


ORIGINAL ARTICLE

Agronomic Application of Genetic Resources

Wheat crop genotype and age prediction using machine learning with multispectral radiometer sensor data

Mutiullah Jamil¹ | Zoha Ahsan¹ | Muhammad Nauman Saeed¹ | Ali Raza² |
 Hazem Migdady³ | Mohammad Sh. Daoud⁴ | Maryam Altalhi⁵ | Absalom E. Ezugwu⁶ |
 Laith Abualigah^{7,8,9,10,11,12} 

¹Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

²Department of Software Engineering, University Of Lahore, Lahore 54000, Pakistan

³CSMIS Department, Oman College of Management and Technology, Barka, Oman

⁴College of Engineering, Al Ain University, Abu Dhabi, United Arab Emirates

⁵Department of Management Information Systems, College of Business Administration, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

⁶Unit for Data Science and Computing, North-West University, Potchefstroom, South Africa

⁷Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan

⁸MEU Research Unit, Middle East University, Amman, Jordan

⁹Computer Science Department, Al al-Bayt University, Mafraq 25113, Jordan

¹⁰Applied Science Research Center, Applied Science Private University, Amman, Jordan

¹¹School of Engineering and Technology, Sunway University Malaysia, Petaling Jaya 27500, Malaysia

¹²Jadara Research Center, Jadara University, Irbid 21110, Jordan

Correspondence

Absalom E. Ezugwu, Unit for Data Science and Computing, North-West University, 11 Hoffman Street, Potchefstroom 2520, South Africa.

Email: Absalom.ezugwu@nwu.ac.za

Assigned to Associate Editor David E. Clay.

Abstract

Wheat (*Triticum aestivum*) yield predictions can be improved by using multispectral remote sensing to identify different genotypes and crop growth stages. We propose an innovative machine learning technique aimed at classifying diverse wheat crop genotypes and providing accurate estimations of plant age. Multispectral reflectance data was obtained from different sites where various wheat genotypes were cultivated. This approach involved analyzing incoming radiation and canopy light reflectance across five distinct spectral bands using a multispectral radiometer. The newly collected remote sensing data was utilized as input for the machine learning algorithm. Impressively, the random forest model achieved an accuracy rate of 98.77% in wheat crop genotype classification. Furthermore, the proposed approach's effectiveness was confirmed through a 10-fold cross-validation mechanism. Moreover, a multiple linear regression model for predicting the age of wheat genotypes explained 91% of the observed variation. These findings signify significant progress in wheat crop genotype and age prediction, ultimately leading to enhanced wheat yield.

Abbreviations: LR, linear regression; MLR, multiple linear regression; NDVI, normalized difference vegetation index; NIR, near infrared; RF, random forest; SVM, support vector machine; SWNIR, short-wave near infrared.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Agronomy Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy.

1 | INTRODUCTION

Wheat (*Triticum aestivum*) grains are rich in essential nutrients, establishing themselves as a valuable nutritional source that enhances diets worldwide (Zahra et al., 2023). Wheat consumption in Pakistan surpasses that of rice, and the per capita consumption of wheat is 124 kg annually (Bakhsh et al., 2003). To effectively address Pakistan's escalating food requirements, advanced agricultural techniques are needed (Panhwar et al., 2021). The research findings by Khan et al. (2015) shed light on the vulnerability of the flowering stage to heat stress. Furthermore, Nabwire et al. (2022) emphasizes the pivotal role of a plant's age in managing water stress and temperature and sourcing essential nutrients from various avenues.

Plant morphology can be used to compare different species, differentiate between various types of plants, or study how plants respond to stimuli (Wyatt, 2016). Some of the most important morphological traits include leaf shape, size, color, texture, angle, and volume. Within the shoot system, leaves adapt to their environment by altering their visual properties, making them recognizable (Yang et al., 2015). Developing alternative phenotypic classification approaches other than physical measurements is important for accelerating breeding, and the prediction of food resources is critical for improving food security.

To comprehend a nation's food resources, it is crucial to conduct a comprehensive assessment of potential crop harvests (Akhter et al., 2023). In this ever-changing landscape, precise and meticulous crop evaluations play a vital role in generating valuable information that informs the strategic management of plant cultivation, allocation of resources, and food security. The intricate interplay between data-driven analysis and farming methods encapsulates the essence of this endeavor, illuminating the path toward sustainable and resilient agricultural systems.

Consequently, this research aims to establish an innovative machine learning-based framework for categorizing wheat genotypes, accompanied by the development of a precise age prediction model for each specific genotype. Optimal yield can be achieved through the cultivation of wheat genotypes in harmony with their respective conducive environmental conditions. The efficacy and precision of our proposed machine learning-based model for wheat genotype classification and age predictions are evaluated through various parameters.

Numerous scholars have contributed to the advancement of wheat genotype classification through a diverse array of approaches. For example, Naser et al. (2020) proposed a model that utilizes the Normalized Difference Vegetation Index (NDVI) to distinguish between wheat genotypes' productivity in dry and wet environments. Their study, conducted in Northeastern Colorado, encompassed various climatic conditions. Employing NDVI data acquired from a prox-

Core Ideas

- This research addresses a significant challenge in wheat crop genotype and age prediction.
- We propose an innovative machine learning methodology to classify different wheat crop genotypes.
- We collected different wheat seed genotype samples using the multispectral radiometer.

imal sensor to gauge the greenness of wheat fields, they also gathered data on grain yield for each wheat genotype. The findings demonstrated a robust correlation between NDVI and grain yield, with higher NDVI readings associated with wheat genotypes exhibiting greater grain yields. Notably, precise measurements of grain yield and effective discrimination of superior wheat genotypes were achieved at non-saturated NDVI values, particularly around the threshold of 0.9. Additionally, they determined that the k-means clustering algorithm could reliably categorize wheat genotypes into three classes of grain yield productivity based on their respective NDVI readings.

A remote sensing study was conducted by Han et al. (2022) to investigate using a random forest (RF) model in monitoring wheat phenology. They discovered that the RF model demonstrated high accuracy in predicting plant nitrogen accumulation, nitrogen nutrition index, aboveground biomass, and nitrogen concentration. The researchers collected multispectral images and crop data at five growth stages.

The study of Raoufi et al. (2018) involved the emulation of growth and harvest patterns of diverse wet rice genotypes at varying seedling ages using the AquaCrop model. The research employed version 4.0 of the AquaCrop model to simulate rice growth. The experimentation spanned 2 years and was carried out at the Haraz Extension and Technology Development Center in Amol, Mazandaran Province, Iran. The study focused on three rice genotypes—Tarom, Ghaem, and Fajr—each exhibiting distinct growth period durations. Raoufi et al. (2018) showed that the model could be used to predict rice yields.

Zhang et al. (2021) used MODIS NDVI time-series satellite data to distinguish winter wheat from other crops. The Heilongjiang region was chosen for winter wheat mapping over four consecutive years (2014–2017). The model employed the peak-slope difference index and the NDVI time-series variation coefficient for wheat crop mapping, specifically utilizing NDVI data from the MOD13Q1 dataset (Hubert-Moy et al., 2019). Landsat-8 multispectral images were acquired from the U.S. Geological Survey (USGS), and sample sites were selected using data from the USGS website, Google Earth

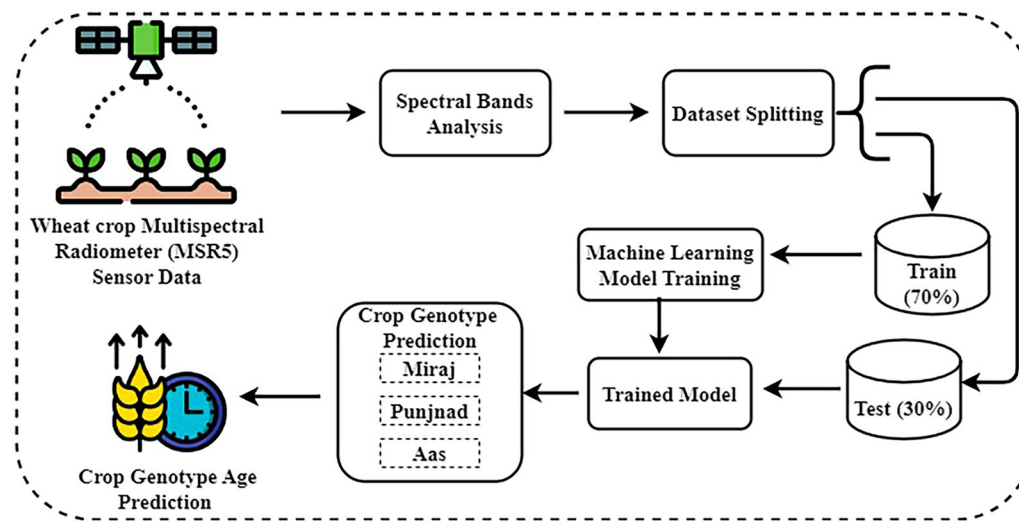


FIGURE 1 Proposed innovative methodology workflow for the prediction of wheat crop genotype and age.

photos, and statistical information. The coefficient of variation (COV) of PSDI demonstrated high user and accuracy rates, achieving 94.10% and 93.74%, respectively.

Das et al. (2021) proposed a methodology to assess water conditions in wheat genotypes using thermal imaging from unmanned aerial vehicles. This approach was valuable in predicting yields in sodic soils. This technique effectively classified agricultural water stress factors and provided biomass and grain production forecasts based on crop water stress indices. Applying classification and regression trees yielded highly accurate predictions for grain yield, root mean square error, and biomass. In the context of sodic soil conditions, wheat genotypes, including Gregory, Bremer, Mace, Lancer, and Mitch, demonstrated greater productivity than Flanker, Gladius, Emu Rock, Scout, and Janz. This research highlights genotype-specific productivity, offering valuable insights for wheat cultivation.

Sandhu et al. (2021) introduced multi-trait machine learning and deep learning models to enhance wheat breeding programs. They observed that the proposed models outperformed genomic best linear unbiased predictor (GBLUP). The authors conducted their study on a dataset comprising wheat genotypes phenotyped for grain yield and grain protein content. Furthermore, the genotypes were assessed for spectral reflectance, which was used to train the machine learning and deep learning models. The authors compared the performance of four uni-trait (UT) and four multi-trait (MT) models. Their findings indicated that the MT and deep learning models surpassed the UT models and the GBLUP methods. The RF and multilayer perceptron models demonstrated the highest performance among the models. The authors concluded that the proposed models represent a promising tool for genomic selection in wheat breeding programs, suggesting their potential in selecting wheat genotypes with superior grain yield and grain protein content.

Fang et al. (2020) used Sentinel-2 imagery with winter wheat. The research conducted in Henan Province, Central China, involved acquiring Sentinel-2 images of winter wheat at a specific phenological stage through Google Earth Engine. Machine learning techniques, including RF, support vector machine (SVM), and classification and regression tree, were employed to identify and map winter wheat across a wide area. Five-fold cross-validation and grid search approaches were utilized to optimize machine learning hyperparameters. The SVM demonstrated superior performance in classifying winter wheat, as indicated by comparing the three algorithms. It achieved an overall accuracy (OA) of 0.94, user's accuracy (UA) of 0.95, producer's accuracy (PA) of 0.95, and Kappa coefficient (Kappa) of 0.92. The results emphasized the SVM's sensitivity to specific parameters (C and gamma), which led to the highest classification accuracy when these hyperparameters were optimized.

Due to the lack of research on genotype classification using multispectral data in the literature, our study aims to address this gap. The primary objective of our research is to design a data acquisition system using multispectral MSRF5 sensors. Additionally, we have developed an automated machine learning-based technique to detect wheat growth stages.

2 | METHODS AND MATERIALS

The research study was conducted in the years 2020 and 2021 under the supervision of the IUB Agriculture Research Center. During data collection, nine plots were harvested, focusing on three types of genotypes and three different conditions of water stress. These conditions included normal watering, a 1-week delay in watering, and a 2-week delay in watering. No fertilization or spraying treatments were applied.

In this study, the selection of wheat crop genotypes for classification is based on diverse criteria, including genetic variability, agronomic performance, and adaptability to specific environmental conditions. This comprehensive methodology allowed for a systematic and rigorous investigation into the classification of wheat crop genotypes, providing valuable insights into their genetic diversity and potential agricultural applications.

Our proposed innovative research methodology (Figure 1) involves architectural analysis. The multispectral radiometer (MSR5)-based sensor data is collected and utilized for building genotype classification and age prediction machine learning models. The collected multispectral radiometer sensor data is preprocessed and converted into five spectral bands. The formatted dataset is then split into training and testing portions. The 70% training portion of the dataset is utilized for training the applied machine learning models. The remaining 30% of the data is used for the evaluation of the machine learning model. The machine learning model is then used for cultivar classification. Following this, SPSS software is employed to predict the cultivar age. Using SPSS software, a multiple linear regression (MLR) model is applied to the classified data for predicting the age of each genotype.

2.1 | Multispectral radiometer sensors data

The study focused on three test genotypes: Miraj, Punjnad, and Aas, each cultivated in pairs, with one plot under water stress. Plots with the dimensions of 3.66 by 3.66 m were established in 2020 and 2021. Plots were planted at a rate determined by each plot size, which measured 2.32 m² (length and width) on 2020 and 2021. After 2 weeks, each genotype underwent 30 MSR5 scans by CROPSCAN, Inc. The process yielded 90 samples at 15-day intervals over 3 months, totaling 540 samples representing six developmental stages, as shown in Figure 2.

2.1.1 | Data collection area

The data collection area chosen is within the Agricultural Research Center located at the Islamia University of Bahawalpur, situated in the dynamic city of Bahawalpur, Punjab, Pakistan. The data collection locations are illustrated in satellite as in Figure 3. This diverse study area encompasses various agro-climates typical of Punjab, where the annual rainfall can be as low as 2 mm (0.1 in.). Among these climates, October records the scantiest rainfall, while July is the wettest month, receiving 61 mm (2.4 in.) of rainfall. Bahawalpur, known for its soaring temperatures, often grapples with water scarcity issues that pose significant challenges.

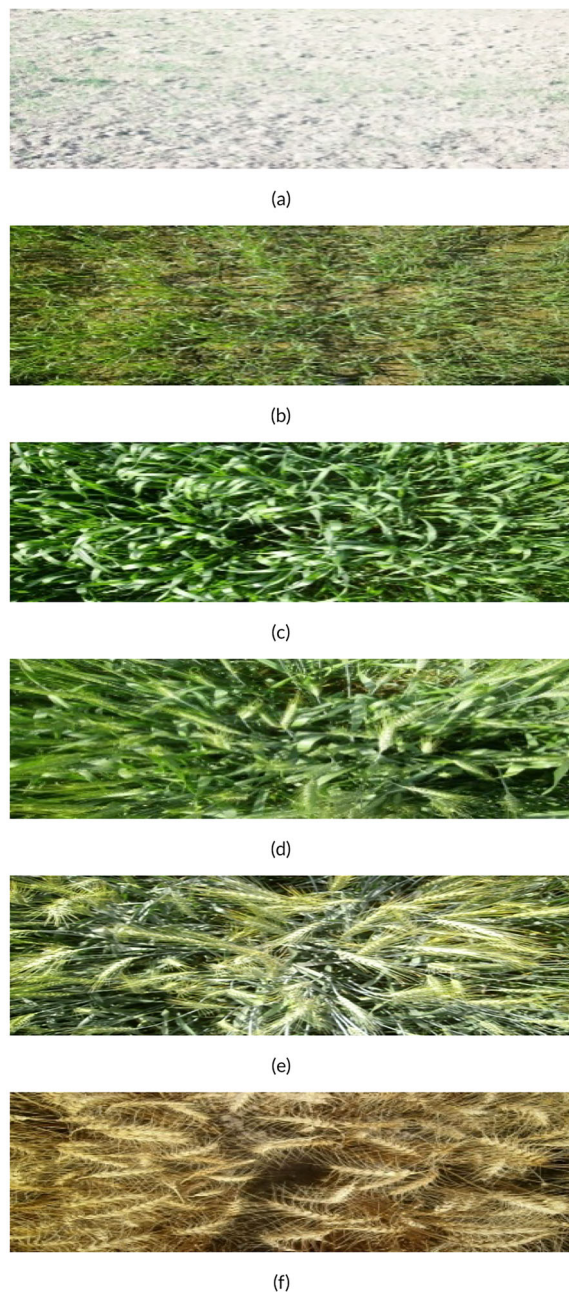


FIGURE 2 The photographic representation of wheat crop of six stages: (a) stage 1, (b) stage 2, (c) stage 3, (d) stage 4, (e) stage 5, and (f) stage 6.

2.1.2 | Data collection experiment design

Observations were made between 2 and 12 weeks. This choice was based on the fact that temperatures below 13°C inhibit flowering, while temperatures exceeding 14°C after flowering and fruit set have negligible effects on plant growth (Noh et al., 2013). The wheat plants were categorized into plants cultivated under optimal growth conditions and plants subjected to high-temperature stress. The selection of plants for the stress group were randomly selected. Each group of

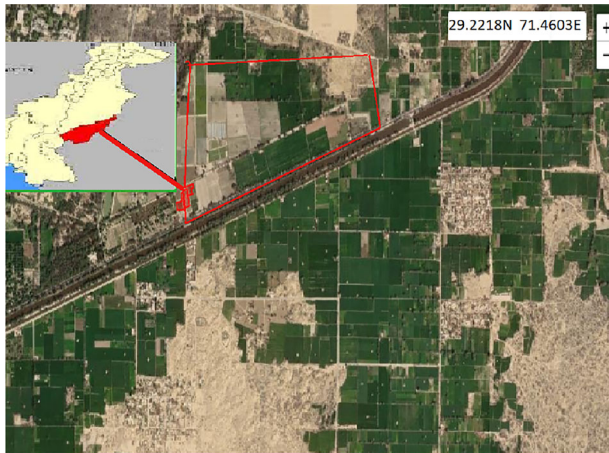


FIGURE 3 Location of the study site using Google Earth View with the map of Pakistan and highlighted in red color ROI at the upper top right corner of the image.

TABLE 1 The soil characteristics analysis during data collection.

Soil characteristics	0–15 cm	15–30 cm
Organic matter (%)	0.79	0.55
pH value	8.4	8.6
Electrical conductivity (dS/m)	250	230
Phosphorus (ppm)	7.1	5.1
Potassium (ppm)	112	114
Saturation characterizing soil texture (%)	36	35

plants was cultivated in dedicated plots, maintaining a consistent relative humidity of 70% throughout the entire growth period. Both the standard and stressed groups adhered to distinct watering schedules, with irrigation administered every 15 days. The irrigation conditions included normal watering, a 1-week delay in watering, and a 2-week delay in watering. Wheat required a total of five irrigations. One irrigation equals 3 ha in., so 15 ha in. was required. No fertilization or spraying treatments were applied during data collection. The ratio of plant population was between 1.2 and 2.0 million seeds per ha. Soil physicochemical analysis was carried out before sowing the crop. Soil samples were taken from 0.0 to 0.15 m and 0.15 to 0.30 m using a soil augur. Soil characteristics (Shah et al., 2020) analysis data are given in Table 1.

2.2 | Multispectral radiometer bands analysis

The multispectral radiometers are utilized to assess incoming radiation and canopy light reflectance across five distinct spectral bands (Qadri et al., 2019; Rehmani et al., 2015).

TABLE 2 The wavelength and spatial resolution for the collected crop scan MSR5 data.

Spectral band	Wavelength (nm)	Spatial resolution (area covered by the sensor)
Band 1 Blue	450–520	1.524 m in radius
Band 2 Green	520–630	1.524 m in radius
Band 3 Red	630–690	1.524 m in radius
Band 4 SNIR	760–900	1.524 m in radius
Band 5 FNIR	1550–1750	1.524 m in radius

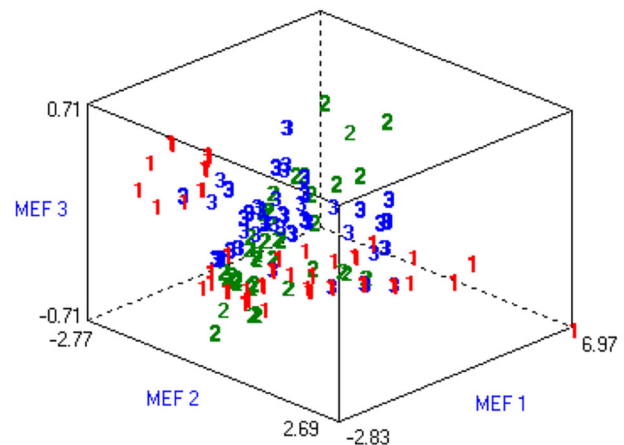


FIGURE 4 The feature space analysis of extracted multispectral bands data; most expressive feature (MEF 1, 2, and 3).

The generated output dataset contained blue (450–520 nm), green (520–600 nm), red (630–690 nm), near-infrared (760–900 nm), and far-infrared wavelengths (1550–1750 nm). Within each specific spectral band, the half-peak width varies, ranging from approximately 5 to 15 nm. This innovative approach, referred to as MSR5, encapsulates an entire scene by utilizing five distinct numerical values, effectively representing five energy bands, as described in Table 2.

2.3 | Feature space analysis

A feature space analysis was conducted to extract the important multispectral bands. The analysis began with the calculation of principal components for feature space analysis. We selected the top five principal components from the band data and illustrated them in Figure 4. This analysis reveals that over 90% of variance is captured in the multispectral bands data. The dataset's feature space exhibits greater linear separability for wheat genotype classification.

2.4 | Applied machine learning methods

2.4.1 | Random forest

Random forest is a commonly used technique for the classification of multispectral data and yields enhanced results compared to other machine learning models (Raza et al., 2023). In the RF model, a value of 100 was utilized for the `n_estimators` parameter, which specifies the number of trees in the RF model.

The RF prediction for the wheat crop genotype can be represented as:

$$RF(X) = \frac{1}{N} \sum_{i=1}^N f_i(X) \quad (1)$$

where N is the number of decision trees in the forest,

X is the feature matrix with n samples and m features,

Y is the target variable representing the wheat crop genotype,

T_i represents the i th decision tree in the forest, and

$f_i(X)$ is the prediction of the i th decision tree.

2.4.2 | Support vector machine

A widely used supervised machine learning technique for classification and regression tasks is known as a SVM (Raza et al., 2022). SVMs have a good ability to differentiate between multiple classes or make precise predictions for continuous values. The SVM model can be represented by the following equation:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (2)$$

2.4.3 | Logistic regression

Logistic regression (LR) (Raza et al., 2023) is a statistical method used to model the relationship between a binary dependent variable and one or more independent variables. The main objective of logistic regression is to estimate the likelihood of a specific outcome based on distinct variables. In contrast to linear regression, which employs a linear equation for modeling variable relationships, logistic regression transforms independent variables into a probability range from 0 to 1 using the logistic function, also known as the sigmoid function. The logistic regression equation is given

by:

$$P(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n)}} \quad (3)$$

2.4.4 | Multiple linear regression

A statistical modeling method known as MLR (Sharma et al., 2022) is employed to investigate the relationship between several independent variables and a dependent variable. This approach extends the principle of simple linear regression to scenarios with multiple independent variables. The aim of MLR is to determine the most accurate linear equation that estimates the value of the dependent variable based on the values of the independent variables. The mathematical equation for MLR can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (4)$$

Tables 3–5 provide a comprehensive array of regression coefficients. In the “Unstandardized coefficients” column, “B” indicates weights. Notably, for the Miraj, Punjnad, and Aas tables, the “B” weights are 53.347, 107.728, and 107.126, respectively, with the “Constant” row representing the intercept. The “B” weight serves as a predictor in conjunction with the slope. A negative slope value implies a negative correlation. These coefficients intricately shape the linear regression equation, providing insight into the relationship. Significance across these three tables is remarkably low, at 0.000, underscoring the influential role of independent variables on the dependent variable.

2.5 | Hyperparameter tuning

The best-fit hyperparameters of the applied machine learning methods are determined, as illustrated in Table 6. In this research study, we employed a grid search approach to optimize the machine learning hyperparameters (Shekar & Dagnev, 2019). The best-fit hyperparameters help us achieve high-performance accuracy scores.

2.6 | Analysis

We used the Python programming language to conduct all research experiments (Hao & Ho, 2019). The Scikit-learn library in Python, version 1.0.2, was utilized to evaluate performance metrics for wheat crop genotype classification. The performance metrics included accuracy, recall, precision, and F1 scores. We have employed several methods to evaluate performance scores, including comparisons of

TABLE 3 The coefficient analysis of the Miraj genotype.

Model	Unstandardized coefficients		Standardized coefficients	Significance
	B	Stdard error	Beta	
Constant	53.347	6.401		0.000
Blue spectral band	-3.119	0.415	-0.614	0.000
Green spectral band	-6.541	1.231	-0.688	0.000
Red spectral band	15.822	0.581	2.576	0.000
Near-infrared spectral band	1.500	0.100	0.694	0.000
Far-infrared spectral band	-5.334	0.301	-1.019	0.000

TABLE 4 The coefficient analysis of the Punjnad genotype.

Model	Unstandardized coefficients		Standardized coefficients	Significance
	B	Stdard error	Beta	
Constant	105.728	11.277		0.000
Blue spectral band	0.692	0.759	0.170	0.364
Green spectral band	-15.830	2.414	-1.947	0.000
Red spectral band	14.848	0.910	2.618	0.000
Near-infrared spectral band	0.670	0.177	0.238	0.000
Far-infrared spectral band	-3.788	0.582	-0.745	0.000

TABLE 5 The coefficient analysis of the Aas genotype.

Model	Unstandardized coefficients		Standardized coefficients	Significance
	B	Stdard error	Beta	
Constant	107.126	8.837		0.000
Blue spectral band	-0.550	0.636	-0.117	0.389
Green spectral band	-15.028	1.703	-1.749	0.000
Red spectral band	16.724	0.636	2.820	0.000
Near-infrared spectral band	0.856	0.168	0.384	0.000
Far-infrared spectral band	-4.858	0.494	-0.899	0.000

TABLE 6 The hyperparameters settings for applied machine learning models.

Method	Hyperparameter description
RF	n_estimators = 100, criterion = "entropy," random_state=1
SVM	kernel = "linear," C = 10, random_state = 3
LR	random_state = 2, max_iter=700

Abbreviations: LR, logistic regression; RF, random forest; SVM, support vector machine.

machine learning model results. This also includes confusion matrix comparisons, *k*-fold cross-validation, and feature space comparisons.

The accuracy metric is calculated using the following equation:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (5)$$

The recall metric is calculated using the following equation:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad (6)$$

The precision metric is calculated using the following equation:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad (7)$$

TABLE 7 Performance analysis of applied machine learning methods for unseen testing data.

Method	Accuracy	Target class	Precision	Recall	F1 score
RF	0.98	Miraj	0.96	1.00	0.98
		Punjnad	1.00	0.97	0.98
		Aas	1.00	1.00	1.00
		Average	0.99	0.99	0.99
SVM	0.90	Miraj	0.98	0.98	0.98
		Punjnad	0.89	0.89	0.89
		Aas	0.86	0.86	0.86
		Average	0.91	0.91	0.91
LR	0.84	Miraj	0.87	0.92	0.89
		Punjnad	0.87	0.84	0.85
		Aas	0.80	0.78	0.79
		Average	0.85	0.85	0.85

Abbreviations: LR, logistic regression; RF, random forest; SVM, support vector machine.

The F1 metric is calculated using the following equation:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

3 | RESULTS

3.1 | Performance analysis of machine learning methods

This analysis provides valuable insights into the performance of various machine learning models (Table 7). The analysis compared: accuracy, precision, recall, and F1 score. More so, the analysis shows that only the LR model achieved moderate performance scores of 0.84 and RF model outperformed the other models (Figure 5).

Furthermore, the comprehensive histogram analysis, depicted in Figure 5 showed that the RF and SVM methods had good precision, recall, and F1 scores. On the other hand, the LR model yielded satisfactory results, indicating its potential for further optimization.

The columns and rows in the confusion matrix (Figure 6) are denoted by 0, 1, and 2, eloquently representing the Miraj, Punjnad, and Aas genotypes, respectively. The diagonal elements within this matrix gracefully show the adeptly classified data, showcasing the proficiency of the RF, SVM, and logistic regression machine learning models. The remaining entries of the matrix affectingly disclose instances where the three genotypes were unfortunately mispredicted. The RF model has successfully classified data with 98.77% accuracy, the SVM accuracy stands at 90.74%, and the accuracy of the logistic regression stands at 84.57% during validation.

TABLE 8 *k*-Fold-based performance validation of applied machine learning methods.

Method	Folds	Accuracy	Standard deviation (+/-)
RF	10	0.99	0.0119
SVM	10	0.90	0.0292
LR	10	0.84	0.0532

Abbreviations: LR, logistic regression; RF, random forest; SVM, support vector machine.

TABLE 9 Performance analysis of the applied multiple regression model.

Genotype	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> ²	Standard error
Miraj <i>Y</i> ₁	0.972	0.946	0.943	6.12
Punjnad <i>Y</i> ₂	0.931	0.867	0.861	9.57
Aas <i>Y</i> ₃	0.963	0.926	0.923	7.12

3.2 | *k*-Fold cross-validation

The performance of the applied machine learning models are rigorously validated through 10-fold cross-validation. This approach enables a comprehensive assessment of how well the applied models handle unseen data. The outcomes of the cross-validation analysis, summarizing the performance of the models across different folds, are presented in Table 8. These findings suggest that ML can be used to predict genotype and plant age.

3.3 | Performance analysis of crop genotype age prediction

The interplay between the coefficients of green and short-wave near infrared (SWNIR) and the age of the wheat crop genotype showcases an inverse correlation. In practical terms, as the plant's age increases, there is a gradual reduction in the intensity of the green spectral band, along with the SWNIR values. Conversely, the red and NIR values demonstrate a direct proportionality with the age of the wheat genotype, exhibiting an upward trend.

The multiple regression model summary in Table 9 depicts the correlation coefficient (*R*) and the *R*² statistic, indicating the "proportionate decrease in error." These values collectively assess the model's performance in predicting the age of the Miraj, Punjnad, and Aas genotypes. A higher *R*² value implies a better fit. The regression models explained at least 86% of the variation, using red, green, near-infrared, and short-wave near-infrared as predictors. The model achieves overall accuracy of over 90%, with a standard error range of 7.1–9 had.

Depicted in the provided Figure 7 is a scatter plot showcasing the age distribution of wheat crops. Notably, this

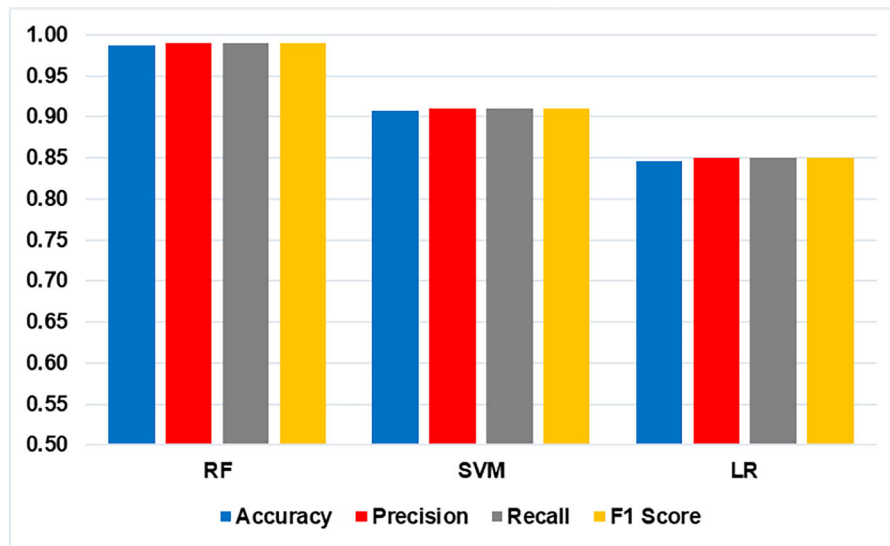


FIGURE 5 The histogram-based performance comparison of machine learning methods. LR, logistic regression; RF, random forest; SVM, support vector machine.

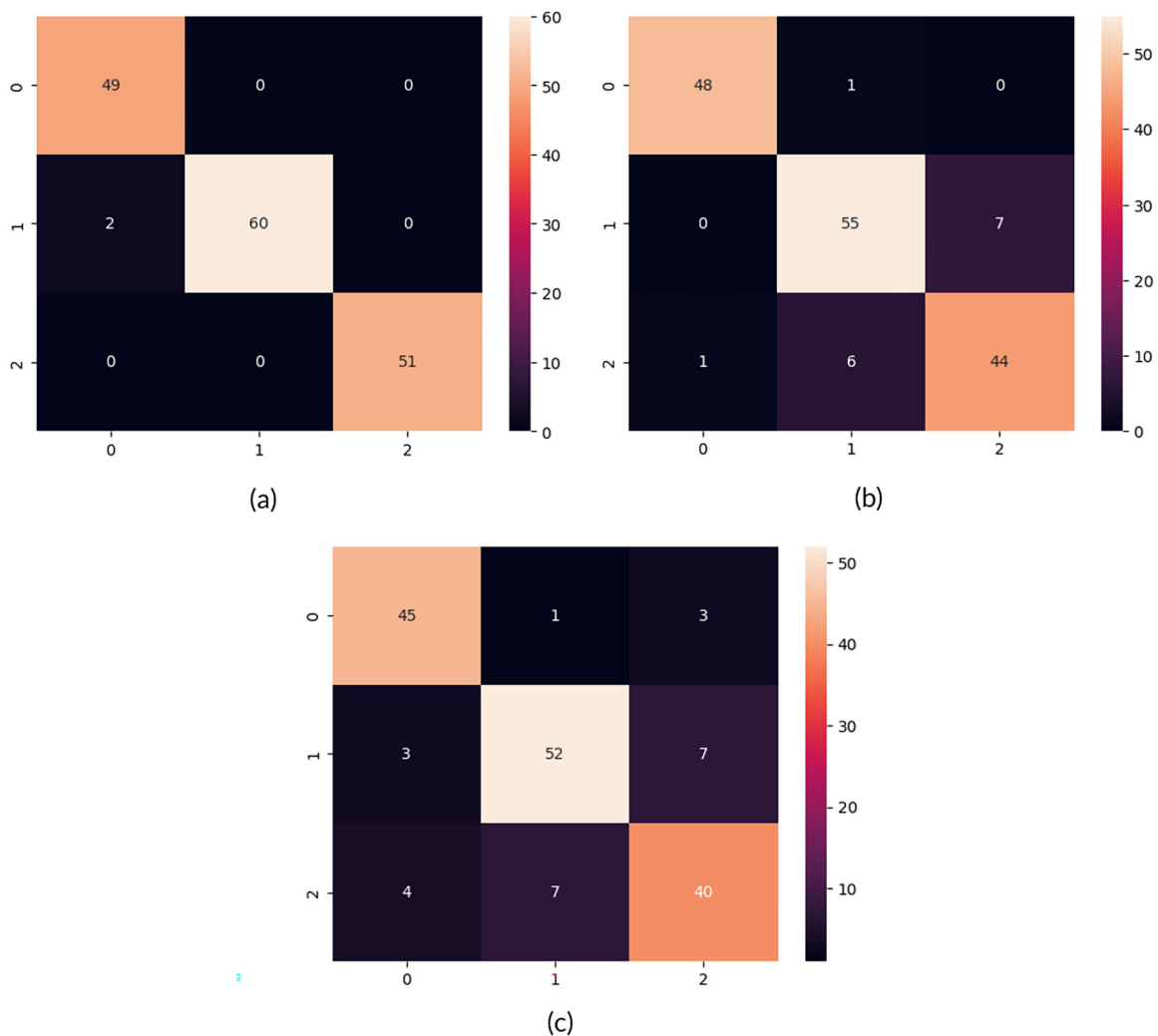


FIGURE 6 The confusion matrix-based performance validations of applied techniques: (a) random forest (RF), (b) support vector machine (SVM), and (c) logistic regression (LR).

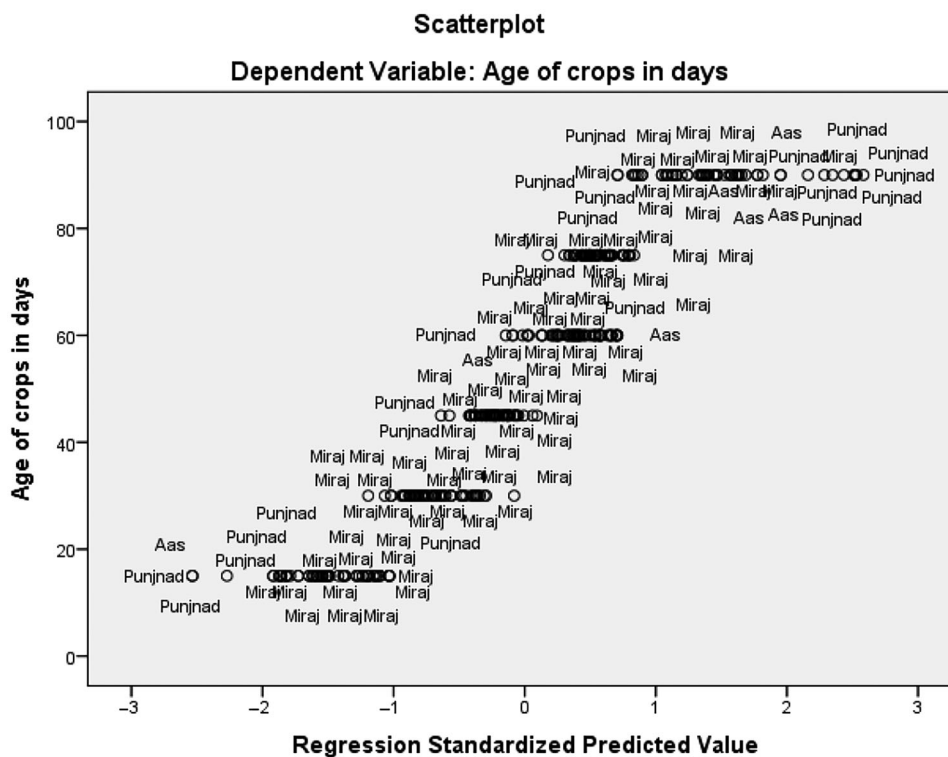


FIGURE 7 The scatter plot of age prediction of wheat genotypes.

TABLE 10 The state of the art approaches comparisons.

Reference	Proposed technique	Performance accuracy
Rehmani et al. (2015)	Artificial neural network (ANN)	0.96
Qadri et al. (2016)	Artificial neural network (ANN)	0.96
Jamil et al. (2023)	Artificial neural network (ANN)	0.97
Jamil et al. (2023)	Random forest (RF)	0.91
This study	Random forest (RF)	0.99

visualization underscores the proficiency of the multiple linear model in predicting the ages of the wheat crop genotypes. The state-of-the-art comparison results are described in Table 10.

3.4 | Discussion

The inclusion of diverse wheat genotypes Miraj, Punjnad, and Aas in our dataset ensures the generalizability of our findings across varieties, making our methodology applicable to a broader range of agricultural settings. The high accuracy achieved by the RF model of 98.77% in wheat crop genotype classification underscores the effectiveness of the machine learning approach. This accuracy is particularly noteworthy as

it provides a reliable means for distinguishing between genotypes. The successful application of diverse machine learning models, including support vector machine and logistic regression, in comparative analyses demonstrates the robustness of our methodology.

The implementation of *k*-fold cross-validation mechanisms further strengthens the credibility of the results. The validation process ensures the generalizability of the models by assessing their performance across various subsets of the dataset. The consistency of high accuracy values across folds substantiates the robustness and reliability of our classification approach.

Our focus on age prediction, a critical aspect of precision agriculture, adds a dimension of practicality to our research. The MLR model developed for age prediction explained 91% of the variability. Such accurate age predictions can significantly contribute to timely and targeted agricultural interventions, optimizing resource management and improving overall crop yield.

4 | CONCLUSIONS

This study demonstrated the effectiveness of multispectral radiometry and machine learning techniques for wheat crop genotype classification and age prediction. The data in this research were collected using a multispectral radiometer encompassing five bands: blue, green, red, near infrared

(NIR), and SWNIR. Among the machine learning models (RF, SVM, and LR), RF excelled in wheat genotype classification, achieving an accuracy rate of 98.77%. The robustness of the classification model is validated through *k*-fold cross-validation. Furthermore, the machine learning model designed to predict additional phenotypic traits, including crop age, exhibited exceptional performance. MLR successfully predicted plant age based on spectral features, achieving over 90% accuracy. Overall, this study establishes the potential of MSR5 spectral bands for estimating the age of wheat crop genotypes. This study can serve as a foundation for the improvement of a real-time monitoring system for wheat crops in high-throughput plant phenotyping facilities.

In the future, we will collect more dataset samples and enhance the wheat genotypes. We will also develop an advanced neural network approach for effective wheat genotype classification. Additionally, we will utilize other sensors similar to MSR.

AUTHOR CONTRIBUTIONS

Mutiullah Jamil: Conceptualization. **Zoha Ahsan:** Conceptualization. **Muhammad Nauman Saeed:** Conceptualization. **Ali Raza:** Conceptualization. **Hazem Migdady:** Conceptualization. **Mohammad Sh. Daoud:** Conceptualization. **Maryam Altalhi:** Conceptualization. **Absalom E. Ezugwu:** Conceptualization; writing—review and editing. **Laith Abualigah:** Conceptualization.

ACKNOWLEDGMENTS

The authors would like to acknowledge Deanship of Graduate Studies and Scientific Research, Taif University for funding this work.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Laith Abualigah  <https://orcid.org/0000-0002-2203-4549>

REFERENCES

- Akhter, M. J., Sonderskov, M., Loddo, D., Ulber, L., Hull, R., & Kudsk, P. (2023). Opportunities and challenges for harvest weed seed control in European cropping systems. *European Journal of Agronomy*, *142*, 126639.
- Bakhsh, A., Hussain, A., & Khan, A. S. (2003). Genetic studies of plant height, yield and its components in bread wheat. *Sarhad Journal of Agriculture*, *19*(4), 529–534.
- Das, S., Christopher, J., Apan, A., Choudhury, M. R., Chapman, S., Menzies, N. W., & Dang, Y. P. (2021). Evaluation of water status of wheat genotypes to aid prediction of yield on sodic soils using UAV-thermal imaging and machine learning. *Agricultural and Forest Meteorology*, *307*, 108477.
- Fang, P., Zhang, X., Wei, P., Wang, Y., Zhang, H., Liu, F., & Zhao, J. (2020). The classification performance and mechanism of machine learning algorithms in winter wheat mapping using Sentinel-2 10 m resolution imagery. *Applied Sciences*, *10*(15), 5075.
- Han, S., Zhao, Y., Cheng, J., Zhao, F., Yang, H., Feng, H., Li, Z., Ma, X., Zhao, C., & Yang, G. (2022). Monitoring key wheat growth variables by integrating phenology and UAV multispectral imagery data into random forest model. *Remote Sensing*, *14*(15), 3723.
- Hao, J., & Ho, T. K. (2019). Machine learning made easy: A review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348–361.
- Hubert-Moy, L., Thibault, J., Fabre, E., Roze, C., Arvor, D., Corpetti, T., & Rapinel, S. (2019). Time-series spectral dataset for croplands in France (2006–2017). *Data in Brief*, *27*, 104810.
- Jamil, M., Rehman, H., Saqlain Zaheer, M., Tariq, A., Iqbal, R., Hasnain, M. U., Majeed, A., Munir, A., Sabagh, A. E., Habib ur Rahman, M., Raza, A., Ali, M. A., & Elshikh, M. S. (2023). The use of Multispectral Radio-Meter (MSR5) data for wheat crop genotypes identification using machine learning models. *Scientific Reports*, *13*(1), 19867.
- Jamil, M., ul Rehman, H., SaleemUllah, Ashraf, I., & Ubaid, S. (2023). Smart techniques for LULC micro class classification using landsat8 imagery. *Computers, Materials & Continua*, *74*(3), 5545–5557. <https://doi.org/10.32604/cmc.2023.033449>
- Khan, S. U., Din, J. U., Qayyum, A., Jaan, N. E., & Jenks, M. A. (2015). Heat tolerance indicators in Pakistani wheat (*Triticum aestivum* L.) genotypes. *Acta Botanica Croatica*, *74*(1), 109–121.
- Nabwire, S., Wakholi, C., Faqeerzada, M. A., Arief, M. A. A., Kim, M. S., Baek, I., & Cho, B.-K. (2022). Estimation of cold stress, plant age, and number of leaves in watermelon plants using image analysis. *Frontiers in Plant Science*, *13*, 847225.
- Naser, M. A., Khosla, R., Longchamps, L., & Dahal, S. (2020). Using NDVI to differentiate wheat genotypes productivity under dryland and irrigated conditions. *Remote Sensing*, *12*(5), 824.
- Noh, J., Kim, J. M., Sheikh, S., Lee, S. G., Lim, J. H., Seong, M. H., & Jung, G. T. (2013). Effect of heat treatment around the fruit set region on growth and yield of watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai]. *Physiology and Molecular Biology of Plants*, *19*, 509–514.
- Panhwar, N. A., Mierzwa-Hersztek, M., Baloch, G. M., Soomro, Z. A., Sial, M. A., Demiraj, E., Panhwar, S. A., Afzal, A., & Lahori, A. H. (2021). Water stress affects the some morpho-physiological traits of twenty wheat (*Triticum aestivum* L.) genotypes under field condition. *Sustainability*, *13*(24), 13736.
- Qadri, S., Furqan Qadri, S., Husnain, M., Saad Missen, M. M., Khan, D. M., Muzammil-Ul-Rehman, Razzaq, A., & Ullah, S. (2019). Machine vision approach for classification of citrus leaves using fused features. *International Journal of Food Properties*, *22*(1), 2072–2089.
- Qadri, S., Khan, D. M., Ahmad, F., Qadri, S. F., Babar, M. E., Shahid, M., Ul-Rehman, M., Razzaq, A., Shah Muhammad, S., Fahad, M., Ahmad, S., Pervez, M. T., Naveed, N., Aslam, N., Jamil, M., Rehmani, E. A., Ahmad, N., & Akhtar Khan, N. (2016). A comparative study of land cover classification by using multispectral and texture data. *BioMed Research International*, *2016*, 8797438. <https://doi.org/10.1155/2016/8797438>
- Raoufi, R., Soufizadeh, S., Amiri Larijani, B., AghaAlikhani, M., & Kambouzia, J. (2018). Simulation of growth and yield of various irrigated rice (*Oryza sativa* L.) genotypes by AquaCrop under different seedling ages. *Natural Resource Modeling*, *31*(2), e12162.
- Raza, A., Munir, K., Almutairi, M. S., & Sehar, R. (2023). Novel class probability features for optimizing network attack detection with machine learning. *IEEE Access*, *11*, 98685–98694. <https://doi.org/10.1109/ACCESS.2023.3313596>

- Raza, A., Rustam, F., Mallampati, B., Gali, P., & Ashraf, I. (2023). Preventing crimes through gunshots recognition using novel feature engineering and meta-learning approach. *IEEE Access*, *11*, 103115–103131. <https://doi.org/10.1109/ACCESS.2023.3316695>
- Raza, A., Siddiqui, H. U. R., Munir, K., Almutairi, M., Rustam, F., & Ashraf, I. (2022). Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *PLOS One*, *17*(11), e0276525.
- Rehmani, E., Naweed, M., Shahid, M., Qadri, S., & Gilani, Z. (2015). A comparative study of crop classification by using radiometric and photographic data. *Sindh University Research Journal (Science Series)*, *47*(2), 335–340.
- Sandhu, K., Patil, S. S., Pumphrey, M., & Carter, A. (2021). Multitrait machine-and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, *14*(3), e20119.
- Shah, M. A. A., Mohsin, M., Chesneau, C., Zulfiqar, A., Jamal, F., Nadeem, K., & Sherwani, R. A. K. (2020). Analysis of factors affecting yield of agricultural crops in bahawalpur district: Analysis of factors of major agricultural crops. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences*, *57*(4), 99–112.
- Sharma, B. P., Zhang, N., Lee, D., Heaton, E., Delucia, E. H., Sacks, E. J., Kantola, I. B., Boersma, N. N., Long, S. P., Voigt, T. B., & Khanna, M. (2022). Responsiveness of miscanthus and switchgrass yields to stand age and nitrogen fertilization: A meta-regression analysis. *GCB Bioenergy*, *14*(5), 539–557.
- Shekar, B., & Dagnew, G. (2019). Grid search-based hyperparameter tuning and classification of microarray cancer data. In *2019 second international conference on advanced computational and communication paradigms (ICACCP)* (pp. 1–8). IEEE.
- Wyatt, J. (2016). Grain and plant morphology of cereals and how characters can be used to identify varieties. *Encyclopedia of Food Grains (Second Edition)*, *1*, 51–72.
- Yang, J., Spicer, R. A., Spicer, T. E., Arens, N. C., Jacques, F. M., Su, T., Kennedy, E. M., Herman, A. B., Steart, D. C., Srivastava, G., Mehrotra, R. C., Valdes, P. J., Mehrotra, N. C., Zhou, Z.-K., & Lai, J.-S. (2015). Leaf form–climate relationships on the global stage: An ensemble of characters. *Global Ecology and Biogeography*, *24*(10), 1113–1125.
- Zahra, N., Hafeez, M. B., Wahid, A., Al Masruri, M. H., Ullah, A., Siddique, K. H., & Farooq, M. (2023). Impact of climate change on wheat grain composition and quality. *Journal of the Science of Food and Agriculture*, *103*(6), 2745–2751.
- Zhang, X., Liu, K., Wang, S., Long, X., & Li, X. (2021). A rapid model (COV_PSDI) for winter wheat mapping in fallow rotation area using MODIS NDVI time-series satellite observations: The case of the Heilonggang region. *Remote Sensing*, *13*(23), 4870.

How to cite this article: Jamil, M., Ahsan, Z., Saeed, M. N., Raza, A., Migdady, H., Daoud, M. S., Altalhi, M., Ezugwu, A. E., & Abualigah, L. (2024). Wheat crop genotype and age prediction using machine learning with multispectral radiometer sensor data. *Agronomy Journal*, 1–12. <https://doi.org/10.1002/agj2.21595>