is applied with the regulation parameter 'alpha' set to 0.1.
- Random Forest: It is a well-known decision tree-based ensemble technique used for ML classification and regression problems [36]. The basic idea is to develop several decision trees, where each decision tree is developed utilizing a subset of features and a random sample of data. In order to predict the target value, each decision tree generates an output and the final value is evaluated by aggregating all generated outputs [37]. Random forest is considered a robust ML model that can deal with noisy data and multiple features without overfitting. To predict wheat grain yield, the random forest is applied with the number of the estimator set to 100 while the 'squared_error' function is employed for assessing the quality of the split.
- Xtreme Gradient Boosting (XGB) Regression: It is another powerful ensemble learning-based ML model employed for both regression and classification [38], [39]. This technique combines different weak models (commonly decision trees) to generate a stronger model. It iteratively builds and combines decision trees to reduce the loss function during each iteration in such a way that the new model attempts to rectify the errors of previous decision trees. In order to find the best hyperparameters of XGB, 'GridSearchCV' is used where the model is tested against several values and combinations of hyperparameters. Subsequently, the

TABLE 1. Vegetation indices.

Sr #	VI Detail	Formula
1	NDVI is used to assess the chlorophyll content in the plants [28].	$NDVI = \frac{NIR - Red}{NIR + Red}$
2	SAVI is used to assess crop vegetation by minimizing soil brightness effect [29].	$SAVI = \frac{NIR - Red}{NIR + Red + L} * (1 + L)$ where L denotes the soil brightness correction factor, L= 0.2 for this study
3	IPVI is used to assess the chlorophyll content in the plants and is functionally equivalent to NDVI; however, it is more efficient and robust [30].	$IPVI = \frac{NIR}{NIR + Red}$
4	OSAVI is used to account for NDVI's deficiencies when confronted with dense vegetation cover, fluctuating soil tint, soil moisture content, and absorption influences [31].	$OSAVI = \frac{NIR - Red}{NIR + Red + 0.16}$
5	TSAVI is another modification of SAVI to minimize the effect of soil reflectance [32].	$TSAVI = \frac{s(NIR - s * Red - a)}{(a * NIR + Red - a * s + X * (1 + s^2))}$ where s and a are, respectively, the slope and intercept of the soil line, and X is an adjustment factor to minimize soil noise. Parameter values used in this study are: s=0.33, a=0.50, and X=1.5
6	SR is the simplest VI that provides information regarding the vigor and greenness [31], [33]	$SR = \frac{NIR}{Red}$

XGB regression model with 'n_estimator' value of 100, and a 'max_depth' value of 3 is used to predict grain yield.

In order to evaluate the performance of the regression models, the following commonly employed evaluation metrics have been used [40]:

- Coefficient of determination (R^2): It is a dimensionless metric in the range from 0 to 1 that is used to assess the ability of the regression model to predict the outcome. It represents the percentage of the variance in the dependent variable (target variable) that can be predicted from the independent variables (feature variables). A higher value of the coefficient implies better prediction performance. It is computed by using Eq 1.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

where n is the number of data points, $\hat{y}_i$ is the predicted value of the dependent variable for the ith data point, y_i is the actual value of the outcome for the ith data point, and $\bar{y}$ is the mean value of the dependent variable.

- Mean Absolute Error (MEA): This metric assesses the performance of the regression model by measuring the average absolute deviation between the predicted values and the actual value. It is evaluated using the Eq 2.

$$MEA = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

where y_i is the actual value of the output variable for the ith data point, $\hat{y}_i$ is the predicted value of the output variable for the ith data point, and n is the number of observations in the dataset.

- Root Mean Square Error (RMSE): It is a commonly used statistical parameter to evaluate the prediction performance of the regression model. It is based on the square root of the average squared difference between the predicted values and the actual values of the output variable. It is calculated by using Eq 3.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (3)$$

where y_i is the actual value of the output variable for the ith data point, $\hat{y}_i$ is the predicted value of the output variable for the ith data point, and n is the number of observations in the dataset.

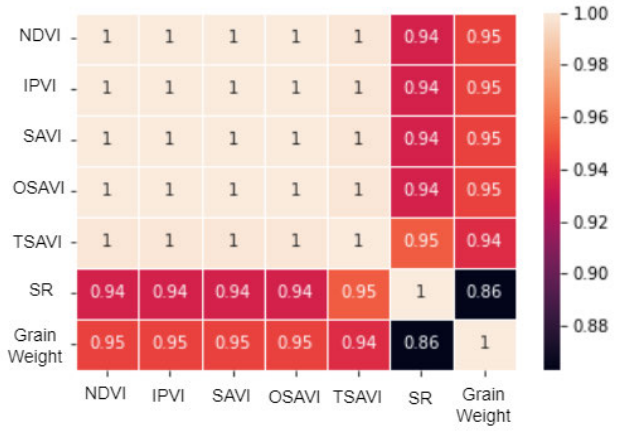


FIGURE 6. Correlation of input features with the target variable in February 2022.

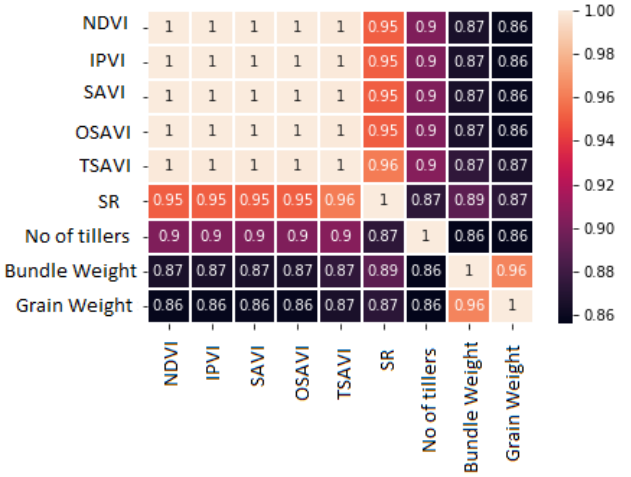


FIGURE 7. Correlation of input features with the target variable in March 2022.

III. RESULTS

In order to predict the wheat grain yield, three regression techniques are selected including random forest, LASSO, and XGB. All three techniques are commonly employed in situations having multiple features while avoiding overfitting. The collected dataset comprised nine features including the number of tillers, bundle weight, SR, SAVI, OSAVI, IPVI, NDVI, TSAVI, and growth stages. The target feature for prediction was grain weight.

To analyze the relationship between the selected feature variables with the target variable, the correlation matrices are created for the three different growth stages of the wheat crop as shown in Figure 6, 7, and 8. However, the values of correlation matrices vary according to the growth stage of the crop. In the month of February, all computed VIs are highly correlated with the target variable; whereas the values of correlation coefficients start to decrease in March and April. These variations in the values of VIs are attributed to the chlorophyll content in the vegetation which keeps on increasing until the wheat crop reaches the grain-filling stage and then starts declining for the rest of the growth cycle.

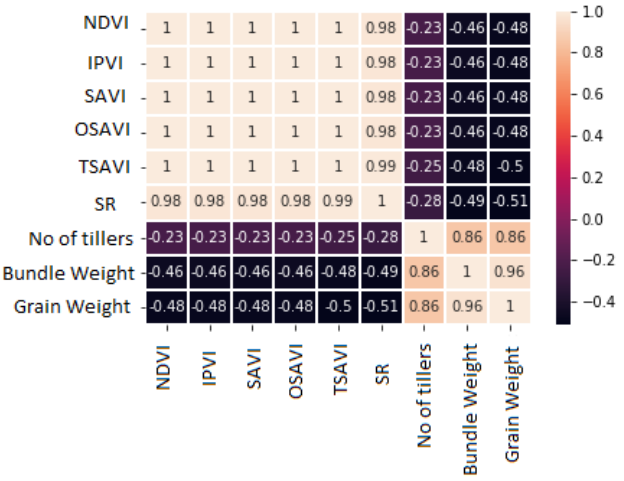


FIGURE 8. Correlation of input features with the target variable in April 2022.

It is noteworthy that the dataset collected in February does not include the parameters 'Bundle weight' and 'Number of tillers', as these particular parameters were not recorded until March 2022.

It is evident from Figure 6, 7, 8 that the feature variables have correlations with each other leading to multicollinearity which makes it difficult to determine the individual effect of the features on the target variable. To address this problem, correlation statistics are used to compute 'k' most important and relevant features. For this purpose, the correlation of each feature variable with the target variable is computed and converted to an F-value representing the feature importance score. Figure 9 shows the F-value of all numerical features in different growth stages of the wheat crop.

The features with the highest F-value are selected for regression, whereas the rest of the features are eliminated from the model to avoid overfitting. In the month of February, the selected features were 'NDVI', 'IPVI', 'SAVI', 'TSAVI', and 'OSAVI'; where the remaining features have been not considered. However, in the month of March, 'bundle weight' has the largest F-value followed by 'SR' and 'TSAVI'. The remaining VI features have an equal F-value; where 'NDVI' is selected after assessing its effect on regression. Likewise, 'Bundle weight', 'No of tillers', 'Growth stage heading complete', and 'NDVI' are the selected features for the month of April. The feature selection phase addressed the first objective of the research i.e., to select the best set of predictors for enhancing the prediction performance.

After feature selection, the regression models are applied in different growth stages of wheat crops and their performance is evaluated using three metrics: R^2, MAE, and RMSE. For this purpose, the dataset comprising all sowing dates is divided into training and testing splits with a ratio of 7:3 respectively. Table 2 shows the performance comparison of the regression models on the data collected in February 2022; whereas Figure 10 illustrates the deviation between the predicted and actual wheat grain yield on a testing split

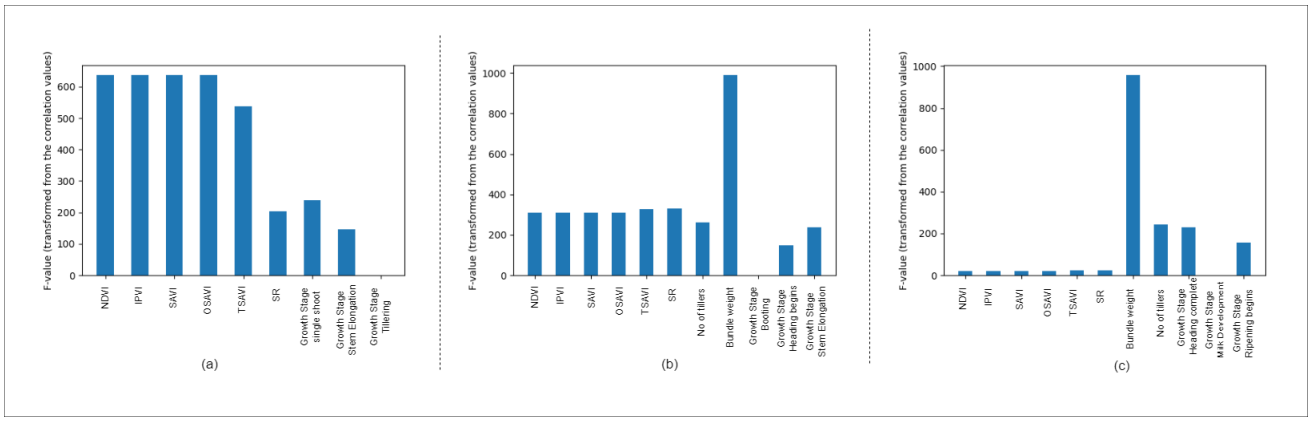


FIGURE 9. Feature selection using correlation statistics (a) February 2022 (b) March 2022 (c) April 2022.

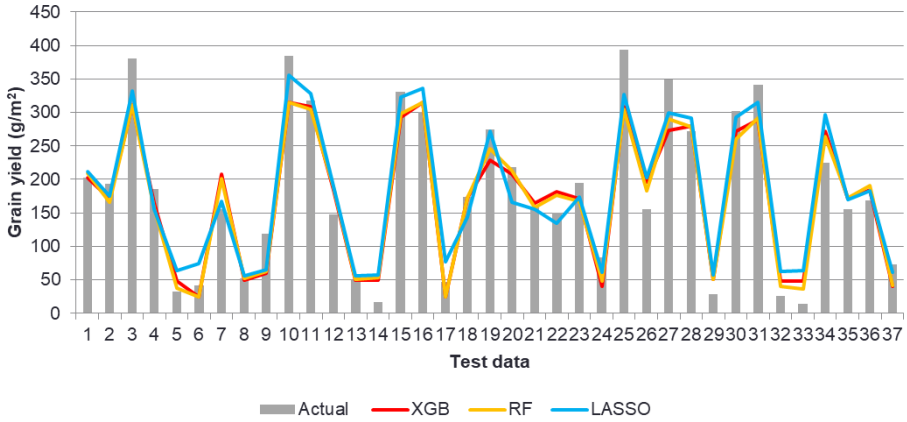


FIGURE 10. Deviation between the actual and the predicted grain yield in February 2022.

TABLE 2. Performance comparison of regression techniques applied to the data collected in February 2022.

Model	R^2	MAE	RMSE
LASSO	0.92	27.95	33.22
Random Forest	0.90	28.49	35.39
XGB	0.89	30.50	37.51

containing 37 data points. It is apparent from Table 2 that LASSO generates the best prediction results with R^2 of 0.92, MAE of 27.95 g/m^2 and RMSE 33.32 g/m^2. However, Random Forest performed better than XGB with R^2 of 0.90, MAE of 28.49 g/m^2.

Similarly, Table 3 presents the performance comparison of regression techniques applied to the dataset collected in March 2022. Best results are again generated by LASSO giving the highest R^2 of 0.93 with the MAE of 22.91 g/m^2 and RMSE of 31.06 g/m^2. Figure 11 shows the predicted grain yield versus actual grain yield against three regression techniques based on observations made in the month of March.

TABLE 3. Performance comparison of regression techniques applied to the data collected in March 2022.

Model	R^2	MAE	RMSE
LASSO	0.93	22.91	31.06
Random Forest	0.89	25.90	37.96
XGB	0.90	26.43	36.9

TABLE 4. Performance comparison of regression techniques applied to the data collected in April 2022.

Model	R^2	MAE	RMSE
LASSO	0.93	21.72	29.32
Random Forest	0.92	23.96	30.77
XGB	0.92	23.38	31.64

Table 4 compares the results of three regression models on the basis of data collected in April 2022. The predicted grain yield for the various regression techniques versus actual yield is shown in Figure 12. These results clearly demonstrate that

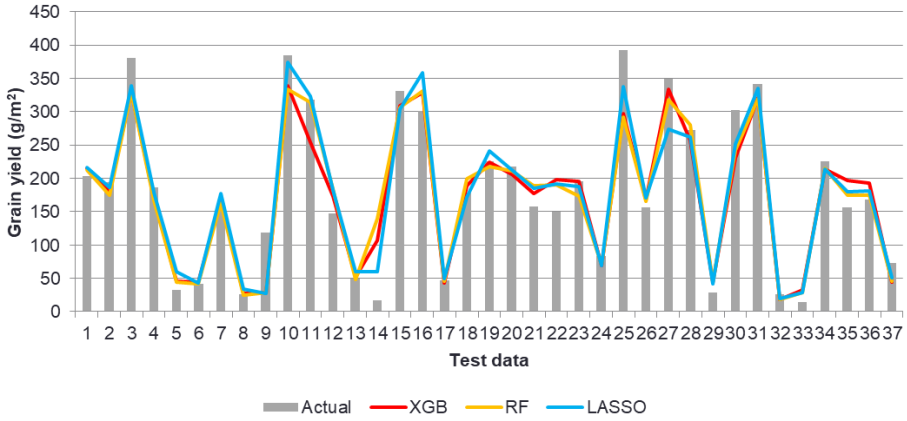


FIGURE 11. Deviation between the actual and the predicted grain yield in March 2022.

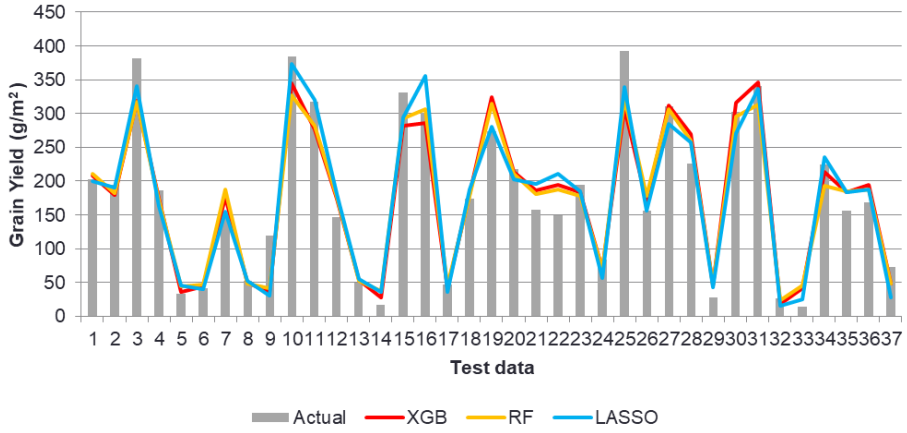


FIGURE 12. Deviation between the actual and the predicted grain yield in April 2022.

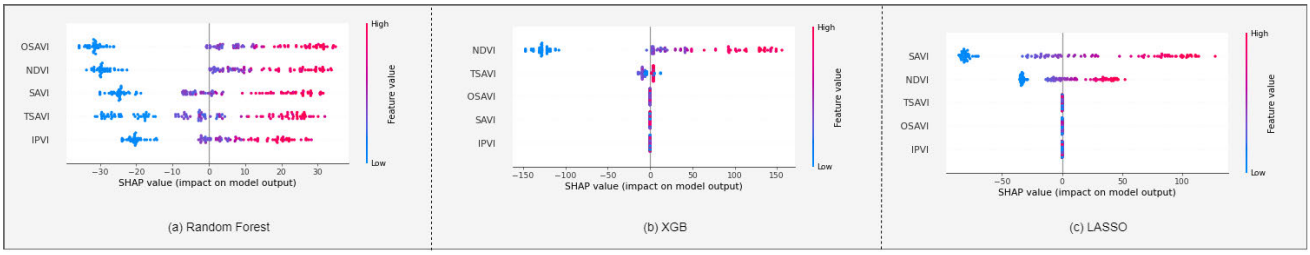


FIGURE 13. Contribution of each feature using SHAP value on data collected in February 2022.

LASSO once again performed better than Random Forest and XGB for the dataset from the month of April.

In order to evaluate the influence of individual features on regression performance, the widely used SHAP (Shapley Additive explanations) method is employed [41]. The SHAP feature importance graphs provide a comprehensive insight into the model's interpretability and sensitivity to individual features and their contributions to predictions. To quantify the performance of each feature on the prediction model, all possible combinations of features are considered, and computing the difference in predictions when a particular feature is included versus when it is excluded. This difference served as SHAP value which signifies the degree of influence a feature wields on a prediction in comparison to its absence. Figure 13, 14, and 15 illustrate SHAP feature importance graphs that are computed on the datasets collected in February 2022, March 2022, and April 2022.

It is evident from the SHAP feature importance graphs that all selected features demonstrate roughly equal contributions in the case of Random forest on the dataset collected in February 2022. However, 'NDVI', and 'SAVI' have a significant contribution in the case of LASSO regression; whereas 'NDVI' has a dominant contribution in the case

FIGURE 14. Contribution of each feature using SHAP value on data collected in March 2022.

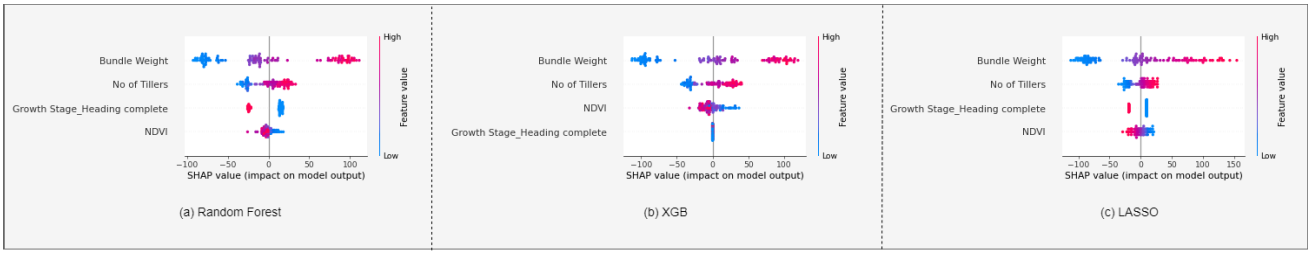


FIGURE 15. Contribution of each feature using SHAP value on data collected in April 2022.

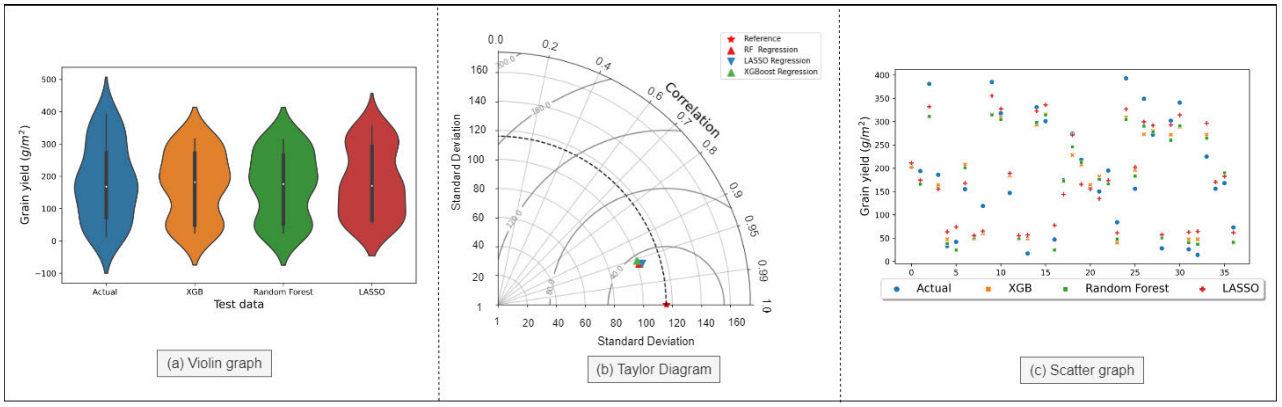


FIGURE 16. Performance comparison of regression techniques in February 2022 using (a) Violin graph showing the statistical summary of predicted and actual wheat grain yield, (b) Taylor diagram exhibiting the model performance in terms of standard deviation ratio, correlation, and centered root-mean-square error from the reference dataset and (c) scatter plot illustrating the difference in actual and predicted values of wheat grain yield on test dataset.

of XGB regression. Conversely, a distinct pattern emerges with the data collected in March 2022 and April 2022. where 'Bundle weight' has more contribution than any other feature. Similarly, the comprehensive visual representation of each feature's contribution across all regression techniques can be observed in Figures 13, 14, and 15. In conclusion, the analysis of these graphs offers valuable insights into the influential dynamics of specific features in shaping the model's outcome. Notably, changes in features with high positive SHAP values (indicated by red color code) can lead to proportionate shifts in predictions, while those with negative values(indicated by blue color code) might cause counteractive shifts. This profound understanding of feature importance not only enhances interpretability but also sheds light on the intricate relationships between input variables and outcomes.

Furthermore, to conduct an in-depth analysis of the performance exhibited by various regression techniques, different graphs have been generated including a violin graph, Taylor diagram, and scatter plot as shown in Figure 16, 17, and 18. The violin graphs provide a statistical summary of actual versus predicted wheat grain yield by different regression techniques. Whereas, the scatter plots illustrate the difference in actual versus predicted wheat grain yield by plotting test data. However, Taylor diagrams provide deeper insights into the model performance in terms of standard deviation ratio, correlation, and centered root-mean-square error from the reference dataset [42]. These graphs offer a holistic view of how well each regression technique matches up against the true data, allowing for a deeper understanding of their relative performance. It is evident from these graphs that the regression results are pretty good on the dataset collected in April 2022 as compared to February 2022 and March 2022. Furthermore, it is notable that the LASSO regression technique produces prediction results that more closely approximate the actual dataset.

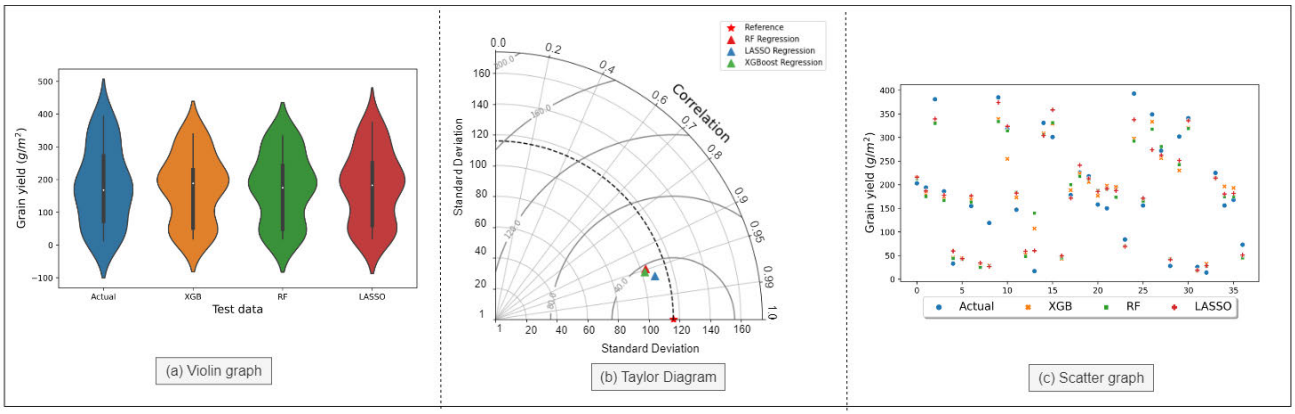


FIGURE 17. Performance comparison of regression techniques in March 2022 using (a) Violin graph showing the statistical summary of predicted and actual wheat grain yield, (b) Taylor diagram exhibiting the model performance in terms of standard deviation ratio, correlation, and centered root-mean-square error from the reference dataset and (c) scatter plot illustrating the difference in actual and predicted values of wheat grain yield on test dataset.

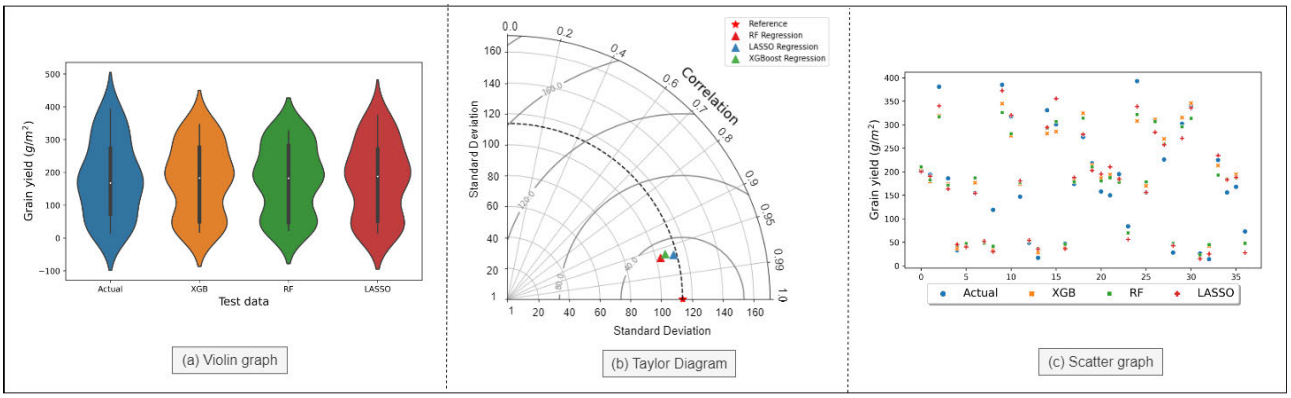


FIGURE 18. Performance comparison of regression techniques in April 2022 using (a) Violin graph showing the statistical summary of predicted and actual wheat grain yield, (b) Taylor diagram exhibiting the model performance in terms of standard deviation ratio, correlation, and centered root-mean-square error from the reference dataset and (c) scatter plot illustrating the difference in actual and predicted values of wheat grain yield on test dataset.

TABLE 5. Comparison of average predicted and actual wheat grain yield in the wheat field with different sowing dates (SD1, SD2, and SD3).

SD	Predicted yield g/m^2	Actual yield g/m^2	Difference g/m^2
SD1	260.54	292.23	31.69
SD2	201.64	180.0	21.64
SD3	47.29	58.25	10.96

It can be observed from the regression results that minimum MAE and RMSE are achieved by LASSO in the month of April 2022. This addresses the second research objective of this study i.e., to identify the suitable time window for accurate wheat yield prediction.

The results from Tables 2, 3 and 4 show that LASSO achieved the best performance due to its ability to deal with high dimensional data and avoid overfitting. This addressed the third objective of the research i.e., to identify the most appropriate prediction model for wheat yield estimation.

A comparison of an average wheat grain yield versus predicted yield for wheat fields with different sowing dates SD1, SD2, and SD3 is given in Table 5. It can be clearly seen that the highest average grain yield is achieved from the wheat field where the crop was sown on 15 November 2021 (SD1). This finding addressed the fourth objective of this research study i.e., to explore the effects of different sowing dates on the crop yield for identifying the best crop sowing time.

In order to further analyze the growth behavior of the crop with different sowing dates, NDVI maps of the wheat field are developed as shown in Figure 19. The maximum greenness is observed in the month of March 2022 in the wheat field with SD1 which depicts the ideal growth behavior for the wheat crop. On the other hand, less vegetation is observed in the wheat field with SD2, and minimal vegetation is seen in the field having SD3. This is due to the fact that wheat, like any other crop, requires a specific temperature profile for its optimal growth. The wheat fields with SD2 and SD3 are sown in the months of December and January respectively. For wheat sown during these months, the temperature profiles do not match the optimal values required for crop growth,

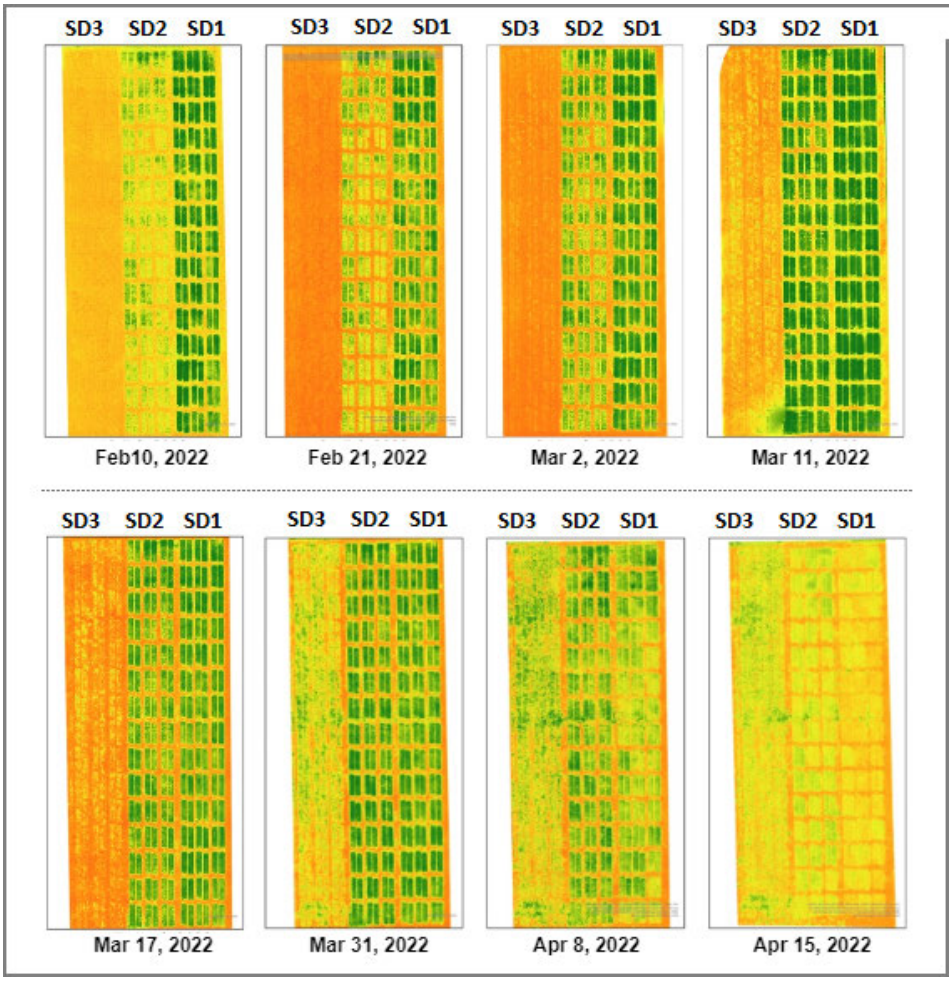


FIGURE 19. NDVI profiles of wheat crop in SD1, SD2, and SD3.

resulting in reduced grain yield. The correspondingly highest yield is obtained for SD1 and wheat production reduces as sowing is delayed beyond the optimal sowing date.

It is worth mentioning that fifteen different wheat varieties were sown on three different dates in the experimental fields. It has been concluded that the highest average yield was obtained for SD1 and production dropped as sowing was delayed. However, the yield was found to vary for different varieties even for the same sowing date. Figures 20-22 show the wheat grain yield obtained by fifteen different wheat varieties on three different sowing dates (SD1, SD2, and SD3).

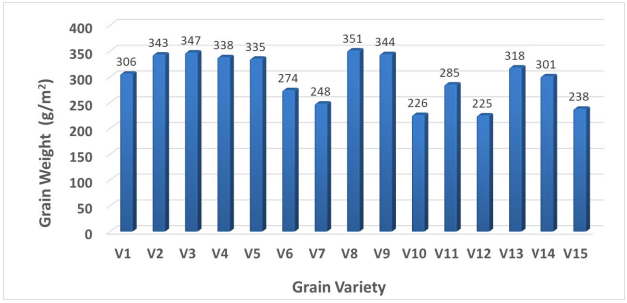


FIGURE 20. Grain yield of different varieties in a wheat field with sowing date SD1.

A. EFFECT OF DIFFERENT SOWING DATES, CLIMATE VARIATIONS AND GENOTYPES ON CROP GROWTH AND YIELD

Varying sowing dates have a great impact on wheat yield. For this purpose, an investigation is carried out to check the response of wheat crop growth and yield of different advanced lines under three diverse sowing dates. The maximum average annual yield is recorded with the sowing date SD1 (292.25 g/m^2), followed by sowing date SD2 (180.0 g/m^2), and the minimum is recorded with the sowing date SD3 (58.25 g/m^2) respectively (Table 5). Among different genotypes (varieties), the maximum yield is recorded with V8 (351 g/m2), followed by V3 (347 g/m2) in the wheat crop with sowing date SD1. Whereas, the minimum yield is recorded with V12 (225 g/m^2) respectively as shown in Figure 20. In the wheat field with sowing date SD2, the highest yield is recorded with V13 (218 g/m^2) followed by V6 (210 g/m^2), and the lowest response is recorded

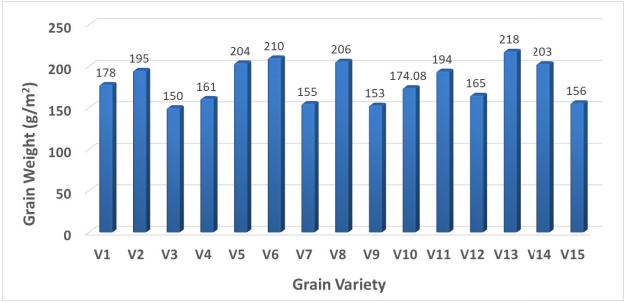


FIGURE 21. Grain yield of different varieties in a wheat field with sowing date SD2.

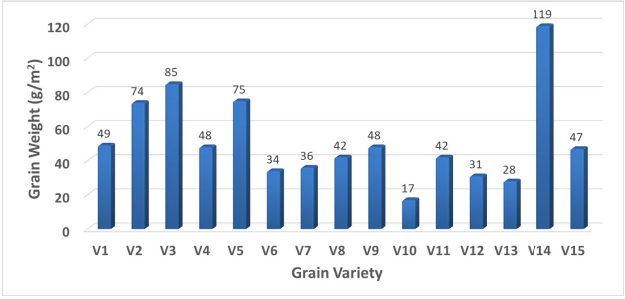


FIGURE 22. Grain yield of different varieties in a wheat field with sowing date SD3.

with V9 (153 g/m^2) as depicted in Figure 21. Similarly, the response of genotypes sown on sowing date SD3 is different in comparison to SD1 and SD2, where the highest yield is recorded with V14 (119 g/m^2), followed by V3 (85 g/m2), and the minimum yield is recorded with V10 (17 g/m^2) as illustrated in Figure 22. It is concluded that the maximum yield is obtained with SD1 as compared with the wheat crops sown in succeeding sowing dates. However, due to the genetic potential of different genotypes, a significant variation is observed in the wheat yield response under different sowing dates.

IV. DISCUSSION

The crop yield prediction holds significant importance in optimizing agricultural resources and boosting overall productivity. To this end, a field experiment is presented to predict wheat grain yield; where different regression techniques have been investigated including LASSO, Random Forest, and XGB regression. For this purpose, the multispectral data is collected by drone in different crop growth stages along with different agronomic traits. Moreover, the effect of different genotypes and the sowing plan on wheat growth and its yield is analyzed by sowing the crop on three different sowing dates. Subsequently, three regression techniques are applied to three different datasets collected in February 2022, March 2022, and April 2022 to determine the best time window to accurately estimate the wheat crop yield.

The results revealed that the best results for wheat grain prediction were observed in the month of April, where LASSO outperformed XGB and Random Forest with the minimum difference between the actual and predicted yield (31.69 g/m^2, 21.64 g/m^2, and 10.96 g/m^2 for SD1, SD2 and SD3 respectively). Figure 16, 17, and 18 illustrate the performance comparison among regression techniques using a violin graph, scatter plot, and Taylor diagram. It is clearly evident that LASSO provides a more accurate estimation of wheat grain yield as compared to XGB and Random Forest. The effectiveness of LASSO in controlling overfitting with limited data, its ability to provide sparse solutions, and its interpretability make it particularly well-suited for addressing this specific problem. In contrast, the Random Forest and XGBoost algorithms, while powerful and capable of handling complex relationships within data, might struggle with limited data. These ensemble methods inherently rely on aggregating multiple decision trees, and their performance typically improves with larger datasets. With a small amount of data, there's a higher risk of overfitting due to the complexity of these models. Additionally, tuning the hyperparameters of these algorithms becomes crucial, and without sufficient data, finding optimal hyperparameters can be challenging, leading to suboptimal performance.

It is observed from the results that the optimal time for sowing the wheat crop is November (SD1); where the maximum grain yield of 351 g/m^2 has been recorded. Whereas, the minimum yield of 17 g/m^2 has been observed for the crop that was sown lately in January (SD3) due to a reduction in the length of the growing season. The suitable time of sowing is imperative to achieve the maximum yield on a sustainable basis because wheat production is highly sensitive to elevated temperatures. Erratic climate has influenced the optimum time of wheat sowing and grain production by variations in temperature during the growth period of the crop. The process of plant development accelerates due to elevation in temperature; however growth parameters reduced such as leaf area, tillers, and length of the spikes which results in a significant reduction of yield [43], [44]. Late sowing seriously affects germination, growth rate, grain development, and reduced tillering in low temperatures and ultimately concealed yield [45]. Similarly, elevation in temperature during vegetative and reproductive growth stages badly affects the emergence of plants and succeeding crop growth stages [46]. For this purpose, an optimum and appropriate environment results in a higher economic yield which aids genotypes to express their full growth potential. Wheat, as a cereal, requires specific environmental conditions for improved growth and production [43] and is vulnerable if exposed to high temperatures through the reproductive phase at grain formation [47]. The favorable temperature that is essential for the anthesis & grain filling phase of wheat ranges from (12 °C - to 22 °C). High temperature accelerates the process of development of grain filling [48], thus resulting in a reduction of assimilation of carbohydrates, deposition of starch in grains, and yield of grains [49]. With the management of the sowing date, potential variety and environmental factors production of

wheat can be increased by 10-80% [50]. Whereas, late planting affects germination, growth, and development of grains and produces poor tillers due to winter injury in low-temperature [45], [51]. Therefore, it is very necessary to find the relationship between varying environments and newly developed genotypes. An appropriate sowing time for wheat plays a significant role in growth and development. However, in varying climatic conditions of Pakistan, it is estimated that yield may be decreased by 58.2 % in delayed sowing practice [52]. The precise and exact information of sowing time of specific variety at a particular location is crucial for meeting the potential yield of grains as discussed in [53].

Optimum environmental conditions are prerequisites for attaining the maximum yield. It has been found from research that each variety has its specific requirements of temperature and light for flowering and development of grains [54], [55]. However, the emergence and number of days to earing for the crop with sowing date SD1 decreased with delayed planting to sowing date SD2. Cultivation of wheat under late sowing results in a reduction of air and soil temperature causing a decrease in the emergence and crop stand establishment [56]. It has been reported in various studies that elevated temperature affects the emergence of crops [57]. Late planted crops decreased no. of tillers due to high temperature during the growth stage of tillering [58] and also decreased the duration of grain filling at the reproductive stage leading to a reduction in enzyme activity and yield of crop [45], [59]. Maximum yield can be obtained when the crop is sown earlier as it received extensive duration of grain filling in comparison with late sowing caused warmer environment.

The results obtained from our analysis provide a foundation for practical implementations in the agricultural domain. The benefits of this research lie in its potential to equip farmers and agronomists with valuable insights that can drive more efficient and productive agricultural practices. The farmers could make decisions related to crop management, resource allocation, and harvest planning. Moreover, optimizing crop yield predictions can lead to more efficient resource utilization and improved crop planning, contributing to increased profitability. By addressing crucial gaps and leveraging data-driven insights, this study has the capacity to catalyze positive transformations within the agricultural landscape. that can drive more efficient and productive agricultural practices.

V. CONCLUSION AND FUTURE WORK

Accurate and timely yield prediction of wheat crops is essential for global food security. Towards this end, a framework for wheat grain yield prediction is presented in this research study. Multispectral data spanning the crop growth cycle from three experimental fields, each planted with the wheat crop at different sowing dates, is collected using drone-based sensors. Following the preprocessing of datasets, the most relevant predictors are identified and three well-known ML regression models including Random Forest, XGB regression, and LASSO regression are employed to estimate crop yield. The results show that LASSO achieved the best prediction performance with R^2 of 0.93, and MAE of 21.72 g/m^2. The annual predicted yield is found to be 260.54 g/m^2, 201.64 g/m^2 and 47.29 g/m^2 for the crop sown in November (SD1), December (SD2) and January (SD3) respectively. Additionally, the best prediction results are obtained from the observations made in the month of April. This research will help farmers and agronomists to timely and accurately estimate crop yields and manage crop resources prior to harvesting.

At present, the estimation of wheat grain yield is accomplished through the use of multispectral data and machine learning techniques. However, in the future, we plan to explore deep learning techniques like CNN, LSTM, etc., to analyze drone optical data for crop yield forecasting. In addition, we plan to integrate more predictors like soil and climate data for enhancing the accuracy of yield estimation.

REFERENCES

[1] U. Grote, A. Fasse, T. T. Nguyen, and O. Erenstein, "Food security and the dynamics of wheat and maize value chains in Africa and Asia," *Frontiers Sustain. Food Syst.*, vol. 4, pp. 1–17, Feb. 2021.

[2] M. J. Puma, S. Bose, S. Y. Chon, and B. I. Cook, "Assessing the evolving fragility of the global food system," *Environ. Res. Lett.*, vol. 10, no. 2, Feb. 2015, Art. no. 024007.

[3] A. K. Srivastava, N. Safaei, S. Khaki, G. Lopez, W. Zeng, F. Ewert, T. Gaiser, and J. Rahimi, "Winter wheat yield prediction using convolutional neural networks from environmental and phenological data," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, Feb. 2022.

[4] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105709.

[5] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, "Maize yield and nitrate loss prediction with machine learning algorithms," *Environ. Res. Lett.*, vol. 14, no. 12, Dec. 2019, Art. no. 124026.

[6] S. J. Maas, "Using satellite data to improve model estimates of crop yield," *Agronomy J.*, vol. 80, no. 4, pp. 655–662, Jul. 1988.

[7] S. Yang, L. Hu, H. Wu, H. Ren, H. Qiao, P. Li, and W. Fan, "Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6253–6269, 2021.

[8] U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, and N. Iqbal, "Precision agriculture techniques and practices: From considerations to applications," *Sensors*, vol. 19, no. 17, p. 3796, 2019.

[9] B. Duan, S. Fang, R. Zhu, X. Wu, S. Wang, Y. Gong, and Y. Peng, "Remote estimation of rice yield with unmanned aerial vehicle (UAV) data and spectral mixture analysis," *Frontiers Plant Sci.*, vol. 10, p. 204, Feb. 2019.

[10] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. San Francisco, CA, USA: Elsevier, 2006.

[11] J. Xue and B. Su, "Significant remote sensing vegetation indices: A review of developments and applications," *J. Sensors*, vol. 2017, pp. 1–17, May 2017.

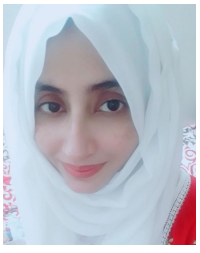
[12] A. Bannari, D. Morin, F. Bonn, and A. R. Huete, "A review of vegetation indices," *Remote Sens. Rev.*, vol. 13, nos. 1–2, pp. 95–120, Aug. 1995.

[13] B. Duan, S. Fang, R. Zhu, X. Wu, S. Wang, Y. Gong, and Y. Peng, "Remote estimation of rice yield with unmanned aerial vehicle (UAV) data and spectral mixture analysis," *Frontiers Plant Sci.*, vol. 10, p. 204, Feb. 2019.

[14] Á. Maresma, M. Ariza, E. Martínez, J. Lloveras, and J. A. Martínez-Casasnovas, "Analysis of vegetation indices to determine nitrogen application and yield prediction in maize (*Zea mays L.*) from a standard UAV service," *Remote Sens.*, vol. 8, no. 12, p. 973, 2016.

[15] P. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104859.

[16] J. Wang, H. Si, Z. Gao, and L. Shi, "Winter wheat yield prediction using an LSTM model from MODIS LAI products," *Agriculture*, vol. 12, no. 10, p. 1707, Oct. 2022.
[17] J. Cao, Z. Zhang, Y. Luo, L. Zhang, J. Zhang, Z. Li, and F. Tao, "Wheat yield predictions at a county and field scale with deep learning, machine learning, and Google Earth engine," *Eur. J. Agronomy*, vol. 123, Feb. 2021, Art. no. 126204.
[18] J. Han, Z. Zhang, J. Cao, Y. Luo, L. Zhang, Z. Li, and J. Zhang, "Prediction of winter wheat yield based on multi-source data and machine learning in China," *Remote Sens.*, vol. 12, no. 2, p. 236, Jan. 2020.
[19] S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, and Y. Ma, "UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," *Precis. Agricult.*, vol. 24, no. 1, pp. 187–212, Feb. 2023.
[20] Y. Cai, K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang, L. You, and B. Peng, "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," *Agricult. Forest Meteorol.*, vol. 274, pp. 144–159, Aug. 2019.
[21] M. L. Hunt, G. A. Blackburn, L. Carrasco, J. W. Redhead, and C. S. Rowland, "High resolution wheat yield mapping using Sentinel-2," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111410.
[22] *WebODM-Drone Mapping Software*. Accessed: Jun. 2022. [Online]. Available: http://opendronemap.org/webodm//
[23] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Elsevier, 2012.
[24] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, vol. 12. New York, NY, USA: Springer, 2011.
[25] C. Seger, "An investigation of categorical variable encoding techniques in machine learning: Binary versus one-hot and feature hashing," School Elect. Eng. Comput. Sci. (EECS), Sweden, Tech. Rep. TRITA-EECS-EX, 2018, p. 596.
[26] I. U. Haq, I. Gondal, P. Vamplew, and S. Brown, "Categorical features transformation with compact one-hot encoder for fraud detection in distributed environment," in *Proc. 16th Australas. Conf. Data Mining (AusDM)*, Bahrurst, NSW, Australia. Singapore: Springer, 2019, pp. 69–80.
[27] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.
[28] J. W. Rouse Jr, R. H. Haas, J. Schell, and D. Deering, "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation," NASA, Washington, DC, USA, Tech. Rep. NASA-CR-132982, 1973.
[29] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, Aug. 1988.
[30] R. Crippen, "Calculating the vegetation index faster," *Remote Sens. Environ.*, vol. 34, no. 1, pp. 71–73, Oct. 1990.
[31] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote Sens. Environ.*, vol. 55, no. 2, pp. 95–107, Feb. 1996.
[32] F. Baret, "TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation," in *Proc. 12th Can. Symp. Remote Sens. (IGARSS)*, Vancouver, BC, Canada, Jul. 1989, pp. 10–14.
[33] A. Erener, "Remote sensing of vegetation health for reclaimed areas of Seyitömer open cast coal mine," *Int. J. Coal Geol.*, vol. 86, no. 1, pp. 20–26, 2011.
[34] J. H. Lee, Z. Shi, and Z. Gao, "On LASSO for predictive regression," *J. Econometrics*, vol. 229, no. 2, pp. 322–349, Aug. 2022.
[35] J. Ranstam and J. Cook, "Lasso regression," *J. Brit. Surg.*, vol. 105, no. 10, p. 1348, 2018.
[36] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. New York, NY, USA: Springer, 2009.
[37] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*. New York, NY, USA: Springer, 2012.
[38] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[39] C. Wade and K. Glynn, *Hands-On Gradient Boosting With XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting With Python*. Birmingham, U.K.: Packt, 2020.
[40] H. Jiawei, K. Micheline, and P. Jian, *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2012.
[41] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values," *J. Theor. Appl. Electron. Commerce Res.*, vol. 16, no. 3, pp. 466–490, Nov. 2020.
[42] K. E. Taylor, "Taylor diagram primer," Citeseer, NJ, USA, Work. Paper, 2005, pp. 1–4.
[43] W. Dabre, S. Lall, and G. Ingole, "Effects of sowing dates on yield, ear number, stomatal frequency and stomatal index in wheat," *J.-Maharashtra Agricult. Univ.*, vol. 18, p. 64, 1993.
[44] M. Mullarkey and P. Jones, "Isolation and analysis of thermotolerant mutants of wheat," *J. Exp. Botany*, vol. 51, no. 342, pp. 139–146, Jan. 2000.
[45] M. Tahir, A. Ali, M. A. Nadeem, A. Hussain, and F. Khalid, "Effect of different sowing dates on growth and yield of wheat (Triticum aestivum L.) varieties in district Jhang, Pakistan," *Pakistan J. Life Soc. Sci.*, vol. 7, no. 1, pp. 66–69, 2009.
[46] S. K. Dwivedi, S. Kumar, and V. Prakash, "Effect of late sowing on yield and yield attributes of wheat genotypes in EIGP," *J. AgriSearch*, vol. 2, no. 4, pp. 304–306, 2015.
[47] N. Kalra et al., "Effect of increasing temperature on yield of some winter crops in northwest India," *Current Sci.*, vol. 94, pp. 82–88, Jan. 2008.
[48] O. Gaju, M. P. Reynolds, D. L. Sparkes, and M. J. Foulkes, "Relationships between large-spike phenotype, grain number, and yield potential in spring wheat," *Crop Sci.*, vol. 49, no. 3, pp. 961–973, May 2009.
[49] M. T. Labuschagne, O. Elago, and E. Koen, "The influence of temperature extremes on some quality and starch characteristics in bread, biscuit and durum wheat," *J. Cereal Sci.*, vol. 49, no. 2, pp. 184–189, Mar. 2009.
[50] D. R. Coventry, R. K. Gupta, A. Yadav, R. S. Poswal, R. S. Chhokar, R. K. Sharma, V. K. Yadav, S. C. Gill, A. Kumar, A. Mehta, S. G. L. Kleemann, A. Bonamano, and J. A. Cummins, "Wheat quality and productivity as affected by varieties and sowing time in haryana, India," *Field Crops Res.*, vol. 123, no. 3, pp. 214–225, Sep. 2011.
[51] M. Khan, "Effects of planting date, chlorotoluron+MCPA and wheat varieties on weed control and wheat yield," *Sarhad J. Agricult., Pakistan*, vol. 18, no. 2, pp. 443–447, 2002.
[52] M. A. Ali, M. Ali, and Q. Din, "Determination of grain yield of different wheat varieties as influenced by planting dates in agro-ecological conditions of Vehari," *Pakistan J. Life Soc. Sci.*, vol. 2, no. 1, pp. 5–8, 2004.
[53] J. I. Ortiz-Monasterio R., S. S. Dhillon, and R. A. Fischer, "Date of sowing effects on grain yield and yield components of irrigated spring wheat cultivars and relationships with radiation and temperature in ludhiana, India," *Field Crops Res.*, vol. 37, no. 3, pp. 169–184, Jun. 1994.
[54] S. Haider, "Growth analysis in relation to sowing dates in four varieties of wheat: A functional approach," *J. Life Earth Sci.*, vol. 2, no. 2, pp. 17–25, Jan. 1970.
[55] F. Aslani and M. R. Mehrvar, "Responses of wheat genotypes as affected by different sowing dates," *Asian J. Agricult. Sci.*, vol. 4, no. 1, pp. 72–74, 2012.
[56] Y. W. Jame and H. W. Cutforth, "Simulating the effects of temperature and seeding depth on germination and emergence of spring wheat," *Agricult. Forest Meteorol.*, vol. 124, nos. 3–4, pp. 207–218, Aug. 2004.
[57] M. A. Shah, M. Farooq, M. Shahzad, M. B. Khan, and M. Hussain, "Yield and phenological responses of BT cotton to different sowing dates in semi-arid climate," *Pakistan J. Agric. Sci.*, vol. 54, no. 2, pp. 233–239, 2017.
[58] S. Tahir, A. Ahmad, T. Khaliq, and M. J. M. Cheema, "Evaluating the impact of seed rate and sowing dates on wheat productivity in semi-arid environment," *Int. J. Agricult. Biol*, vol. 22, no. 1, pp. 57–64, 2019.
[59] B. Barnabás, K. Jäger, and A. Fehér, "The effect of drought and heat stress on reproductive processes in cereals," *Plant, Cell Environ.*, vol. 31, no. 1, pp. 11–38, 2008.



UFERAH SHAFI received the B.Sc. degree in mathematics from Bahauddin Zakariya University, Multan, Pakistan, the M.Sc. degree in information technology from Quaid-i-Azam University, Islamabad, Pakistan, and the M.S. degree in computer science from COMSATS University, Islamabad. She is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, NUST, Pakistan. Her research interests include deep learning, remote sensing, image processing, deep learning, and the Internet of Things (IoT).

RAFIA MUMTAZ (Senior Member, IEEE) received the Ph.D. degree in remote sensing and satellite image processing from the University of Surrey, U.K., in 2010. She is currently a Professor and the Director of the NUST Coventry Internet of Things Laboratory (NCIL), NUST-SEECS. She was a recipient of several national and international research grants worth PKR 134 million. She received the NUST-SEECS Best Researcher Award, in 2019, the Women of Wonder Award, in 2021, the University Best Teacher Award, in 2022, and the HEC Best University Teacher Award, in 2023.

ZAHID ANWAR received the Ph.D. degree in computer science from the University of Illinois, in 2008. He is currently an Associate Professor in cybersecurity with the Department of Computer Science, NDSU, and a Scholar with The Sheila and Robert Challey Institute for Global Innovation and Growth. He has authored more than 100 publications in peer-reviewed conferences and journals. His research interests include cybersecurity policy and innovative cyber defense. He is also a CompTIA-certified penetration tester, security+professional, and an AWS-certified cloud solutions architect. Prior to working in academia, he was a Software Engineer and a Researcher with IBM T. J. Watson, Intel, Motorola, the National Center for Supercomputing Applications, xFlow Research, and CERN.

MUHAMMAD MUZYYAB AJMAL received the bachelor's degree in information technology from Bahauddin Zakariya University, Multan, Pakistan. He is currently pursuing the master's degree in information technology with the School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST). In the past, he has gained professional experience as a data analyst, a database administrator, and a mobile application developer. His research interests include deep learning, remote sensing, image processing, IOMT, and the IoT.

MUHAMMAD AJMAL KHAN has over 30 years of academic, research and development, and leadership experience in the fields of electronics, information, and communication technology. He has worked for over a decade in the avionics industry and has hands-on experience in ground and airborne communication equipment, networks, electronics, and IT systems. As the Founder and the Director of the Information Assurance Research and Development Centre, he led diverse teams for developing indigenous software/hardware solutions for networks and IT systems. He initiated and successfully conducted several projects related to the development of secure software and hardware ICT products. He has been actively involved in the review of various national and organizational level cyber security and IT policies. He has vast teaching and research experience, including the Head of the Department, the Dean, and the Commandant of the PAF College of Aeronautical Engineering. His research interests include machine learning for data processing, secure product development, and embedded systems.

ZAHID MAHMOOD received the Ph.D. degree from Quaid-i-Azam University, Islamabad, Pakistan. He did his thesis research with The University of Queensland, Australia, in plant science. He is currently a Senior Scientific Officer with the Crop Sciences Institute, National Agricultural Research Centre, Islamabad. He has more than 20 years of professional experience in crop improvement. He has published and presented 63 research articles in international journals and conferences. His current research interests include plant genetics, speed breeding, high-throughput genotyping, and high-throughput phenotyping using UAV-based sensors for crop evaluation and development. He also received several international fellowships/trainings at leading institutes and universities in the USA, Australia, China, Mexico, Kenya, and India.

MAQSOOD QAMAR received the M.Sc. degree in crop science from Wageningen University, The Netherlands, and the Ph.D. degree from the University of Muzaffarabad, Azad Jammu and Kashmir, Pakistan. His Ph.D. thesis research from The University of Sydney, Australia, in plant breeding and molecular genetics. He is currently a Principal Scientific Officer/Program Leader of the Wheat Program, National Agricultural Research Centre, Crop Sciences Institute, Islamabad, Pakistan. He has more than 20 years of professional experience in wheat crop improvement, plant breeding, and genetics. He has published more than 58 research articles in national and international journals. His current research interests include the development of climate-resilient and disease-resistant wheat varieties. He also received several international fellowships/training in leading institutes and universities in Australia, China, Turkey, Kenya, Ethiopia, and Taiwan.

HAFIZ MUHAMMAD JHANZAB received the Ph.D. degree in agronomy from PMAS-Arid Agriculture University Rawalpindi. His Ph.D. thesis titled "Assessment of Growth and Physiochemical Responses of Wheat to Chemo-Blended Silver and Iron Nanoparticles. He is currently a Lecturer with the Department of Agronomy, The University of Agriculture, Dera Ismail Khan. He has more than ten years of experience in improving crop production, cropping systems, seed production, variety development, and organic agriculture. He has published 15 research articles in well-reputed international journals. His research interests include the use of nanoparticles as a slow-release fertilizer, chemo-blended nanoparticles for improvement in plant growth, proteomic and metabolomic analysis, and resource conservation technologies for crop production. He received an international fellowship from HEC to work with the University of Tsukuba, Japan.