



OPEN The use of Multispectral Radio-Meter (MSR5) data for wheat crop genotypes identification using machine learning models

Mutiullah Jamil¹, Hafeezur Rehman¹, Muhammad Saqlain Zaheer², Aqil Tariq^{3,4}, Rashid Iqbal⁵, Muhammad Usama Hasnain⁶, Asma Majeed⁷, Awais Munir⁷, Ayman El Sabagh^{8,9}, Muhammad Habib ur Rahman^{6,10}, Ahsan Raza^{10,11}, Mohammad Ajmal Ali¹² & Mohamed S. Elshikh¹²

Satellite remote sensing is widely being used by the researchers and geospatial scientists due to its free data access for land observation and agricultural activities monitoring. The world is suffering from food shortages due to the dramatic increase in population and climate change. Various crop genotypes can survive in harsh climatic conditions and give more production with less disease infection. Remote sensing can play an essential role in crop genotype identification using computer vision. In many studies, different objects, crops, and land cover classification is done successfully, while crop genotypes classification is still a gray area. Despite the importance of genotype identification for production planning, a significant method has yet to be developed to detect the genotypes varieties of crop yield using multispectral radiometer data. In this study, three genotypes of wheat crop (Aas-'2011', 'Miraj-'08', and 'Punjad-1) fields are prepared for the investigation of multispectral radio meter band properties. Temporal data (every 15 days from the height of 10 feet covering 5 feet in the circle in one scan) is collected using an efficient multispectral Radio Meter (MSR5 five bands). Two hundred yield samples of each wheat genotype are acquired and manually labeled accordingly for the training of supervised machine learning models. To find the strength of features (five bands), Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Nonlinear Discernment Analysis (NDA) are performed besides the machine learning models of the Extra Tree Classifier (ETC), Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), k Nearest Neighbor (KNN) and Artificial Neural Network (ANN) with detailed of configuration settings. ANN and random forest algorithm have achieved approximately maximum accuracy of 97% and 96% on the test dataset. It is recommended that digital policymakers from the agriculture department can use ANN and RF to identify the different genotypes at farmer's fields and research centers. These findings can be used for precision identification and management of the crop specific genotypes for optimized resource use efficiency.

¹Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan. ²Department of Agricultural Engineering, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan. ³Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, 775 Stone Boulevard, Mississippi State, MS 39762-9690, USA. ⁴State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China. ⁵Department of Agronomy, Faculty of Agriculture and Environment, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. ⁶Institute of Plant Breeding and Biotechnology, MNS-University of Agriculture, Multan, Pakistan. ⁷Institute of Agro-Industry & Environment, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. ⁸Department of Agronomy, Faculty of Agriculture, Kafrelsheikh University, Kafr El-Shaikh 33516, Egypt. ⁹Department of Field Crops, Faculty of Agriculture, Siirt University, Siirt, Turkey. ¹⁰Crop Science, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany. ¹¹Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374 Müncheberg, Germany. ¹²Department of Botany and Microbiology, College of Science, King Saud University, 11451 Riyadh, Saudi Arabia. ✉email: at2139@msstate.edu; rashid.iqbal@iub.edu.pk; araza@uni-bonn.de

Timely and precise crop yield estimation is the pre-request to define the food availability of a nation. In the modern world, every agricultural country can acknowledge the importance of precise knowledge of crop conditions for plant management. It is the most critical single economic sector, contributing 24% of the national income of Pakistan. Its significance prevails as 48% of the working population is engaged in this sector. Punjab is the country's primary producer of cash crops^{1–3}. In such circumstances, improving agricultural land monitoring is among the most critical and pressing prominent issues that Pakistan must address. Crop patterns have their importance as they are used to develop regional strategies and programs to increase farm production and efficient use of land resources^{4–6}.

One agricultural production goal is maximizing crop yield while minimizing costs. Early detection and management of seasonal crop yield indicator problems can help to increase production and subsequent profit. Crop yield spatial variability can be assessed using remote sensing and global positioning systems (GPS)^{7–13}. Recently, crop yield prediction has been observed in various literature before harvest^{14,15}. Crop yield spectral characteristics without introducing weather noise. A Multispectral Radio Meter with five bands (MSR5) is used for land use land cover (LULC) using K Nearest Neighbor (KNN)^{14,16,17}.

Various types of crops such as sugar cane, potato, tobacco, and land change detection can be made using remotely sensed data such as MSR5, Unmanned Aerial Vehicle (UAV), Landsat 8, and Sentinel. The use of multispectral radiometers in agriculture is becoming a novel approach to identifying insects, pests, and genotypes as well. It is a digital sensor optics with high microelectronics that provides a spectral record of the light to identify the objective.

Although, satellite or Aerial Remote Sensing (ARS) technology can significantly improve the present systems to acquire and generate agricultural resource data^{18–20}. In most of the literature, researchers focus on the identification of different classes (Bare land, water, sugarcane, tomato, potato, etc.) of LULC classification. Another issue in developing countries such as Pakistan, Afghanistan, and India tried to cultivate regional trended crop yield genotypes of wheat and rice, the information of trending crops is gathered through a Decision Tree Classification (DTC) approach that is costly, time-consuming, and with a high error rate but it is not being considered with machine learning methods^{21–23}. Pakistani wheat yield has increased (more than 1%)²⁴. New varieties and genotypes have had an important role in improving crop production in recent decades²⁵. Now it is necessary to adopt new management technologies and policies to adopt the new varieties.

Genotype crop classifications are still ignored while varieties of various crops have a high impact on increasing the production of crop yield^{5,6,17,26–30}. The DTC and RF^{14,31}, models involve various parameters: weather conditions, water stress, rainfall, air humidity, and temperature^{7,32,33}. Wheat varieties are an important parameter that is absent from the Artificial Neural network (ANN) and KNN models due to the unavailability of an identification method. That is the reason; this work mainly focuses on the robust machine-learning model for the identification of wheat genotype and variety.

Multispectral radiometers capture the electromagnetic radiation reflected or emitted by objects in different spectral bands, allowing for the collection of valuable crop-related information. These sensors can capture various spectral bands, ranging from visible to near-infrared and thermal infrared, providing a wealth of data about crop health, vigor, and other important characteristics. Machine learning models, such as support vector machines (SVMs), random forests, and artificial neural networks, have demonstrated great potential in analyzing and interpreting complex data patterns^{34–36}. By training these models on multispectral radiometer data, it becomes possible to develop robust and accurate classification models for identifying different wheat crop genotypes. Number of researchers use different machine learning methods used for classification purposes.

Integrating machine learning models Support Vector Machine (SVM)³⁷, Extra Tree Classifier (ETC)³⁸, Logistic Regression (LR)³⁹, KNN⁴⁰ and Decision Tree (DT)¹⁴ with multispectral radiometer data can potentially revolutionize wheat crop management practices. Accurate and rapid genotype identification can significantly aid plant breeders and agronomist in selecting superior genotypes with desired traits, improving crop yield and quality. Moreover, this approach can contribute to precision agriculture by enabling targeted interventions, such as optimized fertilizer application, pest management, and irrigation strategies.

The application of machine learning models across several domains has extended to the agricultural sector, resulting in notable advantages for this industry. For example, Fei proposed an ensemble framework for wheat yield prediction with different water treatments⁴¹. The authors developed Elastic Net Regression (ELR) for the prediction and deployed it with selected features. The proposed ELR achieved 0.729 R² scores by combining the predicted values of all growth stages. This study proposed a machine learning approach for wheat yield prediction using advanced sensing techniques^{16,25,42}. The authors proposed XY-fused Networks (XY-Fs), supervised Kohen networks, and counter-propagation artificial neural networks (CP-ANNs) models for this purpose and achieved 81.65% accuracy using supervised Kohen networks.

The machine learning model, the satellite, and climatic data are integrated to predict the wheat crop yield⁴³. The source data taken from 2000 to 2014 from Australia is used at the statistical division level. They deployed LASSO, RF, neural networks, and SVM models for the prediction of a significant accuracy score of 0.75 R². The winter wheat yield prediction using machine learning models from multi-source data in China⁴¹. They combined climate, soil, and remote sensing data to predict winter wheat yield based on the Google Earth Engine. The achieved R² > 0.75 with SVM, RF, and Gaussian process regression.

In machine learning models, the solar-induced chlorophyll fluorescence data is used to predict the wheat yield⁴⁴. They deployed LASSO, extreme gradient boosting, Support Vector Regression (SVR), ridge regression, RF regression, and Long Short Term Memory (LSTM). SVR outperforms all other models as well as the deep learning models LSTM with a significant R² of 0.87. UAV hyperspectral and ensemble machine-learning approaches to predict the wheat yield⁴⁵. Three techniques of feature selection such as Brute feature selection, recursive feature elimination, and the Pearson correlation coefficient. They combined four machine-learning models to

make an ensemble: SVM, RF, Gaussian process, and linear ridge regression. The ensemble model achieved a 78% accuracy score.

Environmental and phenological data can predict winter wheat yield using convolutional neural networks⁴⁶. They collect data from 271 counties in Germany and deploy several machine learning and deep learning models. The proposed model convolutional neural networks achieved 7–14% lower RMSE and 3–15% lower MAE. An approach for wheat yield prediction using kernel ridge regression and Satellite-derived predictors⁴⁷. They combined kernel ridge regression, complete ensemble empirical mode decomposition with adaptive noise (CEEM-DAN), and the grey wolf optimizer (GWO-CEEMDAN-KRR). Compared to baseline models, the proposed model reduces the error rate by 20%. Similarly, an approach for regional and local-scale wheat yield prediction using RF in Australia⁴⁸. RF achieved a significant 0.89 R^2 score for Victoria region data.

Multispectral images were collected from a UAV platform to monitor maize growth and nutritional status⁴⁹. The researchers apply radiometric calibration and establish linear regression relationships between SPAD values and spectral/textural indices. Machine learning models, specifically support vector machine (SVM) and random forest (RF), are employed to estimate SPAD values, with SVM performing better ($R^2 = 0.81$, RMSE = 0.14). A comprehensive review of the application of machine learning in agricultural production systems⁵⁰. The review covers various areas such as crop management, livestock management, water management, and soil management. Machine learning techniques include yield prediction, disease detection, weed detection, crop quality assessment, species recognition, and management systems that offer valuable insights and recommendations for informed decision-making by farmers. It is hard for humans to estimate and analyze the crop condition to take the necessary action to save resources with maximum output. In the current era, satellite communication costs have become cheap, and it is the best way to monitor objects and earth situations with increased efficiency and precision.

This research study aims to fill the existing gap in the literature by investigating the impact of multispectral radiometer data on wheat crop genotype identification using state-of-the-art machine learning models. The study will evaluate the performance of different machine learning algorithms, assess the effectiveness of feature extraction techniques, and analyze the influence of varying environmental conditions on classification accuracy. The outcomes of this research can have significant implications for wheat breeding programs, precision agriculture, and crop management practices. By harnessing the power of multispectral radiometer data and machine learning models, accurate and efficient genotype identification can contribute to sustainable agriculture, food security, and the optimization of wheat crop production.

Materials and methods

Study area

The study area is the agricultural research center under the Islamia University of Bahawalpur in Bahawalpur City, Punjab, Pakistan, as shown in Fig. 1. For the present study, the site is located at latitude 29°22'18" N and longitude 71°46'03" E in the agriculture forms of The Islamia University of Bahawalpur^{51–54}. The temperature

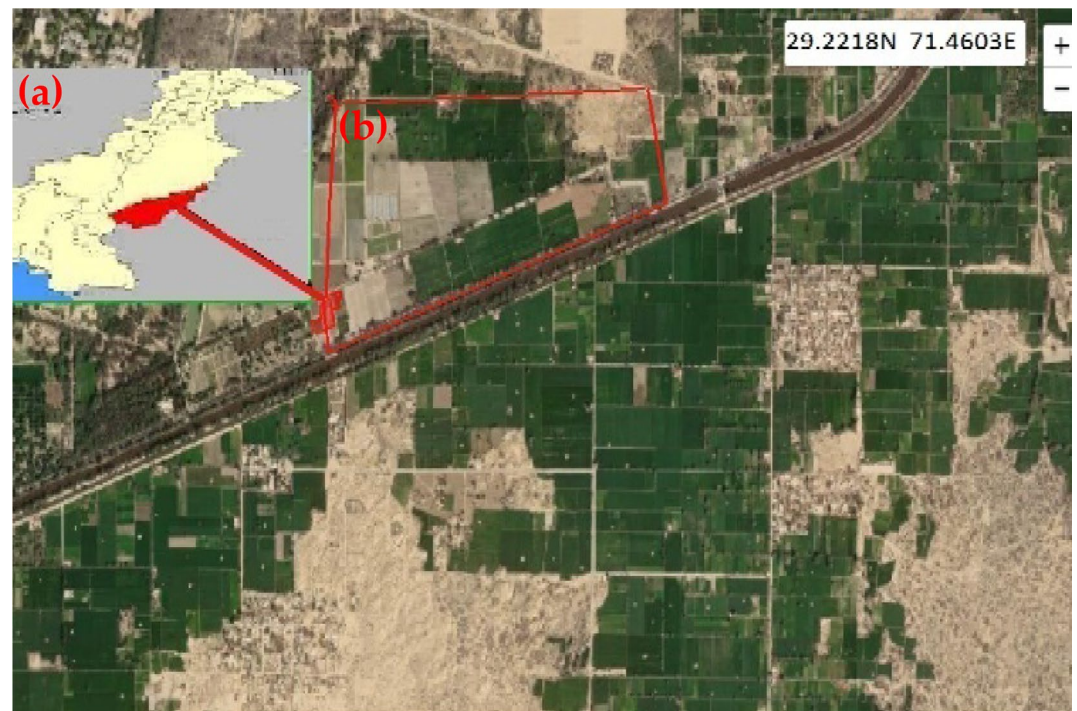


Figure 1. (a) Location of the study site using Google earth view with the map of Pakistan and (b) image highlighted in red color ROI at the upper top left corner of the image. Google Earth 6.0, (2022).

of Bahawalpur is extremely high, and it faces a water stress problem in most of the regions⁵⁵. The study region is very diverse and incorporates Punjab agro-climates with a minimum rainfall range of 2 mm/month in the driest month. October is the driest month, and July is the wettest month, with rainfall of 61 mm. Extremely high temperatures and rain intensity cause much of the rainfall evaporation and runoff⁵⁶. In dry areas, water stress plays an important role in decreasing the production of wheat crops.

Data acquisition and preprocessing

Multispectral radiometric datasets

Satellite remote sensing technology is the modern technology in the development of RS technology, but it has large information with a low resolution of the image. UAVs have low-resolution images compared to satellite images, which cover a large area with special resolution. To accomplish the goal, we need to investigate the power of spectral bands to categorize the genotypes of the wheat crop. Since, we are growing a small amount of wheat for this experiment, using a handheld device is the most efficient way to collect the necessary data⁵⁷. To acquire multispectral radiometric data, the three wheat varieties (genotypes) that were approved by the Punjab seed certification department, Punjab named 'Aas-'2011', 'Miraj-'08', and 'Punjnad-'1' were harvested in three different adjacent plots. Seed grains were provided by the Agricultural Research Department of the Islamia University of Bahawalpur, and each type of wheat variety was kept under the observation of an agricultural research expert. In this way, three plots were built, Ass-2011, Miraj-08, and Punjnand-1, in a row (1 × 3) plot of equal size of 225 square feet each. Additionally, it is ensured that the same human expert does all the harvested processes to reduce the other cropping factors like water, preparation of land, and nutrition supply to the crops.

Various researchers used multispectral radiometers for recording the incoming radiation and light reflectance from the canopy in five spectral bands, similar to Landsat 8 (OLI/TIRS) and Landsat 7 (ETM+) satellites^{58,59}. The output data consists of five bands, detail of which is given in Table 1. Each band has a half-peak band of approximately 5–15 nm, depending on the specific band. In this way, MSR5 describes a complete scene based on five numeric digits, i.e., five energy bands. Previous research shows that only a combination of five bands can classify a complete captured scene. This device has already been used for crop classification⁵⁴ and to efficiently measure nitrogen contents and biomass in plants⁶⁰. Table 1 shows the wavelength and spatial resolution for the wheat crop scan used in this study. To assess the crop field data attained at six stages using crop scan MSR5 (for radiometric data) was acquired from different regions of the crop field.

Six hundred (600) scans from three fields of the crops as mentioned above i.e., three wheat varieties (Ass-2011, Miraj-08, and Punjnand-1), have been acquired at 10 feet from the ground level. The scanned data was stored in the memory of the Data Logger Controller (DLC) device. It was then transferred to a CSV file to analyze the data by using the routines provided by the vendor of MSR5.

Field sample data

The research utilized GPS field surveys and Google Earth images as reference points. It was determined that there are three distinct types of wheat. Visual interpretation of field validation and images from Google Earth were used to select the samples. After that, the ground sample points were arbitrarily divided into sections (80% training and 20%), and the accuracy was computed. The 240 samples out of 300 of each wheat crop variety are taken for training for the possibility of inter-classification of wheat crops using MSR5 data. At the same time, 20% of the whole dataset is randomly selected as test data. RF, SVM, and customized ANN have been selected for classification.

Crop classification method

Features described for crop classification

Previous research demonstrated that the utilization of spectral information to derive the mean, standard deviation, and variation of each band can differentiate between the many characteristics that are associated with crop varieties^{16,61–65}. This information is related to the structure of the target surface and the surrounding environment, which can also indicate spatial variation in land cover. So, statistical, structural, and spectral methods can be used to pull out the information about the texture. Previous research has shown that using spectral data to figure out each band's mean, standard deviation, and variation is a good way to find the difference between the many characteristics of different crop varieties⁶⁶. Significant evidence suggests that the identification of crops can benefit significantly from the use of textural characteristics derived from satellite images^{66,67}. The information about the crop's texture depicts the crop's density as well as its shape. The spectral information of red-edge bands in the MSR52 data demonstrates a possible performance use in determining the growing state of crops.

Spectral band	Wavelength (nm)	Spatial resolution in radios (meter)
Band 1 Blue	450–520 nm	1.524
Band 2 Green	520–630 nm	1.524
Band 3 Red	630–690 nm	1.524
Band 4 SNIR	760–900 nm	1.524
Band 5 FNIR	1550–1750 nm	1.524

Table 1. Wavelength and spatial resolution of the crop scan MSR5 were used in this study.

Classification and assessment accuracy

The retrieved characteristics were used in conjunction with three different advanced machine learning and classification approaches, namely SVM, ANN, and RF. The SVM technique seeks to determine the ideal hyperplane in the n -dimensional space used for classification in order to maximize the margin of separation between classes (the crops)^{68–70}. We employed the SVM classifier by making use of LIBSVM and a radial basis function (RBF) kernel⁶⁸. ANN is able to imitate the recognition structure of the human brain and nervous system while maintaining a high degree of non-high linear classification ability^{31,71}. One sort of neural network that sees widespread use is known as the multi-layered perceptron. This particular variety of ANN typically consists of three or more layers that can partition nonlinear data^{72,73}. It is usual practice to represent the RF classifier as an ensemble of decision trees, with voting serving as the mechanism for assigning class labels. It is capable of dealing with high-dimensional data and is resistant to overfitting to a certain extent⁷⁴. RF is also used to assess the relevance of characteristics in the classification process. These features include texture, spectral, and indices features^{14,75–78}.

This study uses CROPSCAN DATA Inc. 2018 MSR5 multispectral radiometer sample data to train the machine learning model RF, SVM, and ANN with various settings for three different types of wheat crops. A photographic representation of each stage is given in Fig. 2.

Five columns of Table 2, namely “B”, “G”, “R”, “NIR”, and “SIR” represent reflectance bands of the cropped image. The last column of Table 2 represents the labels of three varieties. Figure 3 shows the methodology adopted for wheat crop classification. The collected data is preprocessed, cleaned, and annotated manually for machine learning models.



Figure 2. Photographic representation of Wheat Crop of six stages of MSR5 scan data stage 1–3 left to right in the first row and 4–6 in the second row.

Sr. No.	B	G	R	NIR	SIR	Class
1	2.38	4.23	2.66	48.29	13.28	Aas-2011
2	2.06	4.3	2.51	47.95	13.36	Miraj-08
3	2.57	4.19	3.23	39.4	12.29	Punjnad-1

Table 2. Sample data of multispectral data.

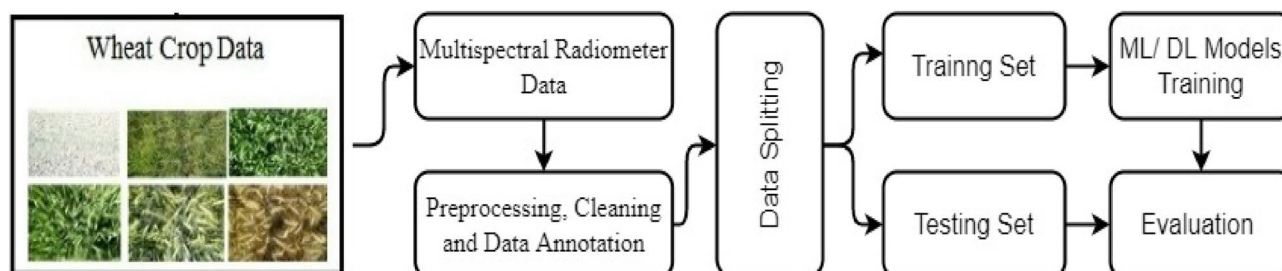


Figure 3. The method used to classify the genotype wheat varieties.

According to the literature of the last five years published on remote sensing, satellite images of Landsat 7, 8, and Sentinel 2 datasets are trained by RF, SVM, and ANN predominantly, so this study selects these models^{14,31,48,51,79}. Most of the researchers prefer to use RF and SVM instead of deep learning models. Usually, a deep learning model is used to enhance the accuracy of classification or mapping of high-resolution images or NDVI image segmentation^{80,81}. In this research, the primary focus is to find the strength of five bands of MSR5 for the inter-classification of wheat crop varieties so that popular machine learning models are implemented as mentioned above.

We used Python statistical software and the Scikit-learn package to implement these classification techniques⁷⁶. Further, thirteen feature scenarios were tested with the machine learning methods. The classification accuracy is reported for each scenario, and the classification results were compared based on accuracy in crop mapping for each crop class, as described in section “Classification results and accuracy assessment by ML models”. Finally, we calculated a confusion matrix for each classification result based on the ground control points. Then, the overall accuracy (OA), Kappa coefficient, producer’s accuracy (PA), and user’s accuracy (UA) were calculated to evaluate the classification results^{76,82}. Another commonly used performance evaluators are accuracy, Precision, and recall, in which accuracy indicates how many of the total predictions were correct. Precision, also known as positive predictive value, tells how many positively predicted instances were actually true.

In contrast, recall, also known as sensitivity or true positive rate, measures how many of the actual positive instances were correctly predicted as positive. Mathematical formulas are given in Eqs. (1, 2 and 3), respectively.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100\% \quad (3)$$

The *F1* measure (Eq. 4) was calculated to evaluate the effectiveness of the crop classification^{83–88}. The *F1* and overall accuracy are considered more meaningful than the Kappa coefficients. The value range of *F1* is from 0 to 1—the larger the *F1* score is, the more accurate the classification results are. The *F1* score is the harmonic mean of U and P as shown in (Eq. 4):

$$F1 = 2 \times \frac{P \times U}{U + P} \quad (4)$$

An additional parameter for image classification accuracy is the Figure of Merit (FoM)^{89,90}. The FoM computes from omission, commission, and overall agreement (Eq. 5):

$$FoM = \frac{\alpha}{o + \alpha + c} \times 100\% \quad (5)$$

In the Eq. (2), α represents overall agreement, o represents overall omission numbers, c represents overall commission numbers.

Plant guidelines

All the plant experiments were in compliance with relevant institutional, national, and international guidelines and legislations.

Results

Dimensionality reduction techniques with graphical representation of data clusters

Principal component analysis (PCA), linear discriminant analysis (LDA), and nonlinear discriminant analysis (NDA) are popular feature reduction techniques with maximum classification accuracy. It can map the input data from the original space to the new feature space so that all classes are duly clustered and well separated using top-ranked minimum features. These are implemented with MSR5 data, which is normalized by dividing the maximum value found in the data. Its graphical representation is given in Fig. 4, and obtained 93%, 94%, and 94% classification accuracy PCA, LDA, and NDA respectively. It means we can train the ML model and achieve more than 94% accuracy, as shown in Table 7.

Classification results and accuracy assessment by ML models

Several researchers published their work in remote sensing and LULC classification using RF, SVM, and ANN machine-learning models for extra classification. Therefore, in the intra-classification of wheat crop varieties classification, we implemented the ANN back propagation machine-learning model and did an empirical analysis of the various configuration of ANN, like the number of iterations, learning rate, and several hidden layers. Detailed experiment results of configuration, training, and testing accuracy percentage are given in Table 3.

Table 3 shows that the learning rate (η) can play an important role in getting the maximum local value of accuracy, which can achieve a very small change of (η) from 0.01 to 0.15. There is no need to jump from 0.05 to 0.99 maximum because the algorithm is very sensitive to small changes. On the other hand, it is observed that after 0.01 to 0.20 outcome of the algorithm is repeated rather than improved in terms of training and testing

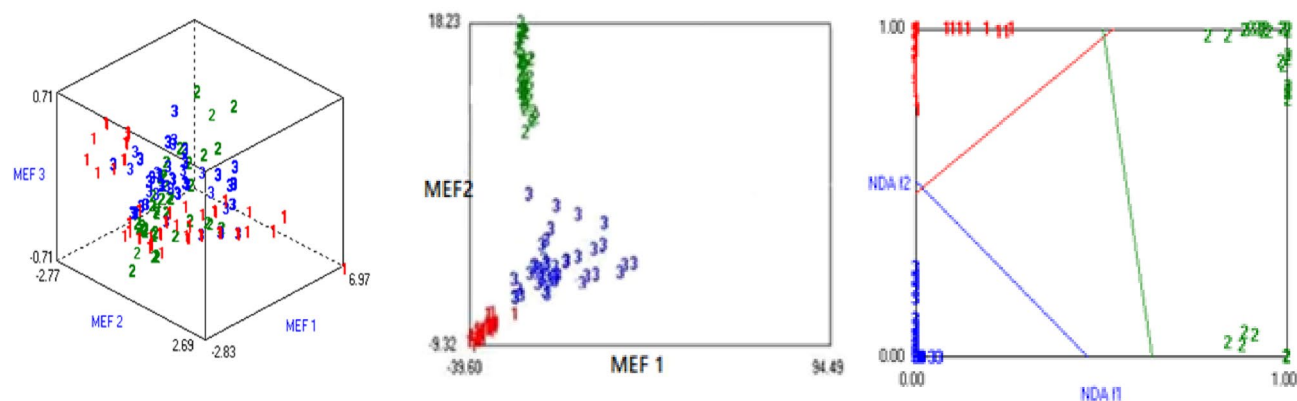


Figure 4. A graphical representation of PCA, LDA, and NDA left to right with Most Expressive Feature (MEF 1, 2 and 3) where 1 = Aas-2011, 2 = Miraj-8 and 3 = Punjnad-1.

S. No.	1st hidden layer	2nd hidden layer	(η)	Iterations	Training accuracy	Testing accuracy
1	1	2	0.05	50	32.50	33.33
2	1	2	0.15	50	64.20	65.55
3	1	2	0.25	50	32.40	33.33
4	1	2	0.35	50	65.50	66.67
5	1	2	0.45	50	57.77	58.33
6	1	2	0.55	50	63.75	76.77
7	1	2	0.65	50	63.45	65.00
8	1	2	0.75	50	63.45	58.33
9	1	2	0.85	50	57.44	61.33
10	1	2	0.99	50	56.70	58.33
11	2	2	0.05	100	94.60	90.00
12	2	2	0.15	100	95.50	90.00
13	2	2	0.25	100	93.40	89.33
14	2	2	0.45	100	95.90	90.00
15	3	2	0.05	100	96.40	90.00
16	3	2	0.15	100	96.40	84.00
17	3	2	0.45	100	95.00	80.00
18	4	2	0.15	100	97.90	85.00
19	4	2	0.25	100	97.91	96.67
20	4	2	0.35	100	97.92	45.00
21	4	2	0.35	300	95.80	51.77
22	4	2	0.35	50	96.70	70.00
23	5	2	0.15	100	97.92	85.00

Table 3. Training and testing accuracy of ANN at various configurations.

accuracy. A number of the first hidden layers are impotent to enhance training and testing accuracy. It is analyzed that when the number of the first hidden layer is increased, the algorithm gives its maximum performance in terms of accuracy with an (η) rate of 0.05 or 0.15 or a maximum at 0.25. When we reached the PCA accuracy, no improvement was found due to the increased number of the hidden layer. Maximum training accuracy is obtained in rows number 20 and 23 in Table 3, also which can be observed in Fig. 5 but the model is over-trained because testing accuracy moves down from 96 to 85%. It means that the model is leading to overfitting.

Neural Network back propagation gives maximum accuracy of 97% and 96% in training and testing datasets, respectively, with a learning rate (η) of 0.25. The confusion matrix of the training and testing data set is given in Table 4, the result of the random forest, and the random forest result in Table 5.

Comparisons of results with models evaluation

To check the model's performance, we implemented another well-known machine-learning model also used in previous research⁹¹. Compared the machine learning models' performance with the proposed ANN model to show the significance of ANN⁹². We used ETC, RF, SVM, DT, LR, and KNN. We deploy these models with their best hyper-parameters settings. RF, ETC are used with 300 estimators indicating that 300 decision trees will

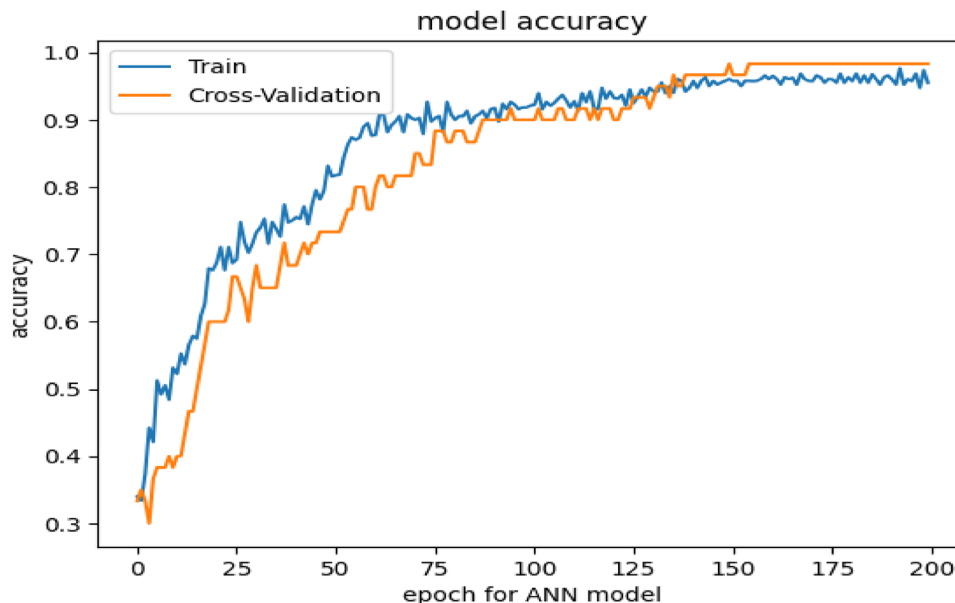


Figure 5. Accuracy score of ANN Machine learning models.

Phase	Class	Aas-2011	Miraj-08	Punjnad-1
Training	Aas-2011	80	0	0
	Miraj-08	0	80	0
	Punjnad-1	5	0	75
Testing	Aas-2011	20	0	0
	Miraj-08	0	20	0
	Punjnad-1	1	1	18

Table 4. Confusion matrix of ANN with training and testing dataset.

Class	Aas-2011	Miraj-08	Punjnad-1
Aas-2011	76	0	4
Miraj-08	0	80	0
Punjnad-1	3	1	76
Correctly classified instances	232		97.0533%
Incorrectly classified instances	8		2.9467%
Kappa statistic	0.8972		
MSE	0.1168		
RMSE	0.169		
Total number of instances	240		

Table 5. The result of the Random Forest decision tree model on train dataset.

be used for weak learners, and each tree will grow to a maximum of 10 level depth because we used ‘max_depth’ parameters with a value of 10. DT is used with only the ‘max_depth’ parameter, which will restrict each model to grow up to a maximum 10-level depth to reduce complexity and overfitting. SVM is used with linear kernel, and LR is used with saga solver. Hyper-parameters for all machine learning models are provided in Table 6.

The results of machine learning models are presented in Table 7; each model’s confusion matrix for detailed accuracy of each class is shown in Fig. 6. According to the results, the performance of machine learning models is also good as tree-based models RF, ETC perform significantly better with 96% and 95% accuracy scores, respectively. RF, ETC are tree-based ensemble models that perform significantly even on small-size datasets. LR and SVM show poor performance because they need a large feature set for the good fit of models.

Model	Hyper-parameters
ETC	n_estimators = 300, max_depth = 10
RF	n_estimators = 300, max_depth = 10
SVM	Kernel = linear, C = 1.0
DT	max_depth = 10
LR	Solver = saga, C = 1.0

Table 6. Hyper-parameters are used for machine learning models.

Model	Accuracy	Precision	Recall	F1 score
ETC	0.95	0.94	0.91	0.89
RF	0.96	0.93	0.92	0.87
SVM	0.84	0.81	0.83	0.82
DT	0.85	0.84	0.83	0.80
LR	0.63	0.71	0.63	0.65
KNN	0.73	0.72	0.78	0.75

Table 7. Performance of machine learning models.

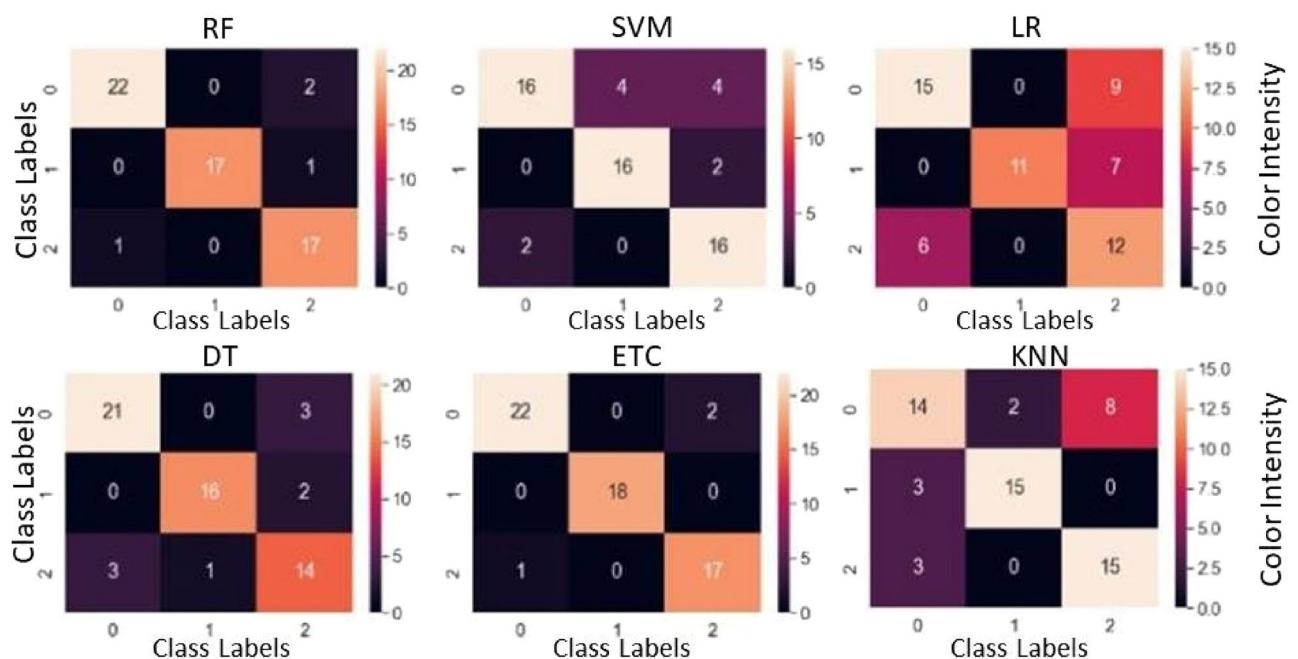


Figure 6. Confusion matrices for machine learning models where the “0” label represent Aas-2011, the “1” label represents Miraj-08 and the “2” label represent Punjnad-01.

The accuracy of the RF algorithm is the second highest in terms of accuracy with Kappa statistics. These studies^{93,94} compared their results with Kappa statistics with less than 88% satisfaction. At the same time, the results show that the samples are more consistent and reliable as compared to the above researchers. It means that MSR5 data has more potential to classify at the micro class classification level without any overlapping of various types of varieties with a minimum error rate. SVM is also a well-known classifier used in LULC classification^{95,96} with various kernels. For this research, the performance of SVM is not significant as its accuracy is 80%, which is only better than KNN and LR.

We deploy several deep learning models to predict wheat yield varieties, such as long short-term memory (LSTM), convolutional neural networks (CNN), and CNN-LSTM. These models are used in comparison with the proposed ANN model. Each model consists of an embedding layer with a vocabulary size of 100,000 and output dimensions of 200. After the embedding layer, the LSTM model contains a dropout layer with a 0.5 dropout rate, which will randomly remove 50% of neurons to reduce the complexity. The LSTM layer with 100 units is

followed by the 100 units and in the end, the LSTM model has a dense layer with three neurons and a Softmax function. CNN model contains a 1D convolutional layer after embedding layer with 128 filters, 3 × 3 kernel size, and ReLU (rectified linear unit). A max-pooling layer with a 3 × 3 pool size is used after the 1D convolutional layer to extract the important feature set. The max-pooling layer follows ReLU activation layer and then a dropout layer is used with a 0.5 dropout rate. A flattening layer is used to convert 3-dimensional data into a 1-dimensional layer. In the end, we use a dense layer with three neurons and a Softmax function. For CNN-LSTM, after the embedding layer we used 1D convolutional layer with a max-pooling layer and activation layer then we used the LSTM layer with 100 units. Similarly, in the end, we used a dense layer with three neurons and a Softmax function. We compile all models with Adam optimizer and 'categorical cross-entropy' loss function. We fitted each model with 200 epochs. The accuracy, precision, recall, and F1 score of deep learning models are given in Fig. 7.

Table 8 contains the results for the deep learning models, which indicate that LSTM achieved 83% accuracy and CNN achieved 88% accuracy, which is better than LSTM. The performance of CNN-LSTM is not good as compared to individual CNN. Overall, the performance of LSTM, CNN, and CNN-LSTM is not good compared to ANN because these models require a large dataset with a large feature set.

Figure 8 shows the confusion matrices for wheat crop variety prediction for LSTM, CNN, and CNN-LSTM models. It can be observed that the number of highest correct predictions come from the CNN model, followed by the CNN-LSTM while the LSTM model gives the lowest number of correct predictions. On average, the performance of deep learning models is inferior to machine learning models^{97–99}.

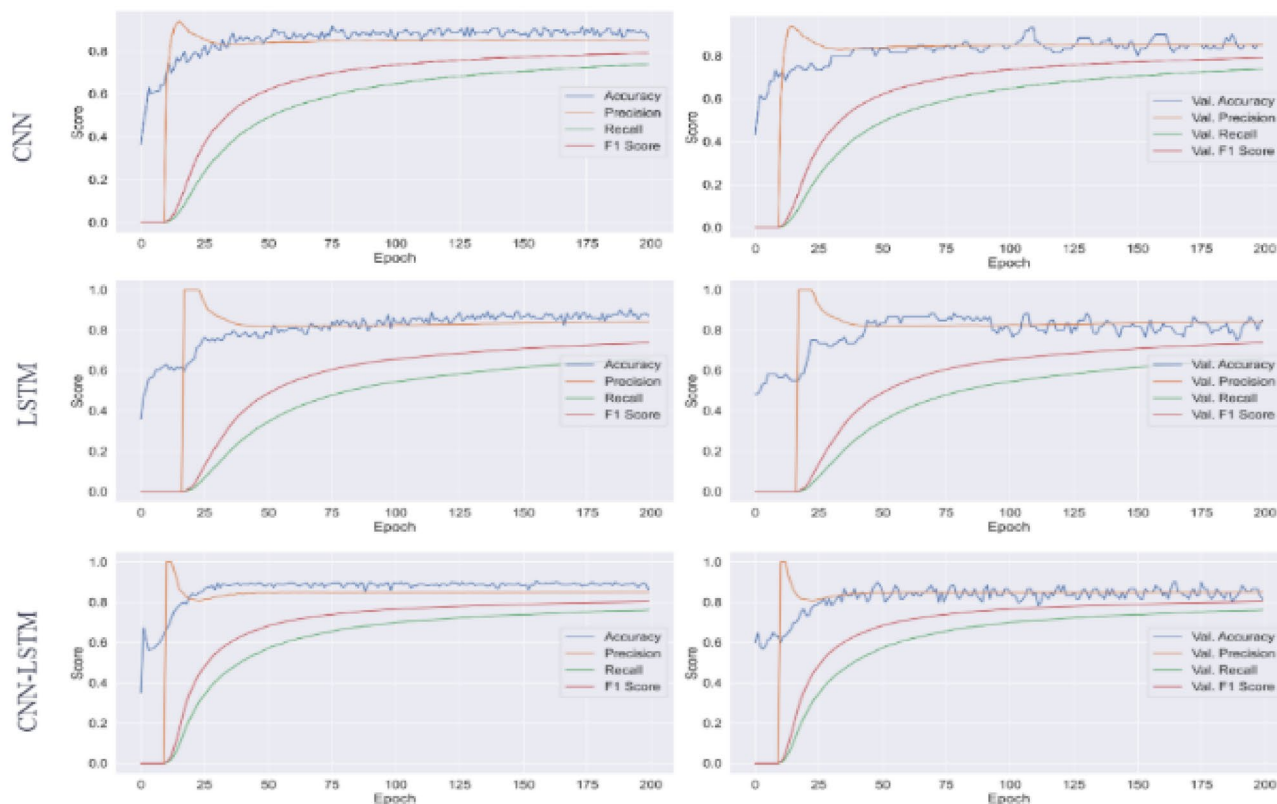


Figure 7. The accuracy, precision, recall, and F1 score of the deep learning models on the training data are presented in the left column, while the results on the testing data are displayed in the right column.

Model	Accuracy	Precision	Recall	F1 Score
LSTM	0.83	0.84	0.80	0.78
CNN	0.88	0.89	0.86	0.84
CNN-LSTM	0.87	0.83	0.84	0.85

Table 8. Performance of deep learning models.

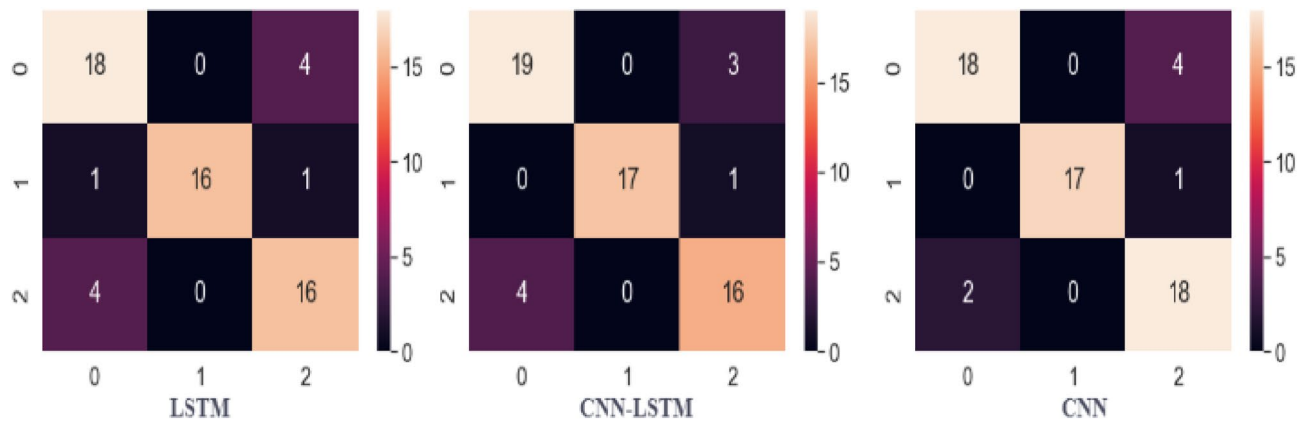


Figure 8. Confusion matrices for deep learning models where the “0” label represents Aas-2011, the “1” label represents Miraj-08, and the “2” label represent Punjnad-01.

Discussion

This case study explores the effectiveness of multispectral radiometers used in remote sensing and crop monitoring system. One of the most important advantages of MSR5 is that it has a fine spatial resolution and is easily implemented in a small study area with a controlled environment^{100,101}. For conducting the pilot study, three plots of wheat crop species were cultivated, and take temporal images after fifteen days to prepare the spectral data set of wheat crop varieties used in remote sensing to find the capacity of MSR5 for micro-class classification. The results are shown in Fig. 4, three clusters of micro-class varieties of wheat crop yield data point using PCA, which indicates that all the sample points are well clustered with low variation within the class. A large distance is also found between class to class ‘Aas-’2011’, ‘Mirage-’08’, and ‘Punjnad-’1’ of wheat varieties. Figure 4 shows that five samples of ‘Aas-’2011’ numbered with labeled one are dispersed from the center of a big cluster of ‘Aas-’2011’. Due to this, the performance of models is slightly affected. Maybe these sample points are recorded with noise due to light or sensor movement during the scanning process or with tree shadow/ appearance of a cloud. However, it shows the maximum potential of MSR5 to classify the wheat crop varieties, which is the first goal of this research^{60,93,94,102–104}.

The second goal of this study is to implement the various traditional machine learning models and try to find the optimal solution in terms of the accuracy and efficiency of the machine learning model achieved by implementing the ANN with various settings. Results show that we can improve the results of ANN by a small increase of the (η) rate, but results are going overfitting or underfitting. So it is proved that (η) rate change greater than 0.5 is a useless activity^{58,59,65}. One to two percent accuracy can be improved by increasing the number of hidden layers that should be less or equal to the number of output classes + 1. There is no need to increase the number of first hidden layers from one to more classes to avoid the overfitting or underfitting of the model. After tuning the ANN compare its performance with a tree-based classifier and support vector machine for doing empirical analysis of various algorithms in which it is analyzed that random forest is the best model in terms of efficiency and ANN is little best in terms of accuracy⁵.

The third goal is achieved by comparing the traditional approach with the classical machine learning method results given in Tables 7 and 8, which show that ANN is better than ETC and CNN, which obtains the best results among machine learning and deep learning models. On the other hand, several researchers apply the classical method for land use land cover classification using spectral images with various indexes of spectral images like NDVI and high-resolution images^{16,59,62}. It is possible that the deep learning models can perform better, with a large dataset with texture features and photographic data using data fusion techniques to improve the model’s accuracy⁶³.

Conclusions

This study demonstrates that multispectral remote sensing MSR5 can be used for micro-classifying wheat crop yield at high spatial and temporal resolution. The Statistical and agriculture related departments can utilize this study for crop mapping and trending crop varieties to get and promote high-quality varieties and increase the country’s production and food security. It is also helpful to find the effect of climate on various types of crop varieties using remote sensing with low cost and the minimum period before the time to manage the need for food. In machine-learning models, RF performs best with approximately 96% accuracy, followed by the ETC with a 95% accuracy score. The best performance is obtained using the ANN which achieves approximately 97% accuracy score. It’s recommended to digital policymakers from the agriculture department can use ANN and RF to identify the different genotypes at farmer’s fields and research centers. The findings showed that multispectral data can map genotype to phenotype and classification of wheat varieties. This methodology may also be used for other crop mapping and genotype identification for accurate area estimation and yield forecasting at regional scale to ensure a better policy for food import and export at national level to ensure food security.

Data availability

All data generated or analysed during this study are included in this published article and data presented in this study are available on request from the first author (Mutiuallahjamil; email: mutiuallahj@gmail.com) and corresponding author (Ahsan Raza; email: araza@uni-bonn.de).

Received: 27 February 2023; Accepted: 7 November 2023

Published online: 14 November 2023

References

1. Tariq, A., Hashemi Beni, L., Ali, S., Adnan, S. & Hatamleh, W. A. An effective geospatial-based flash flood susceptibility assessment with hydrogeomorphic responses on groundwater recharge. *Groundw. Sustain. Dev.* **5**, 100998 (2023).
2. Li, P. *et al.* Soil erosion assessment by RUSLE model using remote sensing and GIS in an arid zone. *Int. J. Dig. Earth* **16**, 3105–3124 (2023).
3. Hussain, S. *et al.* Relation of land surface temperature with different vegetation indices using multi-temporal remote sensing data in Sahiwal region, Pakistan. *Geosci. Lett.* **10**, 33 (2023).
4. Tariq, A. *et al.* Terrestrial and groundwater storage characteristics and their quantification in the Chitral (Pakistan) and Kabul (Afghanistan) river basins using GRACE/GRACE-FO satellite data. *Groundw. Sustain. Dev.* **23**, 100990 (2023).
5. Tariq, A. Quantitative comparison of geostatistical analysis of interpolation techniques and semivariogram spatial dependency parameters for soil atrazine contamination attribute. In *Geoinformatics for Geosciences* (eds. Stathopoulos, N. *et al.*) 261–279 (Elsevier, 2023). <https://doi.org/10.1016/B978-0-323-98983-1.00016-8>.
6. Mary, R. *et al.* Exploring hazard quotient, cancer risk, and health risks of toxic metals of the Mehmood Booti and Lakhodair landfill groundwaters, Pakistan. *Environ. Nanotechnol. Monit. Manag.* **20**, 100838 (2023).
7. Tariq, A. *et al.* Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for burnt and unburnt scars. *Geomat. Nat. Hazards Risk* **14**, 28 (2023).
8. Felegari, S., Sharifi, A. & Moravej, K. Investigation of the relationship between NDVI index, soil moisture, and precipitation data using satellite images. *Sustainability* **2022**, 12 (2022).
9. Tariq, A. *et al.* Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive bayes tree models with optical remote sensing data. *Heliyon* **9**, e13212 (2023).
10. Chen, L. *et al.* Corynoxine protects dopaminergic neurons through inducing autophagy and diminishing neuroinflammation in rotenone-induced animal models of parkinson's disease. *Front. Pharmacol.* **12**, 1–11 (2021).
11. Tian, H. *et al.* Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm. *Remote Sens.* **11**, 820 (2019).
12. Chen, J., Du, L. & Guo, Y. Label constrained convolutional factor analysis for classification with limited training samples. *Inf. Sci. (N. Y.)* **544**, 372–394 (2021).
13. Li, Y., Du, L. & Wei, D. Multiscale CNN based on component analysis for SAR ATR. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–12 (2022).
14. Tariq, A., Yan, J., Gagnon, A. S., Riaz-Khan, M. & Mumtaz, F. Mapping of cropland, cropping patterns and crop types by combining optical remote sensing images with decision tree classifier and random forest. *Geo-spatial Inf. Sci.* **2022**, 1–19 (2022).
15. da-Silva-Monteiro, L. *et al.* Rainfall in the urban area and its impact on climatology and population growth. *Atmos. Basel.* **13**, 1610 (2022).
16. Mousa, B. G., Shu, H., Freeshah, M. & Tariq, A. A novel scheme for merging active and passive satellite soil moisture retrievals based on maximizing the signal to noise ratio. *Remote Sens.* **12**, 1–23 (2020).
17. Liu, J. *et al.* Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods. *Egypt. J. Remote Sens. Sp. Sci.* **26**, 415–426 (2023).
18. Urooj, R. & Ahmad, S. S. Spatio-temporal ecological changes around wetland using multispectral satellite imagery in AJK, Pakistan. *SN Appl. Sci.* **1**, 1–8 (2019).
19. Makarau, A., Richter, R., Schlapfer, D. & Reinartz, P. Combined haze and cirrus removal for multispectral imagery. *IEEE Geosci. Remote Sens. Lett.* **13**, 379–383 (2016).
20. Green, E. P., Mumby, P. J., Edwards, A. J., Clark, C. D. & Ellis, A. C. The assessment of mangrove areas using high resolution multispectral airborne imagery. *J. Coast. Res.* **14**, 483–443 (1998).
21. Guo, Y. *et al.* Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Indic.* **120**, 106935 (2021).
22. Pham, B. T., Tien-Bui, D., Prakash, I. & Dholakia, M. B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **149**, 52–63 (2017).
23. Kobler, A. & Adamic, M. Identifying brown bear habitat by a combined GIS and machine learning method. *Ecol. Modell.* **135**, 291–300 (2000).
24. Kulkarni, N. M. Crop identification using unsupervised ISODATA and K-means from multispectral remote sensing imagery. *Int. J. Eng. Res. Appl.* **07**, 45–49 (2017).
25. Hussain, S. *et al.* Spatiotemporal variation in land use land cover in the response to local climate change using multispectral remote sensing data. *Land* **11**, 595 (2022).
26. Chen, Y. *et al.* Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* **69**, 133–147 (2018).
27. Moulin, S., Bondeau, A. & Delecalle, R. Combining agricultural crop models and satellite observations: From field to regional scales. *Int. J. Remote Sens.* **19**, 1021–1036 (1998).
28. Zheng, B., Myint, S. W., Thenkabail, P. S. & Aggarwal, R. M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf.* **34**, 103–112 (2015).
29. Mingwei, Z. *et al.* Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* **10**, 476–485 (2008).
30. Islam, F. *et al.* Comparative analysis of GIS and RS based models for delineation of groundwater potential zone mapping. *Geomat. Nat. Hazards Risk* **14**, 27 (2023).
31. Erinjery, J. J., Singh, M. & Kent, R. Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. *Remote Sens. Environ.* **216**, 345–354 (2018).
32. Asif, M. *et al.* Modelling of land use and land cover changes and prediction using CA-Markov and Random Forest. *Geocarto Int.* **38**, 1–20 (2023).
33. Tariq, A. & Mumtaz, F. Modeling spatio-temporal assessment of land use land cover of Lahore and its impact on land surface temperature using multi-spectral remote sensing data. *Environ. Sci. Pollut. Res.* **30**, 23908–23924 (2022).
34. Liu, Z. *et al.* Application of machine-learning methods in forest ecology: Recent progress and future challenges. *Environ. Rev.* **26**, 339–350 (2018).

35. Guo, Y. *et al.* Machine learning-based approaches for predicting SPAD values of maize using multi-spectral images. *Remote Sens.* **14**, 748 (2022).
36. Liakos, K. G., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. Machine learning in agriculture: A review. *Sens. (Switzerl.)* **18**, 1–29 (2018).
37. Kavzoglu, T. & Colkesen, I. A kernel functions analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **11**, 352–359 (2009).
38. Ghaderizadeh, S., Abbasi-Moghadam, D., Sharifi, A., Tariq, A. & Qin, S. Multiscale dual-branch residual spectral-spatial network with attention for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 5455–5467 (2022).
39. Tariq, A., Mumtaz, F., Zeng, X., Baloch, M. Y. J. & Moazzam, M. F. U. Spatio-temporal variation of seasonal heat islands mapping of Pakistan during 2000–2019, using day-time and night-time land surface temperatures MODIS and meteorological stations data. *Remote Sens. Appl. Soc. Environ.* **27**, 100779 (2022).
40. Song, W., Li, S., Kang, X. & Huang, K. Hyperspectral image classification based on KNN sparse representation. In *International Geoscience and Remote Sensing Symposium (IGARSS) vols 2016-Novem 2411–2414* (2016).
41. Zhao, F. *et al.* Assessment of the sustainable development of rural minority settlements based on multidimensional data and geographical detector method: A case study in Dehong, China. *Socioecon. Plann. Sci.* **78**, 101066 (2021).
42. Khan, N. *et al.* Prediction of droughts over Pakistan using machine learning algorithms. *Adv. Water Resour.* **139**, 15 (2020).
43. Becker-Reshef, I., Vermote, E., Lindeman, M. & Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **114**, 1312–1323 (2010).
44. Liang, M. *et al.* Deep multiscale spectral-spatial feature fusion for hyperspectral images classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11**, 2911–2924 (2018).
45. Akturk, E. & Altunel, A. O. Accuracy assessment of a low-cost UAV derived digital elevation model (DEM) in a highly broken and vegetated terrain. *Meas. J. Int. Meas. Confed.* **136**, 382–386 (2019).
46. Ashfaq, M., Akram, M., Baig, I. & Saghir, A. Impact of ground water on wheat production in District Jhang, Punjab, Pakistan. *Sarhad J. Agric.* **25**, 121–125 (2009).
47. Masrur Ahmed, A. A. *et al.* Kernel ridge regression hybrid method for wheat yield prediction with satellite-derived predictors. *Remote Sens.* **14**, 1136 (2022).
48. Pang, A., Chang, M. W. L. & Chen, Y. Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia. *Sensors* **22**, 1–19 (2022).
49. Guo, F. *et al.* Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest. *Can. J. For. Res.* **46**, 582–594 (2016).
50. Chalikakis, K., Plagnes, V., Guerin, R., Valois, R. & Bosch, F. P. Contribution of geophysical methods to karst-system exploration: An overview. *Hydrogeol. J.* **19**, 1169–1180 (2011).
51. Hu, P. *et al.* Evaluation of vegetation indices and phenological metrics using time-series modis data for monitoring vegetation change in Punjab, Pakistan. *Water (Switzerl.)* **13**, 1–15 (2021).
52. Sharifi, A., Mahdipour, H., Moradi, E. & Tariq, A. Agricultural field extraction with deep learning algorithm and satellite imagery. *J. Indian Soc. Remote Sens.* **50**, 417–423 (2022).
53. Abbas, I., Liu, J., Amin, M., Tariq, A. & Tunio, M. H. Strawberry fungal leaf scorch disease identification in real-time strawberry field using deep learning architectures. *Plants* **10**, 2643 (2021).
54. Tariq, A., Mumtaz, F., Majeed, M. & Zeng, X. Spatio-temporal assessment of land use land cover based on trajectories and cellular automata Markov modelling and its impact on land surface temperature of Lahore district Pakistan. *Environ. Monit. Assess.* **195**, 114 (2023).
55. Riaz, U. *et al.* Evaluation of ground water quality for irrigation purposes and effect on crop yields: A GIS based study of Bahawalpur. *Pak. J. Agric. Res.* **31**, 1 (2018).
56. Wahla, S. S. *et al.* Assessing spatio-temporal mapping and monitoring of climatic variability using SPEI and RF machine learning models. *Geocarto Int.* **2022**, 1–20 (2022).
57. Hassani, B., Sahebi, M. R. & Asiyabi, R. M. Oil spill four-class classification using UAVSAR polarimetric data. *Ocean Sci. J.* **55**, 433–443 (2020).
58. Tariq, A. & Shu, H. CA-Markov chain analysis of seasonal land surface temperature and land use landcover change using optical multi-temporal satellite data of Faisalabad, Pakistan. *Remote Sens.* **12**, 1–23 (2020).
59. Tariq, A. *et al.* Land surface temperature relation with normalized satellite indices for the estimation of spatio-temporal trends in temperature among various land use land cover classes of an arid Potohar region using Landsat data. *Environ. Earth Sci.* **79**, 40 (2020).
60. Tariq, A. *et al.* Flash flood susceptibility assessment and zonation by integrating analytic hierarchy process and frequency ratio model with diverse spatial data. *Water* **14**, 3069 (2022).
61. Lu, L., Tao, Y. & Di, L. Object-based plastic-mulched landcover extraction using integrated Sentinel-1 and Sentinel-2 data. *Remote Sens.* **10**, 1–18 (2018).
62. Siddiqui, S., Safi, M. W. A., Tariq, A., Rehman, N. U. & Haider, S. W. GIS based universal soil erosion estimation in district Chakwal Punjab, Pakistan. *Int. J. Econ. Environ. Geol.* **11**, 30–36 (2020).
63. Freeshah, M. *et al.* Analysis of atmospheric and ionospheric variations due to impacts of super typhoon Mangkhut (1822) in the Northwest Pacific Ocean. *Remote Sens.* **13**, 661 (2021).
64. Tariq, A. *et al.* Forest fire monitoring using spatial-statistical and Geo-spatial analysis of factors determining forest fire in Margalla Hills, Islamabad, Pakistan. *Geomat. Nat. Hazards Risk* **12**, 1212–1233 (2021).
65. Zainab, N., Tariq, A. & Siddiqui, S. Development of web-based GIS alert system for informing environmental risk of dengue infections in major cities of Pakistan. *Geosfera Indones.* **6**, 77 (2021).
66. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features of image classification. *IEEE Trans. Syst. Man Cybern.* **3**, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314> (1973).
67. Villa, P., Stroppiana, D., Fontanelli, G., Azar, R. & Brivio, P. A. In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features. *Remote Sens.* **7**, 12859–12886 (2015).
68. Huang, C., Davis, L. S. & Townshend, J. R. G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **23**, 725–749 (2002).
69. Jalayer, S., Sharifi, A., Abbasi-Moghadam, D., Tariq, A. & Qin, S. modeling and predicting land use land cover spatiotemporal changes: A case study in Chalus Watershed, Iran. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 5496–5513 (2022).
70. Tariq, A., Siddiqui, S., Sharifi, A. & Shah, S. H. I. A. Impact of spatio-temporal land surface temperature on cropping pattern and land use and land cover changes using satellite imagery, Hafizabad District, Punjab, Province of Pakistan. *Arab. J. Geosci.* **15**, 1045 (2022).
71. Chang, C. C. & Lin, C. J. LIBSVM: A Library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 206 (2011).
72. Jain, A. K., Duin, R. P. W. & Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 4–37 (2000).
73. Mas, J. F. & Flores, J. J. The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens.* **29**, 617–663 (2008).

74. Salman, H. H., Dhafer, M. H. & Al-Shamkhee, H. A. J. Effect of posttreatments on the performance of tungsten carbide (K20) tool while machining (turning) of Inconel 718. *Int. J. Energy Environ.* **6**, 587–596 (2015).
75. Breiman, L. *Random Forests* 5–32 (Springer, 2001).
76. Olofsson, P., Foody, G. M., Stehman, S. V. & Woodcock, C. E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **129**, 122–131 (2013).
77. Majeed, M. *et al.* Spatiotemporal distribution patterns of climbers along an Abiotic Gradient in Jhelum District, Punjab, Pakistan. *Forests* **13**, 1244 (2022).
78. Sadiq-Fareed, M. M. *et al.* Predicting divorce prospect using ensemble learning: Support vector machine, linear model, and neural network. *Comput. Intell. Neurosci.* **2022**, 1–15 (2022).
79. Felegari, S. *et al.* Integration of Sentinel 1 and Sentinel 2 satellite images for crop mapping. *Appl. Sci.* **11**, 10104 (2021).
80. Luo, Z. & Ding, S. Object detection in remote sensing images based on GaN. *ACM Int. Conf. Proc. Ser.* **57**, 499–503 (2019).
81. Kontoes, C. C., Poilvé, H., Florsch, G., Keramitsoglou, I. & Paralikidis, S. A comparative analysis of a fixed thresholding vs a classification tree approach for operational burn scar detection and mapping. *Int. J. Appl. Earth Obs. Geoinf.* **11**, 299–316 (2009).
82. Pontius, R. G. & Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **32**, 4407–4429 (2011).
83. Shafizadeh-Moghadam, H., Tayyebi, A. & Helbich, M. Transition index maps for urban growth simulation: Application of artificial neural networks, weight of evidence and fuzzy multi-criteria evaluation. *Environ. Monit. Assess.* **189**, 8596 (2017).
84. Hamza, S. *et al.* The Relationship between Neighborhood Characteristics and Homicide in Karachi, Pakistan. *Sustainability* **13**, 5520 (2021).
85. Tariq, A. *et al.* Characterization of the 2014 Indus river flood using hydraulic simulations and satellite images. *Remote Sens.* **13**, 2053 (2021).
86. Tariq, A. *et al.* Spatio-temporal analysis of forest fire events in the Margalla Hills, Islamabad, Pakistan using socio-economic and environmental variable data with machine learning methods. *J. For. Res.* **13**, 12 (2021).
87. Mohammadi, M., Sharifi, A., Hosseingholizadeh, M. & Tariq, A. Detection of oil pollution using SAR and optical remote sensing imagery: A case study of the Persian Gulf. *J. Indian Soc. Remote Sens.* **6**, 9 (2021).
88. Tariq, A. *et al.* Quantitative analysis of forest fires in southeastern Australia using SAR data. *Remote Sens.* **13**, 2386 (2021).
89. Gašparović, M. & Jogun, T. The effect of fusing Sentinel-2 bands on land-cover classification. *Int. J. Remote Sens.* **39**, 822–841 (2018).
90. Gašparović, M., Zrinjski, M. & Gudelj, M. Automatic cost-effective method for land cover classification (ALCC). *Comput. Environ. Urban Syst.* **76**, 1–10 (2019).
91. Ghaderizadeh, S., Abbasi-Moghadam, D., Sharifi, A., Zhao, N. & Tariq, A. Hyperspectral image classification using a hybrid 3D–2D convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 7570–7588 (2021).
92. Kalantar, B., Pradhan, B., Amir-Naghbi, S., Motevali, A. & Mansor, S. Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomat. Nat. Hazards Risk* **9**, 49–69 (2018).
93. Khalil, U. *et al.* Developing a spatiotemporal model to forecast land surface temperature: A way forward for better town planning. *Sustainability* **14**, 11873 (2022).
94. Hussain, S. *et al.* Monitoring the dynamic changes in vegetation cover using spatio-temporal remote sensing data from 1984 to 2020. *Atmos. Basel.* **13**, 1609 (2022).
95. Yin, J. *et al.* Monitoring urban expansion and land use/land cover changes of Shanghai metropolitan area during the transitional economy (1979–2009) in China. *Environ. Monit. Assess.* **177**, 609–621 (2011).
96. Zhou, T., Pan, J., Zhang, P., Wei, S. & Han, T. Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sens. Switzerl.* **17**, 1210 (2017).
97. Bera, D. *et al.* Integrated influencing mechanism of potential drivers on seasonal variability of LST in Kolkata Municipal Corporation, India. *Land* **11**, 1461 (2022).
98. Fu, C. *et al.* Timely plastic-mulched cropland extraction method from complex mixed surfaces in arid regions. *Remote Sens.* **14**, 4051 (2022).
99. Haq, S. M. *et al.* Influence of edaphic properties in determining forest community patterns of the Zabarwan mountain range in the Kashmir Himalayas. *Forests* **13**, 1214 (2022).
100. Ahmadi, S. H. & Sedghamiz, A. Geostatistical analysis of spatial and temporal variations of groundwater level. *Environ. Monit. Assess.* **129**, 277–294 (2007).
101. Iqbal, M. F., Khan, M. R. & Malik, A. H. Land use change detection in the limestone exploitation area of Margalla Hills National Park (MHNP), Islamabad, Pakistan using geo-spatial techniques. *J. Himal. Earth Sci.* **46**, 89–98 (2013).
102. Rahman, M. H. *et al.* Multi-model projections of future climate and climate change impacts uncertainty assessment for cotton production in Pakistan. *Agric. Forest Meteorol.* **253**, 94–113 (2022).
103. Habib-ur-Rahman, M. *et al.* Impact of in-field soil heterogeneity on biomass and yield of winter triticale in an intensively cropped hummocky landscape under temperate climate conditions. *Precis. Agric.* **23**, 912–938 (2022).
104. Islam, F. *et al.* Landslide susceptibility mapping (LSM) of Swat District, Hindu Kush Himalayan region of Pakistan, using GIS-based bivariate modeling. *Front. Environ. Sci.* **10**, 1–18 (2022).

Acknowledgements

The authors extend their appreciation to the Researchers supporting project number (RSP2023R306), King Saud University, Riyadh, Saudi Arabia.

Author contributions

M.J.: conceptualization, methodology, software, formal analysis, visualization, data curation, writing—original draft, investigation, validation, writing—review and editing; H.R.: writing—review and editing; M.S.Z.: writing—review and editing; A.T.: methodology, software, formal analysis, visualization, data curation, writing—original draft, investigation, validation, writing—review and editing, supervision; R.I.: writing—review and editing; M.U.H.: methodology, visualization, data curation, writing—review and editing; A.M.: writing—review and editing; A.M.: writing—review and editing; A.E.S.: writing—review and editing; M.H.R., A.R., M.A.A., M.S.E.: methodology, software, formal analysis, visualization, data curation, writing—original draft, investigation, validation, writing—review and editing, and Funding.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.T., R.I. or A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023