PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science (2025) 93:173–196
https://doi.org/10.1007/s41064-024-00329-4

ORIGINAL ARTICLE

Selecting Relevant Features for Random Forest-Based Crop Type Classifications by Spatial Assessments of Backward Feature Reduction

Talha Mahmood[1] · **Muhammad Usman**[1] · **Christopher Conrad**[1]

Received: 13 September 2023 / Accepted: 27 November 2024 / Published online: 9 January 2025

Abstract
Random Forest (RF) is a widely used machine learning algorithm for crop type mapping. RF's variable importance aids in dimension reduction and identifying relevant multisource hyperspectral data. In this study, we examined spatial effects in a sequential backward feature elimination setting using RF variable importance in the example of a large-scale irrigation system in Punjab, Pakistan. We generated a reference classification with RF applied to 122 SAR and optical features from time series data of Sentinel-1 and Sentinel-2, respectively. We ranked features based on variable importance and iteratively repeated the classification by excluding the least important feature, assessing its agreement with the reference classification. McNemar's test identified the critical point where feature reduction significantly affected the RF model's predictions. Additionally, spatial assessment metrics were monitored at the pixel level, including spatial confidence (number of classifications agreeing with the reference map) and spatial instability (number of classes occurring during feature reduction). This process was repeated 10 times with ten distinct stratified random sampling splits, which showed similar variable rankings and critical points. In particular, VH SAR data was selected when cloud-free optical observations were unavailable. Omitting 80% of the features resulted in an insignificant loss of only 2% overall accuracy, while spatial confidence decreased by 5%. Moreover, the crop map at the critical point exhibited an increase in spatial instability from a single crop to 1.28. McNemar's test and the spatial assessment metrics are recommended for optimized feature reduction benchmarks and identifying areas requiring additional ground data to improve the results.

Keywords Backward feature reduction · Crop type mapping · SAR · Random Forest · Spatial accuracy metrics

1 Introduction

Remote sensing can contribute substantially to providing timely and accurate information on agricultural activities with high revisit frequency and spatial resolution (centimeter-scale) (Sishodia et al. 2020). For instance, crop type classifications allow estimation of crop area, crop diversity, and the spatial distribution of cropping patterns in an area (Ibrahim et al. 2021). These data, in turn, are needed for environmental modeling, e.g., monitoring crop growth (Lemoine and Léo 2015), assessment of crop water requirements (Conrad et al. 2013) and the forecasting of crop production to anticipate production shortfalls and food security (Nosratabadi et al. 2021). Despite notable improvements in recent decades, the classification process, particularly the composition of the feature space and the classification algorithm, still require further investigation. For instance, it is important to consider spatially clustered patterns, such as those found in the heterogeneous cropping zones of Punjab, Pakistan (Yang et al. 2020; Yin et al. 2020).

Spectral and temporal features play an important role in distinguishing crop types (e.g., Conrad et al. 2014; Hu et al. 2019). In this context, the increasing amount of accessible optical and microwave remote sensing data and new algorithms have provided unprecedented opportunities for accurate crop-type mapping (Orynbaikyzy et al. 2019). Despite numerous advantages, such as improved differentiation of spectrally similar crop types (e.g., Forkuor et al. 2014) and minor crop types (e.g., Orynbaikyzy et al. 2020), the constriction of extensive feature spaces, including e.g. topography metrics, and crop planting information (Foerster et al. 2012), has further improved the ability to discriminate between different crop types, leading to more accurate re-

✉ Talha Mahmood
talha.mahmood@student.uni-halle.de

[1] Department of Geoecology, Institute of Geosciences and Geography, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany

sults (Mazzia et al. 2020). However, this data-rich situation has greatly extended the computation time required to build a classification model, making the generation of large-scale crop-type maps in heterogeneous landscapes time-consuming (Löw et al. 2013).

Feature selection improves the comprehensibility and processing time of classification models and has been used in machine learning for decades (Vergara and Estévez 2014). Numerous approaches including filter, embedded, and wrapper methods have been tested (Saeys et al. 2007). Effective feature selection methods also eliminate features that may impede the classification process and simultaneously enhance classification accuracy (Hamzeh et al. 2016). Given the extensive research on feature selection over the years (Guyon and Andr'e Elisseeff 2003), but the limited focus on its spatial effects, this study further investigates the sequential backward feature reduction process, in particular its spatial effects, i.e. the spatial variations in mapping accuracy over extensive cropland.

The assessment of features suitable for accurate classification requires comparisons of accuracy measures. In feature selection approaches common accuracy metrics that analyze the confusion matrix such as overall, producer's, and user's accuracies (Olofsson et al. 2014), have been frequently applied. In addition, numerous authors (Löw et al. 2015a; Kumar et al. 2017; Sitokonstantinou et al. 2018; Bueno et al. 2020) have utilized McNemar's test (McNemar 1947) on the confusion matrix derived from the classified validation samples to assess and compare different classification algorithms and feature sets. However, to the best of our knowledge, it has never been applied for supporting optimized feature selection, e.g. to identify significant deviations from the reference crop type pattern received from all input features.

Additionally, investigations on the spatial agreement of thematic maps produced by different algorithms, feature sets, or sample sets are rare, despite their potential to provide valuable insights into mapping quality in terms of reliability and stability (Heupel et al. 2018). For example, Van Tricht et al. (2018) used spatial confidence derived from a random forest (RF) classifier to investigate the impact of adding SAR data to optical data for crop classification in Belgium. Heupel et al. (2018) derived spatial reliability and stability for pixel consistency in progressive crop type classification. No known research has analyzed indicators of spatial agreement between the predicted maps during feature reduction. We hypothesize that utilizing such indicators can further optimize the feature reduction process concerning the resulting crop type maps and spatial crop distributions. This step helps practitioners better understand the randomness of classification decisions introduced by feature reduction and identify relevant temporal windows in which satellite data are important for classification (e.g., Conrad et al. 2014; Yi et al. 2020). This approach may also help map producers increase classification accuracy in distinct cropping zones (Conrad et al. 2016).

This study aims at optimized feature selection and investigates ways to decrease the dimension of the feature space without compromising classification accuracy and spatial consistency. The latter refers to the spatial distribution of crop types in the resulting map in comparison to the map returned from a classification based on all features. Additionally, the important features and their temporal acquisition windows are analyzed. Minor goals included (1) a better description of reduced features and their location in the temporal course of the vegetation period and (2) the observation of the effects caused by feature reduction in different cropping zones. The investigations were carried out based on the RF algorithm (Breiman 2001), with the example of sequential backward elimination of unimportant features utilizing based on RF variable importance rankings. In addition to the standard accuracy metrics, McNemar's statistical test was applied to compare the results with the reference classification. Furthermore, a monitoring of spatial assessment metrics, including spatial confidence and spatial instability between class decisions in the map of all features and the reduced features map, was implemented. This study was applied in a large-scale irrigated agricultural region of Punjab, Pakistan, using satellite data from 2017, i.e., Sentinel-1 time series data combined with multitemporal Sentinel-2 data.

2 Materials and Methods

2.1 Study Area

2.1.1 Irrigation System, Cropping Pattern and Climate

The lower Chenab Canal (LCC) command area is one of the largest human-controlled irrigation areas within the Indus Basin Irrigation System (IBIS). It is located within Rechna Doab region which encompasses the land between the Ravi and Chenab Rivers. The LCC originates from the Khanki headworks on the Chenab River and irrigates approximately 12,400 km^2 of cultivated land on the eastern and western sides Fig. 1. The climate of the study area is arid to semi-arid, with large seasonal fluctuations in temperature and rainfall. The summer season is from April to September, with daily temperatures ranging from 21–49 °C. The winter season ranges from December to February, with temperatures ranging from 5–27 °C. The mean annual rainfall follows a gradient from 290–1050 mm from south to north in Rechna Doab (Usman et al. 2015). Most of the rainfall, approximately 70%, occurs during the monsoon season, i.e., in July August (Mujtaba et al. 2022).

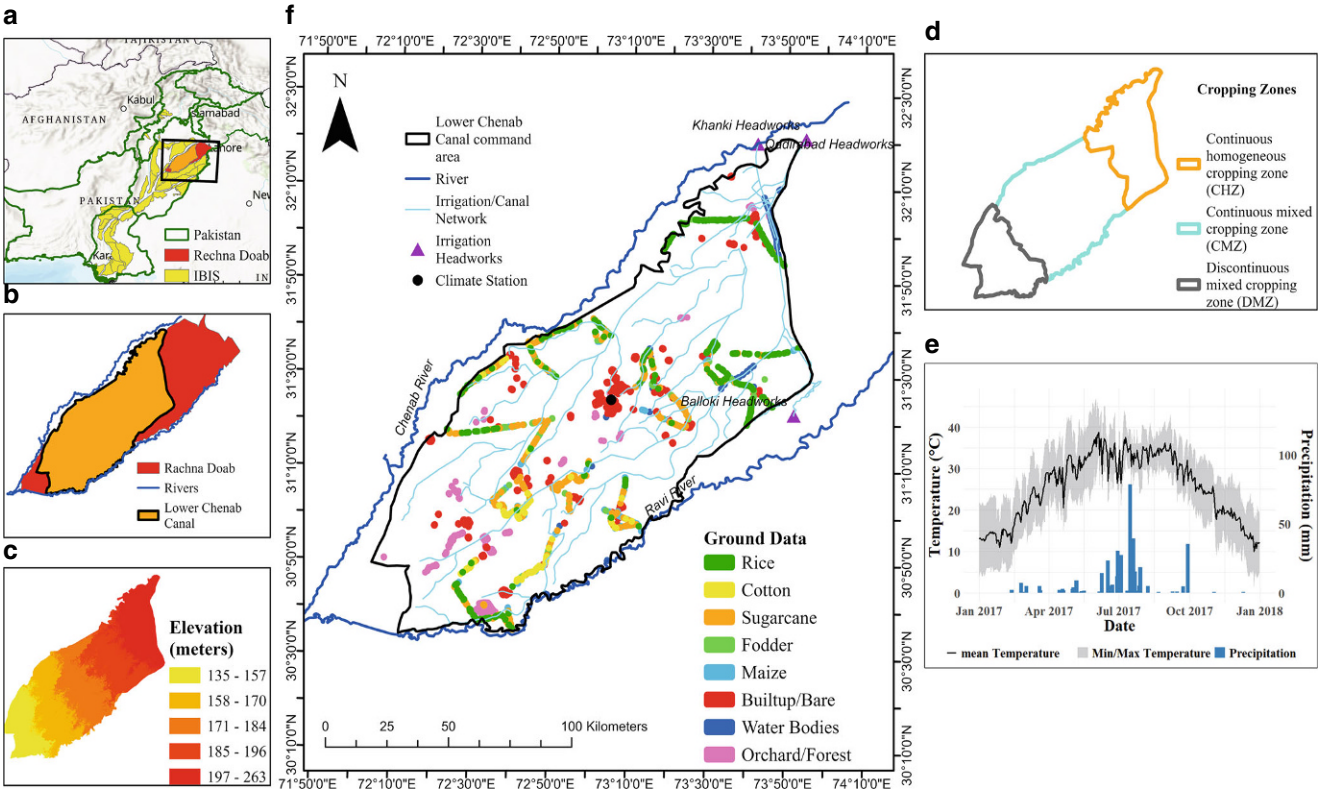

Fig. 1 The geographical location of the study area (Lower Chenab Canal command area) (**a**) in Pakistan and irrigated Indus Basin irrigation systems (IBIS); (**b**) in Rechna doab; (**c**) spatial distribution of elevation in the study area (source: Shuttle Radar Topography Mission); (**d**) three cropping zones in the study area—*orange:* continuous homogenous zone (CHZ), *violet:* continuous mixed cropping zone (CMZ), *gray:* discontinuous mixed cropping zone (DMZ); (**e**) daily temperature and precipitation at climate stations during 2017 (source: Pakistan Meteorological Department); (**f**) ground-truth data (*field polygons*) collected during a field visit in 2017 (Table 2)

The area has two major cropping seasons, one in the summer, known locally as the *Kharif* season, and another in the winter, referred to as the *Rabi* season. Rice and cotton are the main crop types during the *Kharif* season, whereas wheat is the major crop type during the *Rabi* season. Sugarcane is an annual crop that is cultivated mainly in September and February and is considered the third major crop in the region (Usman et al. 2015). Other prominent crop types include fodder, maize, and vegetables. For this study vegetables and fodder are treated as single crops due to their similar phenologies and occurrence in scattered small-scale farming systems.

The current research focuses on crop type classification during the *Kharif* season due to increased cropping heterogeneity, whereas wheat is the single dominant crop during the *Rabi* season. Moreover, varying cropping patterns and the influence of monsoons season present additional technical challenges for accurate crop mapping. To investigate the impact of cropping practices on crop mapping accuracy and feature reduction in the study area, three cropping zones were identified and are displayed in Fig. 1. These cropping zones are described in Sect. 2.1.2. The cropping calendar for the *Kharif* season being followed in the region is given in Table 1.

Table 1 Cropping calendar in the Lower Chenab Canal command area for the summer (*Kharif*) season

Summer Crops	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rice	–				–	Sowing	Transplanting	Growth			Harvest	–
Cotton	–				Sowing		Growth			Harvest		–
Sugarcane	Harvest	Sowing		Initial Growth			Maturity					
Maize	–					Sowning	Growth		Harvest	–	–	–
Fodder *Kharif*	–			Start		Multi Cut Growth				End	–	

2.1.2 Cropping Zones with Different Heterogeneity

In LCC, prior crop mapping efforts conducted by Awan and Ismaeel (2020) and Usman et al. (2015) relied on medium-resolution data from MODIS. Different crops are grown in various regions of the area due to varying environmental (e.g., soil type, climate, rainfall), water requirement, and availability (canal, groundwater and rainfall) factors. Specifically, rice and cotton, two major crops in the region, have distinct growth requirements and are cultivated in different zones (Usman et al. 2015).

To investigate the spatial effects of feature reduction under different spatial configurations of an agricultural landscape, the study area was divided into distinct cropping zones, based on the prevailing cropping practices. These zones are as follows: continuous homogenous zone (CHZ), continuous mixed cropping zone (CMZ), and discontinuous mixed cropping zone (DMZ) (Fig. 1). These zones are also in accordance with the distribution of cropping areas in Punjab, Pakistan, as derived from MODIS time series data by Usman et al. (2018). CHZ is dominated by rice crops, with fodder, maize, and sugarcane present in patches. The CMZ zone includes all the crops in the study area. Sugarcane is the dominant crop and exceeds the area share of other crops. In this zone, cotton, fodder, and maize are cultivated mostly on small scale. The spectral mixing of fodder and maize due to similar phenologies and the smaller scale of cultivation makes cropping in this zone more complex in comparison to the CHZ. Most of the middle part of the DMZ is barren but interspersed with forests and orchards. The southern (right and left) of this zone is cultivated with Rice, sugarcane, and cotton.

2.2 Data

2.2.1 Reference Data

Ground truth data on crop types were collected during a field campaign in October 2017. The geographical locations of the 894 crop fields were collected via a GPS device. The distribution of the samples followed the proportion of each crop-type present in the region. Objects representing the landscape surrounding the cropland, i.e., forest/orchard areas, and water bodies, including canal networks, settlements, and bare soil, were sampled from high-resolution Google Earth images. The high-resolution remote sensing data in Google Earth can be very useful for visual interpretation of such general land cover classes (Miyazaki et al. 2011). The details of the ground truth data are presented in Table 2.

A total of 894 crop-type polygons were collected. Of these, 626 (70%) were used for training, while the remaining 268 (30%) were used for validation purposes. A stratified random sampling design was used to split the reference data into two sets (i.e., training and validation). A stratified random split was used to maintain a 70:30 proportion of each crop type in the training and validation datasets. The number of validation samples in three cropping zones is provided in Table 3. Due to an insufficient number of samples, spatial and accuracy comparisons for cotton from the CHZ and fodder and maize from the DMZ were not included. In addition, these crops are cultivated in small patches in these zones and are difficult to find in the area.

Table 2 Number of crop-type polygons used for training and validation

	Total Fields/ Polygons	Training	Validation
Crop Type Classes			
Rice	274 (30.6%)	192	82
Cotton	84 (9.4%)	59	25
Sugarcane	287 (32.1%)	201	86
Maize	135 (15.1%)	95	41
Fodder	114 (12.8%)	80	34
Other Land Cover Classes			
Orchard/Forest	134	94	40
Urban/Bare	135	95	41
Water Bodies	115	81	35

The percentage next to each crop-type polygon indicates its share relative to the total crop-type data collected

Irrigation from different sources, including canals, groundwater, and rainfall, combined with traditional methods of seeding and fertilization, may not be used uniformly leading to spectral variability among fields of the same crop type (Aduvukha et al. 2021). Thus, instead of using a point directly, a polygon of the surrounding field was drawn to consider the within-field variability of the sample. The field boundaries were also extracted from Google Earth. However, despite other studies successfully utilizing aggregated field information as feature values for classification (Conrad et al. 2014; Ghazaryan et al. 2018), we observed that the average values of pixels inside individual field polygons for training the classifier can increase the confusion among crops with similar spectral behavior. Hence, we applied the classification at the pixel level, which is also per current practices (e.g., Orynbaikyzy et al. 2020). To reduce spatial autocorrelation and to maintain manageable computational effort, we extracted a small subset of random points. This subset is taken following the field size, as given in Table 3. The table also shows the resulting training and validation points in each cropping zone.

These points were generated after splitting the initial ground data (Table 2), ensuring that the random points inside the training crop polygon/fields should remain in training and vice versa. The resulting number of training and validation samples for the crop classes was 2250.

Table 3 Criteria for random points in the field/polygons and the resulting number of training and validation points in each cropping zone

No. Of pixels in polygon	No. Of random points inside polygon	Classes	CHZ		CMZ		DMZ	
1–2	1	–	*T.P*	*V.P*	*T.P*	*V.P*	*T.P*	*V.P*
3–5	2	Rice	330	149	121	46	93	38
6–8	3	Cotton	0	0	77	37	83	32
9–15	4	Sugarcane	41	15	355	160	105	41
16–20	5	Maize	33	16	117	69	1	4
>20	6	Fodder	44	22	161	46	12	2

T.P training points, *V.P* validation points

2.2.2 Satellite Data Preprocessing

The optical data from the Multi-Spectral Imager (MSI) onboard Sentinel-2A and B (S-2) and the C-band synthetic aperture radar (SAR) data from Sentinel-1A (S-1) were obtained via the Copernicus open access hub (https://scihub.copernicus.eu/). The acquisition dates are listed in Table 4. Cloud-free S-2 data was obtained on six distinct days during the observation period from May-October 2017 (*Kharif* season). The level-1C top-of-atmosphere reflectance data were converted to surface reflectance using Sen2Cor in the Sentinel Application Platform (SNAP) environment. Bands 1 (coastal), 9 (water vapor), and 10 (cirrus clouds) were excluded because of their irrelevance to crop mapping. The S-2 bands used in this study include 2, 3, and 4 (visible), 5, 6, 7 (red-edge), 8 and 8A (near-infrared; NIR), and 11 and 12 (SWIR), which lie in the visible to infrared window of the electromagnetic spectrum and are very important for studying land and vegetation processes (Gascon et al. 2017). In addition, the normalized difference vegetation index (NDVI) was calculated using band 8 and band 4 of S-2 using Eq. 1. The original spatial resolution of the S-2 bands used was 20 m, except for bands 2, 3, 4, and 8, which are at 10 meters. Previous studies (Immitzer et al. 2016; Orynbaikyzy et al. 2020) have highlighted the importance of the red edge and SWIR bands for crop separability; thus, all layers were resampled to 20 m for spatial consistency among different bands. Furthermore, the 20 m resolution is a significant improvement over previous crop mapping efforts in the study area (e.g., Cheema and Bastiaanssen 2010; Usman et al. 2015; Awan and Ismaeel 2020).

$$NDVI = \frac{Band8 - Band4}{Band8 + Band4} \quad (1)$$

C-band level 1 ground range detected (GRD) data in interferometric wide (IW) swath acquisition mode from S-1 were downloaded in 14-day time steps from May to October 2017 in the same orbit (107) and orientation (descending). The preprocessing of the data for the retrieval of backscatter values (σ^0) in VV and VH polarizations included thermal noise removal, radiometric calibration, speckle filtering, and terrain correction using digital elevation model from the Shuttle Radar Topography Mission (SRTM).

The resulting σ^0 values were converted to decibels. To reduce the effect caused by variations in the incidence angle on backscatter values, the incidence angle was normalized using double cosine correction provided by Ulaby et al. (1982) according to Eq. 2. For normalization, the mean incidence angle in the area was 40^0 (θ_{ref}).

$$\sigma^0_{ref} = \frac{\sigma^0_\theta \cos^2(\theta_{ref})}{\cos^2(\theta)} \quad (2)$$

where θ is the incidence angle and σ^0_{ref} is normalized the backscatter under θ_{ref}.

The gray-level co-occurrence matrix (GLCM), as developed by Haralick et al. (1973), is a widely applied method for the calculation of texture features in remote sensing data. Up to seven GLCM textural features can be derived from the GLCM matrix, including contrast dissimilarity, regular second moment, entropy, homogeneity variance and

Table 4 Acquisition dates of the optical and SAR data used in this study

	Sentinel-1A		Sentinel-2 (A&B)
1	20170506	1	20170506
2	20170518		
3	20170530	2	20170526
4	20170611	3	20170615
5	20170623		
6	20170705		
7	20170717		
8	20170729		
9	20170810	4	20170809
10	20170903		
11	20170915	5	20170918
12	20170927		
13	20171009	6	20171008
14	20171021		

Date Format: YYYYMMDD

GLCM mean (Shi et al. 2022). In this study, the GLCM mean texture feature was used as it was found to be the most suitable for classification of SAR image (Chen et al. 2018), improving the accuracy in separating crop types in comparison to other GLCM features (Treitz et al. 2014). GLCM mean texture measures the average gray-level value of the pixel pairs that constitute the GLCM matrix. We used SNAP to derive GLCM mean textures from σ^0 images in both polarizations (VV and VH) with a 9 × 9 moving window in all four directions (0^0, 45^0, 90^0, and 135^0). The GLCM mean texture was calculated using Eq. 3.

$$\mu = \sum_{i,j=0}^{N-1} i \left(P_{i,j}\right) \tag{3}$$

$P_{i,j}$ is a normalized gray tone spatial dependence matrix with and as the row and column indices of the GLCM, respectively. The mean texture is represented by μ, and the number of distinct gray levels in the GLCM is denoted by N.

2.3 Methodology

2.3.1 Feature Selection, Classification, and Variable Importance Algorithms

The sequential backward feature elimination method employed in this study belongs to the category of wrapper methods, that use classifier results as part of the feature evaluation process (Guyon and Elisseeff 2003). Backward feature elimination starts with all available features and iteratively eliminates the features. In contrast, e.g. the forward selection method starts from a single feature, progressively increasing the number of features. Notably, both techniques suffer from a nesting effect, i.e., a feature once added in forward selection cannot be removed again, and a feature removed in backward elimination cannot be added again (Vergara and Estévez 2014).

A hierarchal classification approach was employed (e.g., Forkuor et al. 2014). First, the cropland class was masked out against urban, water, and tree-covered areas using all features. Afterward, the cropland class was disaggregated into crop types, i.e., rice, cotton, sugarcane, fodder, and maize. The analytical focus was set on this second step, i.e., crop type mapping based on existing cropland boundaries, as applied by Heupel et al. (2018).

For classification, we used the RF classifier, which has not only demonstrated competitive results but also greater interpretability than other machine learning algorithms (Sheykhmousa et al. 2020). It has been increasingly used for classification purposes because of its accurate results, ability to handle high-dimensional and multicollinear data, and relative robustness against overfitting (Belgiu and Drăgu 2016). In recent years, there has been a notable increase in interest in the use of RF-based variable importance, due to successfully application of RF in the ranking of high-dimensional feature sets (Immitzer et al. 2012; Belgiu and Drăgu 2016). Hence, the use of RF variable importance with backward and forward feature elimination approaches for the identification of relevant features is a widely used practice to reduce the amount of input data as well as the computation time and effort (Speiser et al. 2019; de Moraes and Gradvohl 2021).

It is a nonparametric algorithm that builds and analyses ensembles of decision trees established using bootstrapped sampling. Each decision tree predicts a target class for each training sample, and the class with the highest number of votes is selected. We used fast implementation of RF in the statistical software R, provided in the package "ranger" (Wright and Ziegler 2017).

The algorithm has two hyperparameters, the number of trees (ntree) and the number of predictors (mtry), which are randomly selected for each decision tree in the RF. The default values for ntree and mtry are 500 and the square root of the input features, respectively. The majority of studies reviewed by Belgiu and Drăgu (2016) have reported that error stabilization before reaching 500 trees and increasing the mtry value, results in increased computational time (Gislason et al. 2006). Accordingly, ntree was maintained at 500, and mtry was adjusted based on the square root of the number of features during each feature elimination step.

Backward feature elimination requires a ranking of variables and builds on the idea that removing unimportant features from the classification does not affect the classification accuracy (Vergara and Estévez 2014; de Moraes and Gradvohl 2021). We used impurity-based feature importance, called Gini importance (also known as the mean decrease in impurity), from the RF for feature ranking. The Gini importance of a feature is defined as the total reduction in node impurity achieved by using that feature to split the data, averaged over all trees (Breiman et al. 1984). A high Gini importance score of a feature means that this specific feature is more important for classification and vice versa.

2.3.2 Error Matrix

The accuracy metrics applied in this study include the overall accuracy (OA) and the kappa coefficient (K) (Congalton and Green 2008). The OA is the ratio of the correctly classified pixels to the total number of pixels. The K value further indicates whether the classification was significantly better than the random classification (Cohen 1960). The K value ranges from 0–1, where a 0 corresponds to a total random classification, whereas 1 indicates perfect agreement between the reference data and classification. McNemar's statistical test was used to test the differences between the

two classifications for significance (McNemar 1947). McNemar's test compares each predicted sample between two classifications and provides the opportunity to compare the statistical similarity of confusion matrices. The McNemar test is a standardized normal chi-square (X^2) statistic computed from a 2 × 2 matrix of correctly and incorrectly classified samples in two classifications, calculated using Eq. 4.

$$X^2 = \frac{(f12 - f21)^2}{(f12 + f21)} \tag{4}$$

where f_{12} represents the pixels correctly classified by the 1st classification but incorrectly classified by the 2nd classification and f_{21} represents the pixels correctly classified by the 2nd classification and incorrectly classified by the 1st classification. A value of X^2 greater than 3.84 indicates that the two classifications are significantly different from one another at 95% confidence interval (Kumar et al. 2017).

For class-wise comparison of the accuracies, the error matrix was further analyzed using the Fscore (Inglada et al. 2017). This measure combines both the user (UA) and producer (PA) accuracies from the error matrix for each class according to the relationship given in Eq. 5. The Fscore range from 0–1, where 1 is considered the optimal result for a particular class.

$$Fscore = 2 * \frac{UA * PA}{UA + PA} \tag{5}$$

2.3.3 Spatial Assessment Metrics

Spatial confidence and spatial instability were calculated to investigate the spatial impacts of the feature selection process on the resulting crop type map. These two metrics are interconnected and can be used to identify hotspots of unclear classification decisions, which in turn require more attention for better accuracy, e.g., by increased sampling.

At each feature reduction step, spatial confidence was averaged over validation samples in the backward feature elimination process by comparing all sequential classifications starting from the reference classification (Sect. 3.1). In other words, after reducing n features, $n+1$ layers were integrated to calculate spatial confidence. It is defined as the ratio between the number of classifications that have the same class at corresponding pixels to the total number of maps computed at each step. High values of spatial confidence correspond to more reliable classification and vice versa.

Spatial instability was assessed by counting the number of unique crops/classes occurring at each pixel and averaging over all the validation samples during the feature reduction process. A spatial instability value of one indicates that a pixel does not change during the feature reduction process, indicating a single crop. A higher value indicates more variability and randomness in the classification decision. For example, an average value of 2 for spatial instability means that a pixel was classified into two distinct classes during the feature reduction process. Spatial instability was used to identify the randomness of classification decisions in space. The more frequently different crop types appear in a pixel during the feature reduction process, the more random the resulting classification.

Figure 2 presents an example demonstrating the calculation of spatial confidence and instability between a reference classification (with n features) and a reduced feature classification (with n-2 features), resulting in three consecutive classifications at this step. Shaded pixels indicate changes in classification output during the feature reduction process, resulting in reduced spatial confidence and increased spatial instability. Specifically, pixels shaded in green represent the most inconsistent classifications, exhibiting the lowest spatial confidence value (0) and highest spatial instability (3) as defined above and illustrated in the figure below. The average values (spatial confidence: 0.90; spatial instability:1.25) are derived by calculating the mean across all pixels.

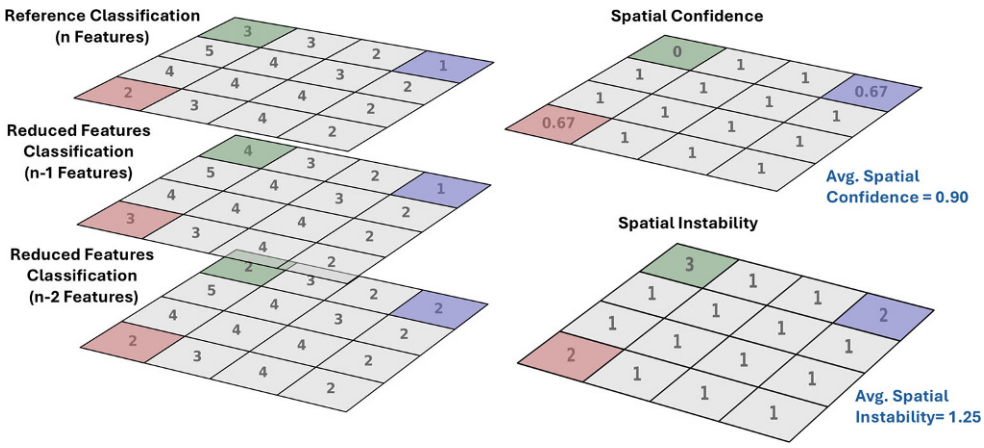

Fig. 2 Example calculation of spatial confidence and instability after reducing two features, comparing the reference classification with classifications based on the reduced feature set

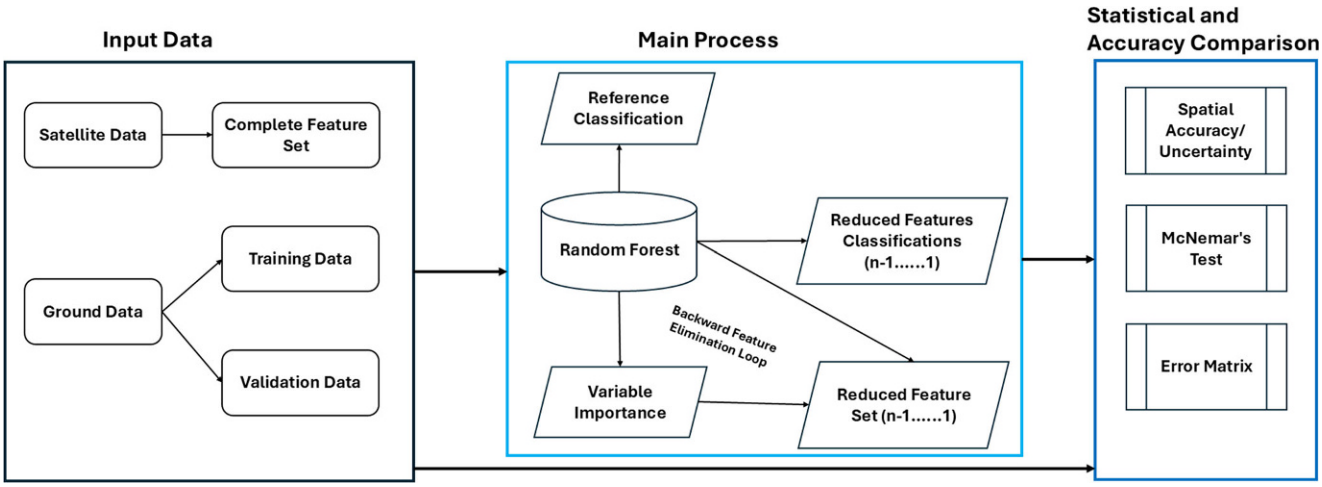

Fig. 3 Schematic view of the workflow for the assessment of backward feature elimination used in this study

2.4 Analysis Workflow

The analytical workflow of the study is illustrated in Fig. 3. The workflow includes two primary steps, i.e., the main process and a subsequent analysis phase involving statistical and accuracy comparisons between reference and reduced feature classifications. The main process starts with the classification of the entire feature set with RF to obtain the reference map. Based on the variable importance analysis, the least important feature is removed (reduced feature set), and the RF is applied again, which results in a reduced feature classification. Afterward, in the second step, the error matrices of the reference and reduced feature classifications were analyzed, and the McNemar test and the error matrix were applied (Sect. 2.3.2), and the results were entered into the accuracy metrics table that allows for tracing the accuracy development in the feature reduction procedure. The process was repeated until only one feature, i.e., the most important feature, remained. After applying the spatial assessment metrics, two uncertainty measures (Sect. 2.3.3) are derived from the reduced feature classifications. The reference data (Table 2) was split into two sets (i.e., training and validation) as explained in the Sect. 2.2.1.

To investigate the effect of the samples on the RF models and variable importance and hence to increase the reliability and robustness of the process as suggested by Stehman and Foody (2019), the entire procedure was repeated 10 times by randomly splitting the ground samples in the training and validation datasets. Ten different sampling splits were further repeated 10 times with a backward feature reduction process, which resulted in 100 elimination instances (rankings) of each feature across 10 sampling splits. These ranking eliminations across 10 different sampling splits were compared through the Kruskal–Wallis test (Kruskal and Wallis 1952). The Kruskal–Wallis test is a widely used statistical test for comparing the machine learning algorithms (da Silva et al. 2022). It is a nonparametric test used for data that is not normally distributed, where t-tests may not be appropriate. The Kruskal–Wallis test was performed to assess the difference in the ranking distribution of selected features during the backward feature reduction process across 10 distinct sampling splits. The null hypothesis (H_0) of the Kruskal–Wallis test is that all sampling splits result in equivalent rankings. The null hypothesis is rejected when the p-value associated with the Kruskal–Wallis test statistic is smaller than the chosen significance level (α), which is set at 0.05.

The Results section is structured as follows: Sects. 3.1 and 3.2 describe the outcomes of 10 repetitions from single-sampling splits with RF variable importance and a backward feature reduction process along with the calculation of McNemar's critical point. Section 3.3 presents the results of the feature ranking analysis from single splits during 10 repetitions, as well as the effect of 10 different sampling splits on those rankings. Section 3.4 provides a crop-type map based on the final selected features, and an overall and class-wise accuracy analysis across the entire study area. The accuracy analysis (spatial accuracy and error matrix) is detailed in Sects. 3.5 and 3.6 for each cropping zone and crop type.

2.4.1 Generation of Reference Map

We set the crop type classification based on the complete feature set as the reference map of our experiment for two reasons. First, using all features is usually close to the optimum (e.g., Löw et al. 2013; Yin et al. 2020; Conrad et al. 2014), and second, a systematic experiment requires a clear reference map (Stehman and Foody 2019).

2.4.2 Generation and Evaluation of Reduced Feature Maps

Starting from the reference map, backward feature elimination was applied. It is an iterative procedure in which each step reduces the feature set by dropping out the least important feature received from the variable importance assessment. From each resulting reduced feature map (Fig. 5), OA, and K, were calculated. Furthermore, McNemar's test, in addition to spatial confidence and instability, was calculated by comparing the reduced feature maps at each stage with the reference map and entered into a table. The entire process, with the application of RF variable importance, backward elimination of features, and comparison with the reference map, was repeated 10 times because the repetition of the accuracy assessment protocol increases the reliability of the process (Stehman and Foody 2019).

2.4.3 Detection of Optimal Feature Set and Analysis of Spatial Effects

One set of features should supply both reduced computational costs and high accuracy, comparable to the reference map, in terms of OA and spatial distribution of crop types (Hu et al. 2019). The latter is urgently needed to increase the acceptance of crop-type maps in practice. Therefore, a compromise for the accuracy metrics was suggested by identifying a critical point. This critical point is achieved when McNemar's statistics constantly indicate significant differences between the reference crop type map and the reduced feature crop type map. Constantly means that at least two consecutive McNemar's test values exceed 3.84 during feature reduction.

To trace the spatial effects of the feature reduction process, spatial confidence, and instability measures were applied in each feature reduction step, and uncertainty measures were produced and analyzed. The zones in which the LCC was categorized (Sect. 2.1.2) further helped us to identify configurations in the agricultural system that are affected by the feature reduction process and require attention. Thus, for further analysis, the accuracy measures were also computed for the three spatial zones during each instance of the feature reduction and compared with the results of accuracy metrics derived for the entire LCC.

3 Results

3.1 Reference Classification

The accuracy metrics of the reference classification using all features varied only slightly among the 10 applied repetitions (Fig. 4a). The results obtained from repetitions 6 and 10 were found to be similar to the average of all 10 repetitions. This indicates that these two repetitions were less prone to variations and would likely be easier to reproduce.

Except for four comparisons (3 and 5; 5 and 7; 7 and 8; 7 and 9), all McNemar's X^2 between the reference classifications in 10 repetitions show insignificant differences (Fig. 4b). McNemar's X^2 value between paired data indicates the difference between two classifications; a higher value indicates a greater difference, and vice versa. The reference classification, obtained from repetition 6 showed the closest statistically similarity to the other repetitions, with the highest X^2 value of 1.50, which was still lower than the highest values observed in the other repetitions.

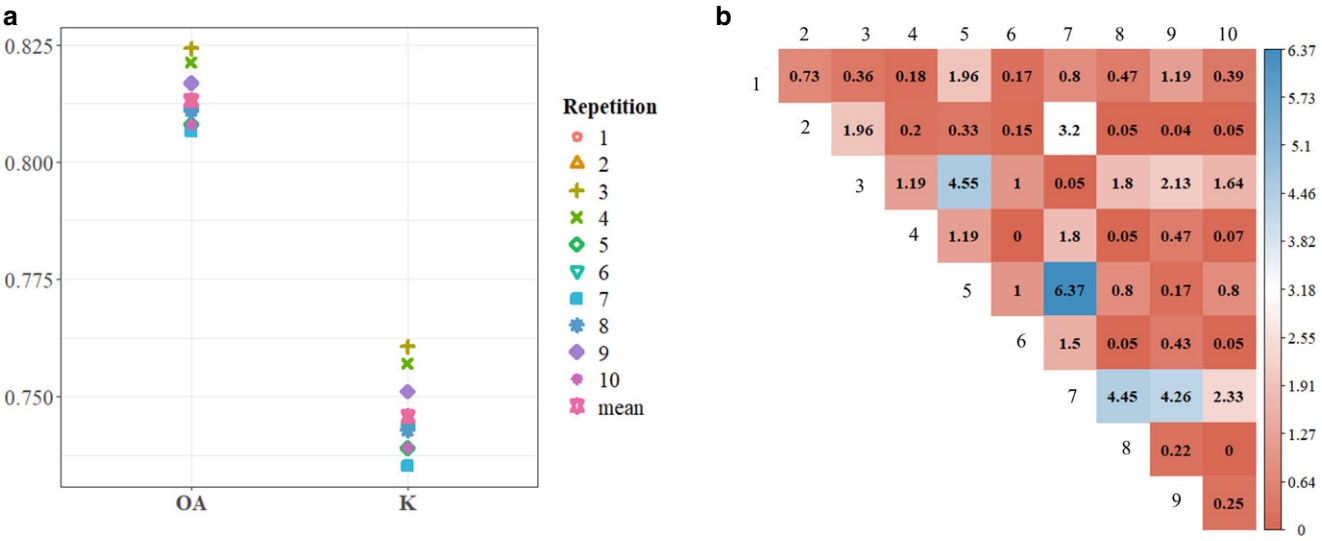

Fig. 4 Overall accuracy (OA), Kappa coefficient (K), and their means during 10 repetitions of classifying all features (reference classifications); (**b**) McNemar's test between reference classifications in 10 repeated runs; comparisons resulting in McNemar's test value ≥ 3.84 are significantly different (*blue shades*)

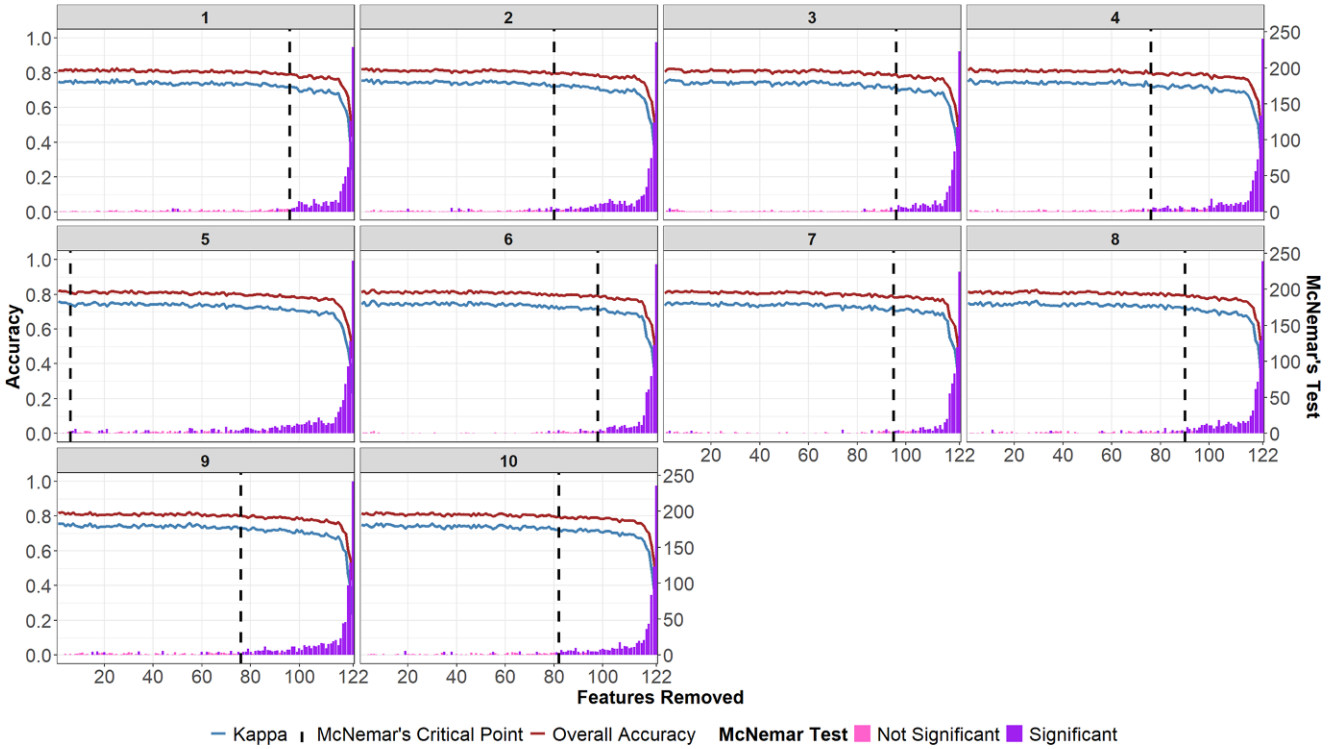

Fig. 5 Overall accuracy (*brown*), kappa (*light blue*), and McNemar's test critical points (*black vertical dotted line*) obtained during the 10 repetitions. McNemar's test value at each feature elimination step is also represented using a bar plot

Therefore, repetition 6 was selected as a reference map for further analysis despite minor randomness indicating some classification uncertainty in the crop maps.

3.2 Accuracy Metrics and Critical Points

The iterative process of backward feature selection was repeated 10 times. Figure 5 shows the OA and K after the least important features was eliminated the at each step for all 10 repetitions. The pink (not significant) and purple (significant) bars highlight the results of McNemar's test between the reference map and the reduced feature maps at each step, representing the significance of the eliminated feature at each step of feature reduction. The vertical black dotted line shows the critical point based on McNemar's test criteria (Sect. 2.4.3). Apart from repetition, five critical points occurred in a range of 76–98 reduced features (Table 5). Repetition 5 shows two consecutive McNemar's X^2 values greater than 3.84 at feature reduction step 7 and then behaves similarly to other repetitions. The OA ranged from 0.42–0.82 as the classification was performed using all 122 features down to a single feature. The maximum standard deviations of OA and K across all reduction steps observed were 2.1% and 2.9%, respectively, during the 10 repetitions.

Table 5 McNemar's critical points and number of important features across the 10 repetitions in backward feature elimination workflow

Repetition	Critical Point	Important Features
1	96	26
2	80	42
3	96	26
4	76	46
5	7	115
6	98	24
7	95	27
8	90	32
9	76	46
10	82	40

The reduction in features also harmed the spatial assessment metrics presented in Fig. 6. This effect was less pronounced for spatial confidence than for spatial instability. Using fewer than 10 features (x-axis 111-121) remarkably increases the probability of varying class decisions. However, a sharp increase in spatial instability occurred in all 10 repetitions beyond the critical point, as indicated by the McNemar test. The maximum standard deviations for spatial confidence and spatial instability received during the 10 repetitions were 3.4% and 3.9%, respectively.

3.3 Variable Ranking Analysis

This section analyses the order in which the features were removed during the 10 repetitions with a particular focus on the 6th repetition, as explained in Sect. 3.1. The order

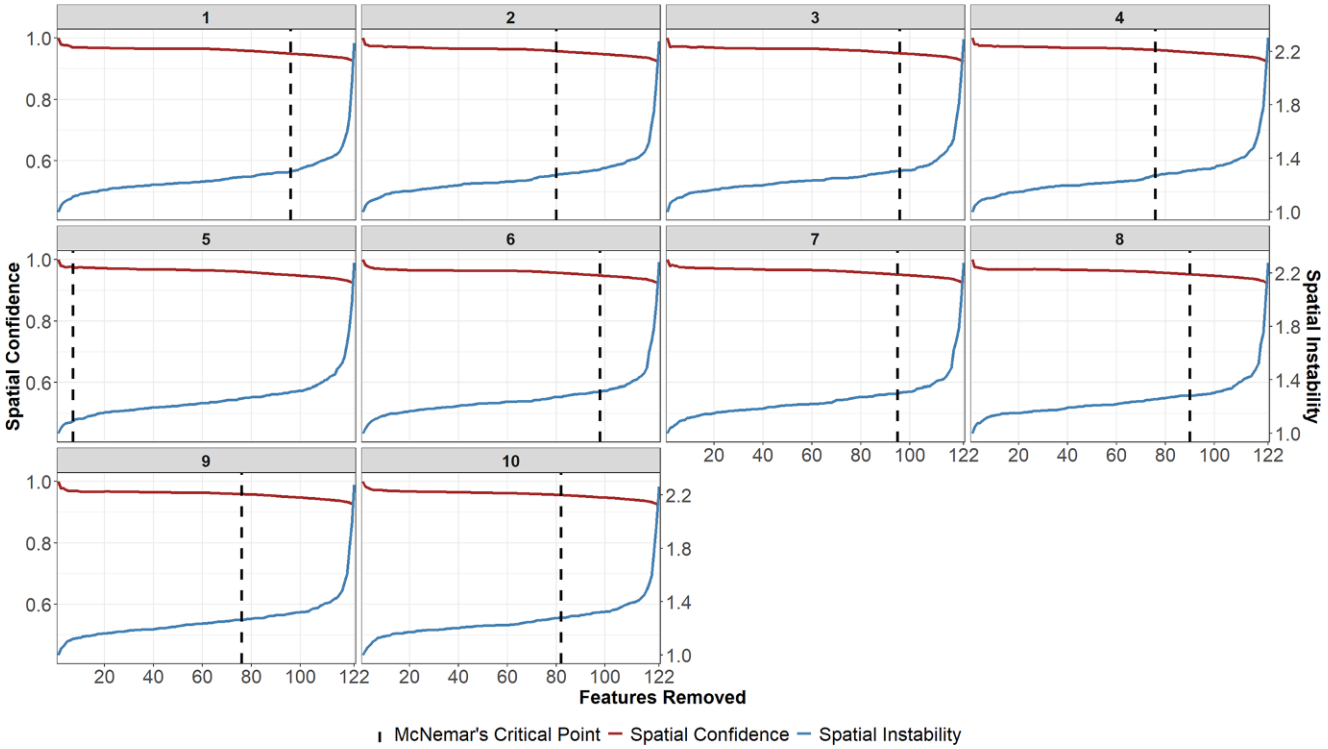

Fig. 6 Spatial confidence (*brown; y-axis*) and spatial instability (*light blue; secondary y-axis*), 10 times along with the critical points

of feature elimination during all 10 repetitions is explicitly given in Fig. 13. The Annex also labels the feature to the sensor (S-2-optical or S-1 SAR) and presents the modal value (i.e., the most frequently occurring rank position at which a feature is removed) and maximal distance (to this modal value) of each feature. Figure 7 shows the results of feature ranking from the 6th repetition run, with the x-axis positions reflecting the elimination order during this specific run. The points represent the modal values of the feature elimination positions, with red indicating the optical and light blue representing the SAR feature. These values were calculated from ten repetitions. The distance from the 1:1 line indicates deviations observed from the common ranking observed in repetition six. The size of the bars in Fig. 7 (distance) indicates the absolute difference between the modal value and the maximum value of the elimination instance of each feature from 10 repetitions. It is evident from the figure that features after the critical point were more consist in elimination during the process. The colors indicate the data source, i.e., optical or SAR.

Low (high) distance values show consistent (inconsistent) elimination positions of the features across the 10 repetitions. Among the 24 non-eliminated (after the critical point) features, 20 exhibit a maximum distance of three or less, which also strengthens the argument that important features are more likely to be eliminated at a similar instance. Most of the features (86 out of 122) show a consistent position at which they are eliminated during the classification process, with a maximum distance of nine or less. However, five features exhibit highly inconsistent behavior, with a maximum distance of 20 or more. This may be because these features have a greater degree of correlation with one another, which in turn leads to variations in their elimination position across different repetitions.

Figure 8 underpins this assumption of high correlation among these features, except for optical features. For instance, S-2 bands 8A (“20170506. B8A”) and 3 (“20170615. B3”), recorded on May 6 and June 15 respectively, are exceptions to this trend.

Figure 9a presents the variability in elimination instances of the features, which may depend on the specific training dataset used in the model run as well as the resulting variable importance. It displays the distribution of 100 elimination instances of each feature resulting from 10 repetitions of 10 different sampling splits with a backward feature reduction process, using individual box plots for all 122 features. The variation in the box plot is visible from the spread of the data, with higher plot indicating a greater degree of variation and vice versa. The features after McNemar’s critical point exhibit less variation than the features before the critical point. This indicate that these features are important regardless of the samples used, suggesting that the selected features were not selected by chance.

The effect of the training samples used for RF and variable importance on features that remained after McNemar’s critical point (selected or important features) during back-

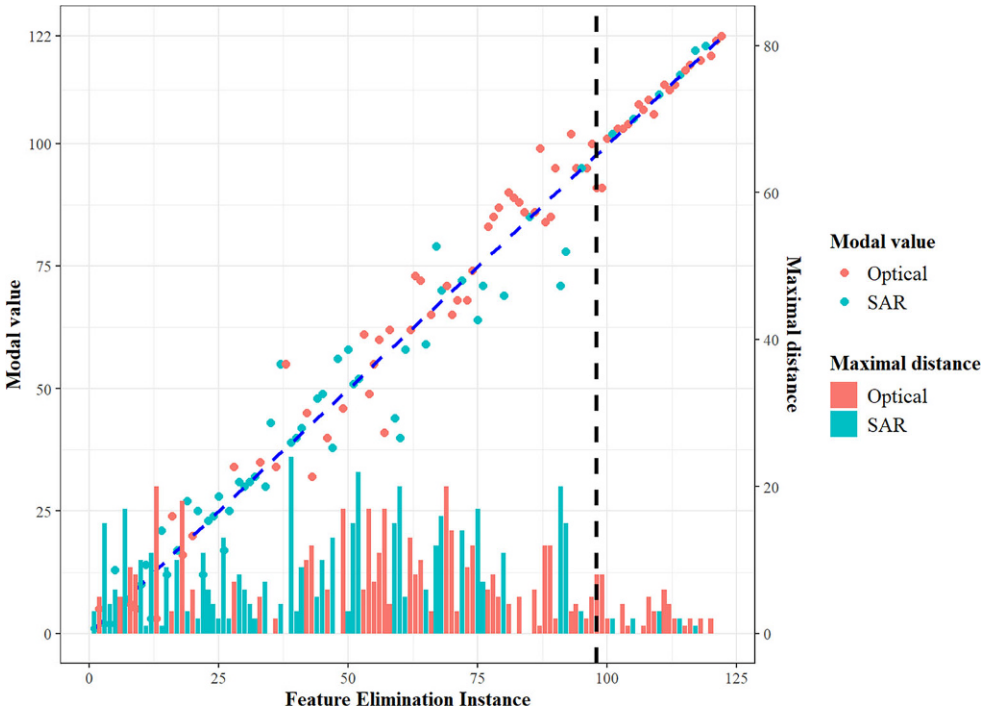

Fig. 7 Modal values (*y-axis*) of feature ranking during 10 repetitions ordered by the ranking positions achieved in repetition 6 (Fig. 13: *x-axis*); bars show the maximal distance to the modal value (*secondary y-axis*), the black dotted line shows McNemar's critical point during 6th repetition (Table 5), and the blue diagonal dotted line represents the 1:1 line

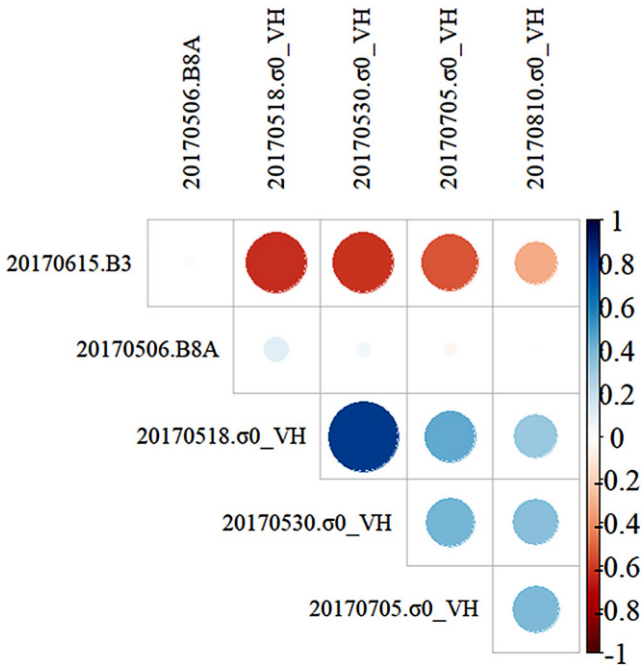

Fig. 8 Correlation coefficient between features with a maximal distance of 20 or more

ward feature reduction were further analyzed through the Kruskal–Wallis test. The modal ranking values of selected features from 10 repetitions across 10 different sampling splits were used for comparison among the sampling splits. Kruskal–Wallis test p values of pairwise comparisons of ten different sampling splits are presented in Fig. 9b. The color gradient (from brown to blue) represents associated p-p-values obtained through the Kruskal–Wallis test among ten different sampling splits, from lowest to highest. The Kruskal–Wallis test showed that there was no significant difference between sampling splits ($p > \alpha$), as presented in Fig. 9b. Therefore, we fail to reject H_0 among the comparisons between distinct sampling splits, indicating that the samples used to train the models do not have a significant impact on the outcome of the selected features.

3.4 Crop Type Map Based On the Optimal Feature Set

Table 6 list the variables according to the order of elimination, and the features highlighted in bold font are the final selected features of repetition run six (Fig. 5) based on variable importance. The features that range above 98 (critical points) are important features that still produce a classification statistically similar to the reference classification. The 24 features used for final crop mapping included 18 optical features and 6 SAR features. It is evident (Fig. 7) that on average, optical features (S-2) are removed later than SAR features (S-1).

The important features after the critical point that produced accuracy statistically similar to that of the complete feature set included the SWIR (B11, B12), red edge (B05), red (B04), and NDVI from S-2. Notably, these important features from S-2 span all months of the *Kharif* growing season from May–September, except for July, when S-2 data is unavailable due to cloud cover during the monsoon.

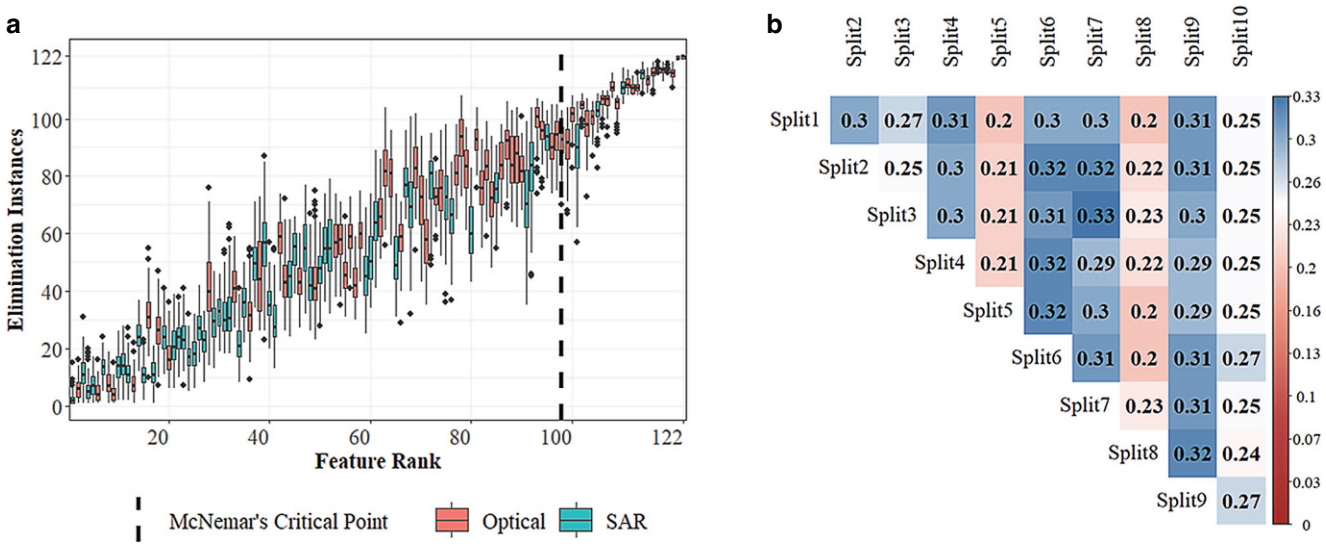

Fig. 9 Analysis of selected features across ten distinct sampling splits. (**a**) Variability in elimination instances of features across ten different sampling splits (feature rank corresponds to feature elimination instances during 6th repetition (Fig. 13)). (**b**) Kruskal–Wallis test *p*-value comparing mode values of important features across ten repetitions of ten different sampling splits

All six important S-1 features were derived from VH polarization. Five out of the six important SAR features were from the monsoon season (July), during which S-2 data is unavailable. Among these six features, four were mean GLCM texture features, derived from VH polarization backscatter. Additionally, almost half (10 out of 22) of the selected features are from July and the early phase of August, when most of the crops in the study area are at their full development stage.

A direct comparison between the reference and the selected reduced feature classification revealed a decrease in OA from 0.81 to 0.79, which represent a 2% loss in accuracy with an almost 80% reduction in the number of features according to McNemar's critical point criterion (Table 7). The K and Fscore decreased by approximately 4% during this process. The spatial confidence decreased to 0.95, and the spatial instability increased to an average of 1.28 crops per pixel from a single crop during the feature reduction process from the reference classification to the critical point classification. Table 7 also indicates that rice is least affected by the feature reduction process and achieved maximum Fscore of 0.93 and 0.92 for all features and important features, respectively. The Fscores of sugarcane and maize decreased by 2% during the feature reduction process. Cotton and fodder were more susceptible to feature reduction, with 7% and 5% decline of accuracy, respectively.

3.5 Spatial Uncertainty Assessment Per Cropping Zone

The resulting crop type map (Fig. 10) shows the three zones mentioned in Sect. 2.1.2, i.e., the northern part of the study area, which is dominated by rice (CHZ); the middle part, which is an intensively used mixed cropping zone (CMZ) with a patchy cultivation pattern (Fig. 10e); and the southern part (DMZ), which features sparse agricultural areas interspersed with widespread forest/orchard, urban, and barren patches. The OA levels in these zones resembled each other during feature reduction but at two different levels. Although the accuracy levels in the CHZ and the DMZ exceeded 0.80 in all reduction steps until the critical point, this value could not be achieved in the CMZ. The OA in the CMZ decreased to 0.76 until reaching the critical point, which was 0.84 and 0.82 for the CHZ and DMZ zones, respectively. This observation may be caused by an increased likelihood of mixed pixels in mixed cropping areas, which always leads to decreases in classification accuracy (Löw and Duveiller 2014).

Figure 10b–d also depicts OA and spatial assessment metrics (as defined in Sect. 2.3.3) during backward feature reduction in each of these three zones. The graphs show the mean values derived from the validation samples in each zone at each feature reduction step. Zone-wise OA analysis revealed distinct differences between the CHZ and the CMZ together with the DMZ. OA demonstrated a decline, reaching 0.80 and 0.76 in the DMZ and CMZ, respectively, compared to the value of 0.84 observed in the CHZ. The spatial confidence between the classification results based on the reduced feature sets declines sharply in the CMZ after a few features. In the same zone, the probability of a pixel being classified into a class differing from the reference map is always greater than that in the two more homogenous zones. The CHZ and DMZ had spatial instability values of 1.21 and 1.26 and spatial confidence values

Table 6 Order of elimination of features used in this study, e.g., 122 shows the most important feature or feature that remained in the last classification when all other features were removed sequentially (S-1 = Sentinel-1 SAR data, and S-2 = Sentinel-2 optical data)

Datasets	06. May	18. May	26. May	30. May	11. Jun	15. Jun	23. Jun	05. Jul	17. Jul	29. Jul	09. Aug	10. Aug	03. Sep	15. Sep	18. Sep	27. Sep	08. Oct	09. Oct	21. Oct
σ0VH	76	91	–	39	1	–	68	52	**114**	**110**	–	60	61	35	–	34	–	29	24
σ0VV	26	3		11	17		7	21	47	19		95	65	44		41		22	40
GLCM VH	72	**105**		75	4		92	**101**	**117**	**119**		50	67	51		32		30	25
GLCM VV	14	12		15	5		10	23	48	31		85	59	37		80		27	45
NDVI	**115**	–	**120**	–	–	**116**	–	–			96	–	–	–	**100**	–	37	–	–
B02	70		87			89					97				98		97		
B03	38		83			69					**99**				**109**		73		
B04	28		**106**			**107**					90				**111**		**100**		
B05	20		79			93					**112**				**104**		13		
B06	9		54			42					77				86		74		
B07	6		74			88					73				57		58		
B08	2		62			63					84				46		43		
B08A	13		71			78					**102**				55		24		
B11	8		49			64					**121**				**122**		17		
B12	36		**103**			94					**118**				**113**		53		

Elimination ranks of important/selected features are highlighted in bold font

Table 7 Overall accuracy, kappa coefficient, spatial confidence, spatial instability and class-wise Fscore comparison between reference and reduced feature classifications

	Fscore	
	Reference Classification	Important Features Classification
Rice	0.93	0.92
Cotton	0.82	0.75
Sugarcane	0.85	0.83
Maize	0.64	0.62
Fodder	0.51	0.46
Overall Accuracy	0.81	0.79
Kappa	0.75	0.71
Spatial Confidence	1	0.95
Spatial Instability	1	1.28

of 0.96 and 0.95, respectively, indicating that on average, 96% and 95% of the 98 removed features did not affect the final classification of the CHZ and DMZ, respectively. The spatial instability increased to 1.37, while the spatial confidence dropped to 0.94 up to the critical point in the CMZ. The visual illustration of spatial comparison between the reference and important feature maps in the three cropping zones is provided in Fig. 11.

Figure 11 presents a close-in view between the reference map and the important feature map across the three cropping zones. Pixels other than the crop classes were masked for better visual interpretation and understanding. The overall results indicate notable similarities between the reference and important feature classifications across all three zones. Nevertheless, a slight increase in class mixing is observed in the CMZ when compared to the CHZ and DMZ. These observations support the results provided above that feature reduction, which indicate that feature reduction has a more pronounced impact on the CMZ.

3.6 Class-Wise Assessments Per Cropping Zone

Figure 12 shows the Fscore, spatial confidence and spatial instability of the final selected classification per crop type within each zone. Due to the absence of and very few ground truth points for cotton in the CHZ and fodder and maize in the CMZ (Table 3), the results of these crop types were excluded from these zones. In terms of Fscore, rice

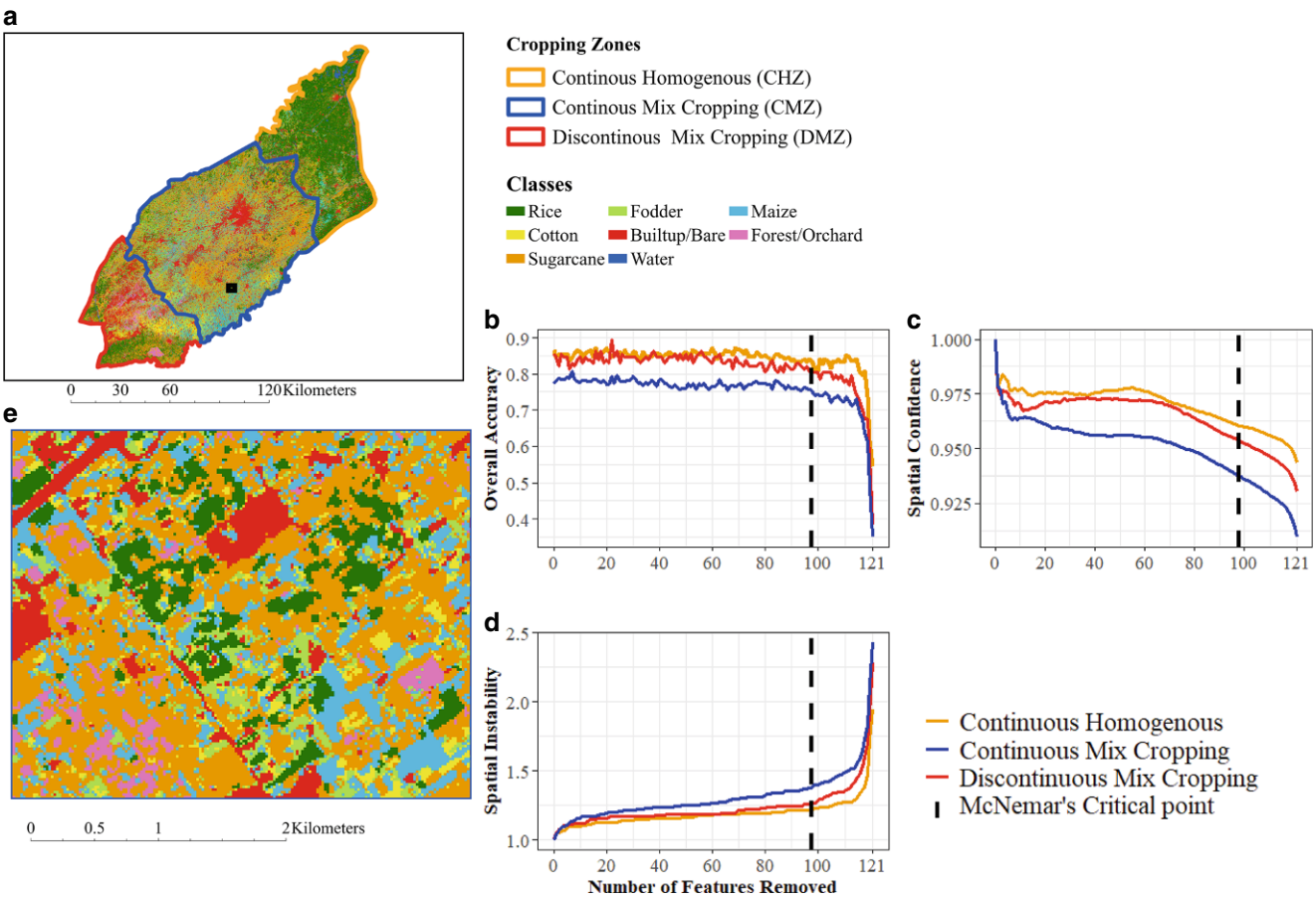

Fig. 10 (**a**) Final crop type map based on important features remaining after the critical point; (**b**) zone-wise overall accuracy; (**c**) zone-wise spatial confidence; (**d**) spatial instability; (**e**) close-in view (*black rectangle* in Fig. 10a) showing crop distribution in the continuous mixed cropping zone (CMZ)

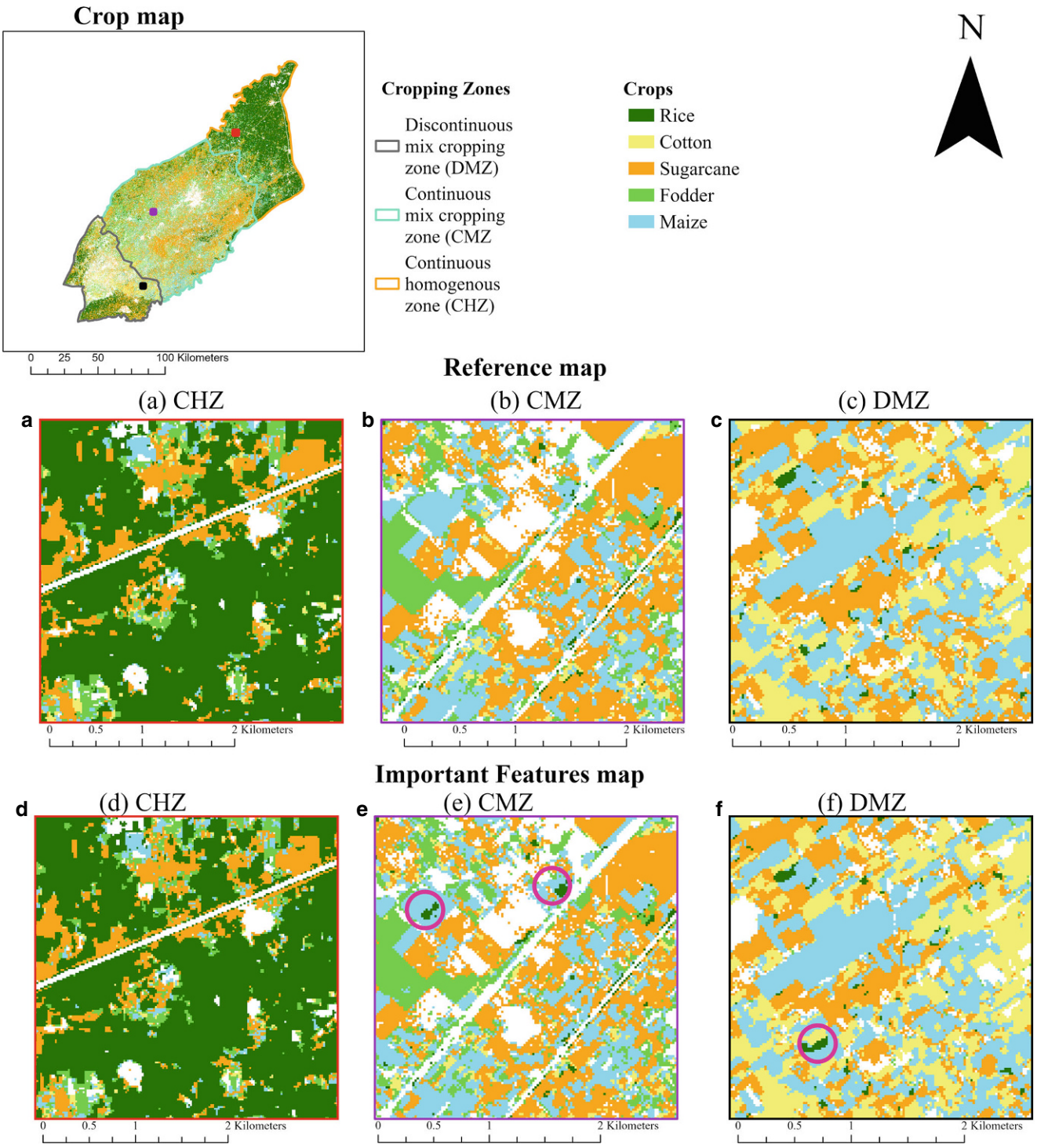

Fig. 11 Close-in-view comparison between reference and important feature maps across three cropping zones. **a**, **d** CHZ: continuous homogenous zone; **b**, **e** CMZ: continuous mixed cropping zone; **c**, **f** DMZ: discontinuous mixed cropping zone (*Red, purple, and black rectangles* indicate the zoomed locations in the CHZ, CMZ and DMZ, respectively. *Pink circles* highlight the mixing of different classes in important features classification)

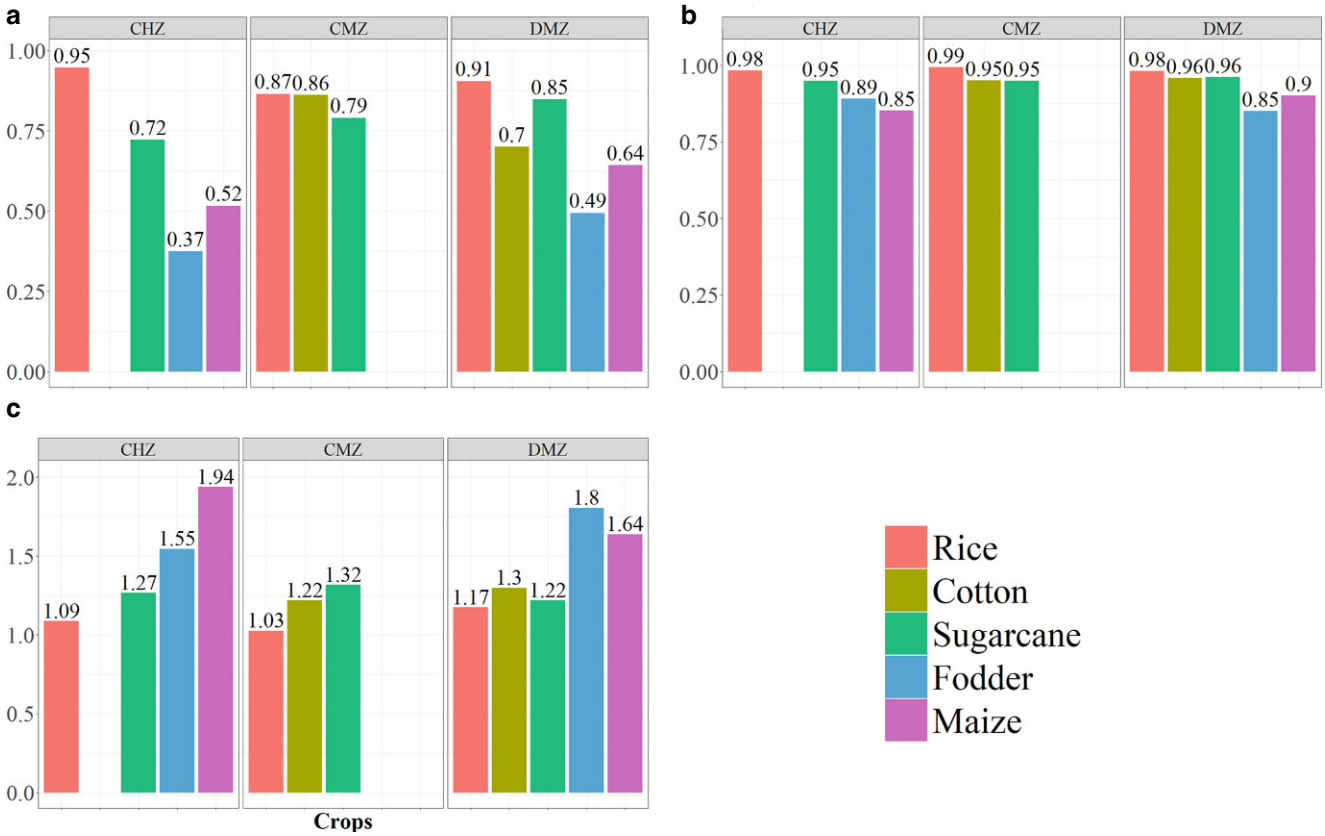

Fig. 12 Class-wise (**a**) Fscore, (**b**) spatial confidence and (**c**) spatial instability for each cropping zone in LCC with an important feature set (*CHZ* continuous homogenous zone, *CMZ* continuous mixed cropping zone, *DMZ* discontinuous mixed cropping zone)

outperformed the other crops in all cropping zones. It was followed by sugarcane, which achieved a higher Fscore in the CHZ and CMZ zones compared to other crops, while cotton had a better accuracy in the DMZ (Fscore: 0.86). However, accuracy of fodder and maize was found to be lower than that of the other crops in both zones (CHZ and CMZ).

Similar to Fscore, the most spatially confident and least spatially instable classification was achieved for rice during the feature reduction process (Fig. 12b,c). Rice had the lowest average spatial instability and the highest spatial confidence in each zone with a reduced number of features. The lowest confidence and highest instability were achieved for fodder, with a spatial confidence of 85% and an average instability of 1.8 crops per pixel during the feature reduction process in the DMZ.

4 Discussion

In the example, the LCC command area effects of feature reduction for random forest (RF) classifications of Sentinel-1 (S-1) and Sentinel-2 (S-2) data were investigated to optimize both the classification accuracy and processing time. Accuracy metrics indicate only a 2% loss of OA, with more than 80% reductions in features allowing predictions that are only insignificantly differ from a classification based on the full set of features. In contrast to previous studies (e.g., Löw et al. 2015b; Cui et al. 2020), reduced number of features did not significantly exceed the classification accuracy when using all features. However, comparable OA levels from zero to 80% feature reduction indicate redundancy and correlation among the features, as also observed by Hu et al. (2019).

The reference classification achieved an accuracy 2% higher in comparison to the reduced features with greater computational time. However, this difference in accuracy is statistically insignificant based on McNemar's test. This aligns with McNemar's test criteria established in previous studies (Kumar et al. 2017; Sitokonstantinou et al. 2018; Bueno et al. 2020) for concluding the performance of classification algorithms and datasets. Additionally, Gilbertson and van Niekerk (2017) reported that it is not necessary for feature reduction to increase the crop classification accuracy. Moreover, with a statistical benchmark, we identified the areas and crops that suffer most from feature reduction and require further consideration and optimization in

the feature reduction process (e.g., spatial accuracy assessment).

Our findings indicate that McNemar's test can identify a crucial point during the stepwise feature reduction process, which separates the important and unimportant features, beyond which the deviation of the reduced feature map and the reference map based on all features becomes continuously significant. The approach offers another perspective on feature reduction because even if OA and K indicate acceptable values, e.g., sufficient to select an optimum feature set (He et al. 2022; Htitiou et al. 2022), the reproductivity of the reduced feature map and the reference map may not be given. Especially in practical applications, it would be difficult to first offer a mapping result to a user and then a significantly different update with the same OA, possibly for a larger area, but with fewer features. The user would find that many classification decisions are different between the maps. Additionally, the area statistics differ with significantly varying error matrices, which in turn may affect land management decisions. The users may therefore ask which of the classifications is correct and possibly question the complete product.

The difference of 2% in overall accuracy between classifications based on all features and reduced features was accompanied by a decrease of spatial confidence by 5% and an increase in classification randomness. Monitoring spatial assessment metrics indicate the impacts of the feature reduction process on the mapped cropping pattern. For instance, regions characterized by high heterogeneity in crop types (such as CMZ and DMZ in our study area), and spectrally similar classes (e.g., cotton, fodder, and maize) exhibit lower reliability and greater instability than areas with uniform cropping and easily distinguishable crops (Fig. 12c). This is, e.g., in line with the findings reported by Heupel et al. (2018), and van Oort et al. (2004). For practitioners, the proposed monitoring method locates the randomness of classification, i.e., areas with mixed pixels or areas that have spectral-temporal properties but were not explicitly sampled.

4.1 Selected Features and Separability of Crop Types

Ten repetitions of 10 different sample compositions and runs showed, with few exceptions, relatively high congruence in the order of feature reduction, i.e., most features were eliminated in a certain position range of the reduction process. Interestingly, the remaining ca. 20 features were more likely to be eliminated at a similar position in the feature reduction process during all 10 repetition runs. The fact that important features covered the entire growing season from the start to the end underpinned previous observations that higher temporal resolution of satellite data during growing season is important for distinguishing different crops (Meng et al. 2020). This finding partially contradicts previous statements that satellite images representing important phenological stages are more important than the number of images in a year (Foerster et al. 2012; Conrad et al. 2014), but may be attributed to differences in the study region.

The variable importance rank and results in this study showed the high importance of optical features compared to SAR. These results are confirmed by some previous studies, where optical data also turns out to be more relevant for crop mapping than SAR data (Van Tricht et al. 2018; Demarez et al. 2019). The important features spanned the entire growing season with the highest number of features occurring in July and early August. This period corresponds to the phenological development phase in the study area, and Orynbaikyzy et al. (2020) also highlighted the importance of features during this period for crop type mapping. The selected features related to optical data are mainly from the SWIR, red and red-edge, which are very helpful for crop classification, as previously discussed, e.g., by Immitzer et al. (2012) or Orynbaikyzy et al. (2020). Additionally, the study confirmed the importance of the NDVI for crop mapping, which has been well documented by various researchers (Georganos et al. 2018; Orynbaikyzy et al. 2020), as four out of six available NDVI features were present in the final important feature set (Table 6).

The results showed that SAR features were more likely to be eliminated early (avg. S-1 Modal value: 47; avg. S-2 Modal value: 74) and showed increased variability in the feature reduction process (avg. S-1 Maximal distance: 7.8; avg. S-2 Maximal distance: 6.3) compared to optical features (Fig. 7). This observation could be attributed to their strong correlation, which was also reported by Holtgrave et al. (2020). The features (σ^0 and GLCM) from the VH polarization of the S-1 SAR data were superior to those from the VV polarization, given that all six of the selected S-1 features were derived from the VH polarization (Table 6). This is in accordance with the findings of (Chen et al. 2020), who also reported that the VH polarization outperformed VV polarization for crop mapping purposes. They also reported that GLCM texture features derived from VH backscatter performed better compared to those derived from VV backscatter.

The results of this study showed that incorporating SAR data in addition to optical data was beneficial, as was evident in the final selected features (Table 6). The presence of SAR data in selected features highlights the advantage of using both types of data, as it provides a more complete understanding of the plant structure (Veloso et al. 2017). Moreover, Forkuor et al. (2014) demonstrated that the use of SAR data reduces confusion between cotton and maize crops. In this research, a case study on crop type mapping in LCC showed a high overall accuracy, which is better than

the results of previous studies by Cheema and Bastiaanssen (2010) and Usman et al. (2015), which were based solely on optical data. In addition, optical data from S-2 were not available in July due to the monsoon season (Table 6), and having data during this time could improve the ability to distinguish between different crop types (Steinhausen et al. 2018). This is consistent with findings from previous studies (Forkuor et al. 2014; Muthukumarasamy et al. 2019), which also showed that SAR data (S-1) can compensate for the absence of optical data during these periods.

4.2 Spatial Effects of the Feature Selection Process and Implications for Field Sampling

Our results supports previous observations that mapping accuracy in heterogeneous regions and among crops with similar growth patterns is more challenging compared to the homogenous area and crops that are easily separable (Aduvukha et al. 2021). In the present study, rice exhibited highest accuracy compared to other crops due to flooding and transplanting in the early stages. This led to unique water influence signal in the satellite data, which underpinned the observations made by Yin et al. (2020). Maize and fodder (e.g., sorghum and millet) are scattered throughout the LCC and follow similar growth patterns, which causes mixed pixels and affects the accuracy of crop mapping (Cheema and Bastiaanssen 2010). Hence, because of the small field size in comparison to moderate-resolution sensors, previous studies on crop mapping in LCC, e.g., Cheema and Bastiaanssen (2010) and Usman et al. (2015), merged these phenologically similar crop classes into a single class. Similarly, the homogeneity caused by monocropping (mainly rice) resulted in the accuracy in the CHZ being better that in the CMZ and DMZ (Fig. 10).

The observations in this study suggest that the increased mapping uncertainties, in addition to the similar phenological behaviors of these crops or pixel heterogeneities, can also be attributed to the utilized ground data, as previously discussed by Löw et al. (2015a). The varying class accuracies among the detected cropping zones of LCC indicated that an increased number of training and validation samples of minor classes may have been useful but were difficult to realize under the time and financial constraints of field sampling, which coincides with the constraints observed in similar studies of extensive cropping (Ibrahim et al. 2021; Burke and Lobell 2017). For instance, the Fscore for both fodder and maize in the CMZ exceeded that in the CHZ (Fig. 12a). This is attributed to the better ground data coverage in the CMZ, highlighting that, particularly for mixed cropping regions, the use of large training datasets are desirable for improving classifier accuracy (Heydari and Mountrakis 2018). Monitoring spatial assessment metrics, spatial instability and confidence can help to determine those areas where additional ground data are needed to improve the results.

In contrast, our results indicated that the CMZ exhibited lower spatial confidence and higher spatial instability compared to the DMZ (Fig. 10d), despite having more ground truth points. This could be due to the complexity of mixed pixels in the area and the fact that simply having a larger number of samples may not be adequate for resolving this issue. It remains important to consider the quality of the ground truth data and, in particular, the representativeness and effectiveness of the applied classification model, as demonstrated by Foody and Mathur (2004) in their work on support vector machines.

4.3 Research Perspective

Further steps could involve implementing a wider range of RF-based feature selection techniques (Speiser et al. 2019) with the inclusion of spatial assessment metrics to assess their spatial effects. Additionally, the phenology-based feature reduction method may address correlated features more effectively than the RF-based feature reduction process (Hariharan et al. 2018). By diversifying feature reduction methods and integrating spatial assessment metrics, different aspects of feature importance and interactions can be highlighted, leading to a more comprehensive identification of relevant features.

The ground data used in this study was particularly limited for minor crops such as fodder and maize due to insufficient resources and the strategic challenges in obtaining data for these crops. This limitation can introduce systematic uncertainties and biases (Foody 2010). While the reproducibility of results related to accuracy and feature rankings demonstrates the reliability of the methods used and results presented, this may still impact the model's performance and its ability to accurately assess variable importance, potentially leading to a biased or incomplete representation of the key features relevant to the study (Millard and Richardson 2015). Therefore, a well-designed and systematic sampling method that includes spatially distributed samples across the study area, with an adequate number of observations across different strata, leads to improved accuracy and smaller standard errors (Stehman and Foody 2019). Additionally, advanced technologies provide access to various high-resolution images with sufficient temporal representation, allowing the selection of large sample sizes for improved accuracy in assessing large geographic areas with heterogeneous or mixed classes (Wu et al. 2017; Ramezan et al. 2019).

We utilized 10 distinct stratified random sampling splits to assess the RF variable importance based on backward feature reduction, to evaluate the stability and effectiveness of the selected features across different data partitions.

Although this approach demonstrated reliability through reproducibility, incorporating additional methods such as leave-one-out (LOO) cross-validation or k-fold cross-validation could further enhance its robustness (Ramezan et al. 2019). Known for its superior performance over other sampling techniques, k-fold cross-validation can reduce the variability in feature importance and performance estimates caused by random data partitioning (Basha and Rajput 2018). Additionally, integrating an entropy-based measure to assess the stability of feature selection methods concerning perturbations in the data (Křížek et al. 2007) could provide further insights into the reliability of the feature selection process.

The accuracy of a model usually refers to the area and the year in which it was trained. Spatial or temporal transferability is usually limited by seasonal variability, crop management practices, weather conditions, and phenological development. Additionally, technical aspects such as image acquisition dates or overfitting of the models must also be considered (e.g., Meyer et al. 2019; Orynbaikyzy et al. 2022). Thus, although the analysis of the feature selection process can support the generation of accurate models for reduced feature sets, the spatial and temporal transfer of the classification model remains an open task.

5 Conclusions

With a growing amount of optical and SAR data, increasing possibilities for integrating multispectral and spatial satellite data, and improved machine learning algorithms, the need to use feature selection processes has become very important in remote sensing classification applications. With a focus on crop mapping, we proposed a feature selection process that involves the machine learning algorithm RF and a statistical analysis at the example of the LCC command area in Pakistan. The RF variable importance coupled with McNemar's statistical test was used for backward feature elimination of unimportant features. Furthermore, we monitored spatial consistency and instability during and after the feature reduction process to better understand its effects on the mapping results. This was achieved by incorporating spatial accuracy analysis along with error matrix. The study has pointed out the following conclusions.

1. The proposed feature reduction process indicates consistency during repetition runs, and the use of McNemar's test can be recommended to identify critical points beyond which standard metrics such as overall accuracy may be high, but the resulting maps significantly differ from a map based on all features.
2. In addition to achieving reduced loss in overall accuracy, the feature selection process can strongly impact the spatial confidence of the produced maps and increase the randomness of a classification, particularly in mixed cropping zones or parts of the study area with reduced quantity or quality of ground reference data.
3. The spatial accuracy metrics between classifications produced using all available features and important features provided additional insight for the selection of important features based on an existing reference dataset, particularly when looking at different configurations of cropping zones, i.e., field sizes, compactness and crop composition.
4. The use of spatial accuracy metrics can provide better information on which crops and cropping zones require more attention for planning ground data collection.
5. At least in the large-scale study area in Pakistan, consistent temporal data available during the cropping season are very important for accurate crop mapping. The optical data proved to be more important than the SAR data; however, adding SAR data during the monsoon period, when optical data was unavailable, improved the overall accuracy of crop mapping in the study area.

6 Appendix

Name of Features	Repetitions 1	2	3	4	5	6	7	8	9	10	Modal Value	Maximal Distance	Source
20170611.σ0_VH	4	1	4	1	1	1	2	1	3	1	1	3	SAR
20170506.B8	5	6	10	9	9	2	3	8	5	10	5	5	Optical
20170518.σ0_VV	11	8	2	4	2	3	12	9	15	17	2	15	SAR
20170611.GLCM_σ0_VH	2	2	6	2	5	4	1	4	1	5	2	4	SAR
20170611.GLCM_σ0_VV	13	19	1	10	18	5	11	11	13	9	13	6	SAR
20170506.B7	3	11	7	5	6	6	10	6	6	2	6	5	Optical
20170623.σ0_VV	24	18	21	11	7	7	17	18	11	7	7	17	SAR
20170506.B11	6	3	8	15	4	8	6	5	7	3	6	9	Optical
20170506.B6	1	5	5	13	10	9	4	2	2	4	5	8	Optical
20170623.GLCM_σ0_VV	10	16	11	7	20	10	15	10	10	8	10	10	SAR
20170530.σ0_VV	9	14	15	8	12	11	5	14	12	15	14	1	SAR
20170518.GLCM_σ0_VV	7	9	3	6	11	12	8	3	14	6	3	11	SAR
20170506.B8A	14	4	23	3	3	13	7	7	8	11	3	20	Optical
20170506.GLCM_σ0_VV	21	21	20	12	8	14	14	13	9	22	21	1	SAR
20170530.GLCM_σ0_VV	8	12	17	21	13	15	13	15	4	12	12	9	SAR
20171008.B2	18	13	27	19	24	16	24	24	17	23	24	3	Optical
20170611.σ0_VV	17	20	13	18	19	17	16	16	27	19	17	10	SAR
20171008.B4	16	10	34	20	15	18	19	19	16	21	16	18	Optical
20170729.σ0_VV	27	27	24	30	26	19	21	25	21	25	27	3	SAR
20170506.B5	26	7	9	14	16	20	9	20	20	16	20	6	Optical
20170705.σ0_VV	25	22	16	27	25	21	27	21	23	14	25	2	SAR
20171009.σ0_VV	23	17	12	22	14	22	20	12	18	18	12	11	SAR
20170705.GLCM_σ0_VV	15	23	22	24	22	23	23	22	19	29	23	6	SAR
20171021.σ0_VH	20	24	19	28	27	24	25	23	26	26	24	4	SAR
20171021.GLCM_σ0_VH	19	28	28	25	23	25	30	28	28	20	28	2	SAR
20170506.σ0_VV	12	15	14	17	17	26	26	27	30	13	17	13	SAR
20171009.GLCM_σ0_VV	22	25	25	16	21	27	18	17	25	24	25	2	SAR
20170506.B4	36	31	41	34	32	28	34	41	33	34	34	7	Optical
20171009.σ0_VH	31	39	31	33	33	29	32	35	24	31	31	8	SAR
20171009.GLCM_σ0_VH	30	29	26	26	35	30	31	29	36	27	30	6	SAR
20170729.GLCM_σ0_VV	29	32	29	32	31	31	35	26	31	28	31	4	SAR
20170927.GLCM_σ0_VH	34	26	32	29	29	32	33	33	32	33	32	2	SAR
20171008.ndvi	35	36	33	35	28	33	22	31	40	30	35	5	Optical
20170927.σ0_VH	28	30	30	31	30	34	36	37	29	36	30	7	SAR
20170915.σ0_VH	43	38	40	43	36	35	37	38	43	43	43	0	SAR
20170506.B12	33	34	18	23	34	36	28	30	22	32	34	2	Optical
20170915.GLCM_σ0_VV	51	47	59	56	55	37	48	53	34	55	55	4	SAR
20170506.B3	55	44	47	55	49	38	49	44	45	46	55	0	Optical
20170530.σ0_VH	53	41	39	39	63	39	43	51	60	57	39	24	SAR
20171021.σ0_VV	39	33	35	40	43	40	40	36	41	35	40	3	SAR
20170927.σ0_VV	42	51	37	41	40	41	47	40	42	37	42	9	SAR
20170615.B6	45	54	54	45	46	42	45	55	49	50	45	10	Optical
20171008.B3	32	35	44	36	37	43	29	32	35	39	32	12	Optical
20170915.σ0_VV	47	48	52	37	51	44	53	43	53	48	48	5	SAR
20171021.GLCM_σ0_VV	59	56	45	49	48	45	38	49	39	49	49	10	SAR
20170918.B8	40	45	43	38	41	46	39	42	37	44	40	6	Optical
20170717.σ0_VV	38	37	42	48	44	47	42	39	38	51	38	13	SAR
20170717.GLCM_σ0_VV	56	43	46	47	38	48	41	45	52	42	56	0	SAR
20170526.B11	54	42	38	63	45	49	46	46	50	58	46	17	Optical
20170810.GLCM_σ0_VH	52	58	49	50	61	50	58	61	48	45	58	3	SAR
20170915.GLCM_σ0_VH	61	57	51	51	52	51	56	48	66	47	51	15	SAR
20170705.σ0_VH	48	66	63	52	39	52	64	67	62	74	52	22	SAR
20171008.B8	50	67	62	61	54	53	61	63	58	52	61	6	Optical
20170526.B6	49	49	57	65	60	54	66	59	57	56	49	17	Optical
20170918.B8A	57	52	55	53	50	55	62	50	47	53	55	7	Optical
20171008.B8A	60	65	66	54	64	56	57	71	59	60	60	11	Optical
20170918.B7	41	55	36	58	42	57	50	57	46	41	41	17	Optical
20171008.B7	63	62	61	44	66	58	52	56	61	62	62	4	Optical
20170903.GLCM_σ0_VV	44	46	56	46	57	59	44	47	51	38	44	15	SAR
20170810.σ0_VH	37	40	50	42	53	60	55	34	44	40	40	20	SAR
20170903.σ0_VH	58	60	60	57	58	61	51	52	55	63	58	5	SAR
20170526.B8	75	64	69	62	70	62	59	70	56	66	62	13	Optical
20170615.B8	81	59	73	73	72	63	73	65	78	79	73	8	Optical
20170615.B11	66	72	71	70	80	64	60	82	70	72	72	10	Optical

20170903.σ0_VV	62	50	53	59	65	65	54	58	54	59	59	6	SAR
20171008.B6	65	68	64	66	62	66	65	54	64	61	65	3	Optical
20170903.GLCM_σ0_VH	76	74	79	80	82	67	79	77	82	91	79	12	SAR
20170623.σ0_VH	70	73	78	64	77	68	75	72	86	67	70	16	SAR
20170615.B3	83	84	75	82	91	69	71	85	67	71	71	20	Optical
20170506.B2	73	76	65	71	75	70	68	62	79	65	65	14	Optical
20170526.B8A	68	63	48	68	47	71	67	64	63	54	68	3	Optical
20170506.GLCM_σ0_VH	80	82	72	83	78	72	83	86	72	68	72	14	SAR
20170809.B7	67	70	68	67	59	73	77	68	68	70	68	9	Optical
20170526.B7	77	61	74	74	74	74	76	73	75	86	74	12	Optical
20170530.GLCM_σ0_VH	64	53	77	60	79	75	63	81	65	69	64	17	SAR
20170506.σ0_VH	46	71	58	78	71	76	70	60	69	64	71	7	SAR
20170809.B6	82	85	76	72	89	77	80	83	73	83	83	6	Optical
20170615.B8A	93	79	83	85	85	78	88	75	81	75	85	8	Optical
20170526.B5	79	86	87	92	90	79	89	87	89	87	87	5	Optical
20170927.GLCM_σ0_VV	69	69	80	69	76	80	69	69	77	78	69	11	SAR
20171008.B5	90	89	93	90	94	81	94	90	84	93	90	4	Optical
20171008.B12	89	80	67	81	67	82	78	74	87	89	89	0	Optical
20170526.B3	87	88	88	88	81	83	93	66	74	90	88	5	Optical
20170809.B8	74	83	86	86	68	84	82	79	76	81	86	0	Optical
20170810.GLCM_σ0_VV	72	81	82	77	69	85	74	80	80	85	85	0	SAR
20170918.B6	85	77	81	76	73	86	86	92	88	80	86	6	Optical
20170526.B2	99	75	98	93	100	87	99	99	94	96	99	1	Optical
20170615.B7	84	91	96	79	84	88	84	89	85	73	84	12	Optical
20170615.B2	86	95	85	94	97	89	85	84	83	77	85	12	Optical
20170809.B4	96	92	89	87	86	90	95	88	95	97	95	2	Optical
20170518.σ0_VH	71	87	70	75	56	91	72	76	71	76	71	20	SAR
20170623.GLCM_σ0_VH	78	93	90	84	83	92	81	78	90	84	78	15	SAR
20170615.B5	102	105	99	102	99	93	104	101	101	98	102	3	Optical
20170615.B12	95	78	84	89	87	94	90	93	99	82	95	4	Optical
20170810.σ0_VV	98	94	91	98	92	95	96	95	96	95	95	3	SAR
20170809.ndvi	88	90	95	95	96	96	91	94	97	94	95	2	Optical
20170809.B2	100	97	100	99	98	97	101	98	92	105	100	5	Optical
20170918.B2	91	99	92	96	88	98	97	97	91	92	91	8	Optical
20170809.B3	92	96	94	91	93	99	87	91	93	88	91	8	Optical
20170918.ndvi	97	101	103	103	95	100	92	102	98	101	101	2	Optical
20170705.GLCM_σ0_VH	104	98	97	100	104	101	102	96	102	102	102	2	SAR
20170809.B8A	101	100	102	97	101	102	100	103	103	103	103	0	Optical
20170526.B12	94	103	104	107	103	103	103	104	104	104	103	4	Optical
20170918.B5	103	104	105	101	102	104	98	105	100	100	104	1	Optical
20170518.GLCM_σ0_VH	107	102	101	106	105	105	105	100	105	99	105	2	SAR
20170526.B4	108	108	108	105	108	106	108	107	107	106	108	0	Optical
20170615.B4	106	106	107	104	107	107	107	108	108	107	107	1	Optical
20171008.B11	109	110	112	109	109	108	109	114	110	110	109	5	Optical
20170918.B3	105	107	106	108	106	109	106	106	106	108	106	3	Optical
20170729.σ0_VH	110	109	109	111	113	110	110	110	109	109	110	3	SAR
20170918.B4	113	112	111	112	116	111	118	109	111	112	112	6	Optical
20170809.B5	111	113	113	114	115	112	111	112	112	111	111	4	Optical
20170918.B12	112	114	110	110	112	113	112	111	113	114	112	2	Optical
20170717.σ0_VH	114	111	116	115	110	114	113	113	114	113	114	2	SAR
20170506.ndvi	115	116	114	113	111	115	114	115	115	115	115	1	Optical
20170615.ndvi	116	117	115	118	118	116	115	116	117	116	116	2	Optical
20170717.GLCM_σ0_VH	119	120	119	120	119	117	116	120	116	118	119	1	SAR
20170809.B12	118	115	117	117	117	118	119	117	118	117	117	2	Optical
20170729.GLCM_σ0_VH	120	119	120	119	120	119	120	119	120	120	120	0	SAR
20170526.ndvi	117	118	118	116	114	120	117	118	119	119	118	2	Optical
20170809.B11	121	121	121	121	121	121	121	121	121	121	121	0	Optical
20170918.B11	122	122	122	122	122	122	122	122	122	122	122	0	Optical

Maximal Distance
0-2
3-5
6-9
> 9

Un-Important Features
Important Features

Fig. 13 Rank of features in the ten implemented repetitions of backward feature reduction (Final selected Repetition order removed (6th Repetition)), Format: YYYYMMDD.Featurename

Acknowledgements The first author is thankful to the Higher Education Commission (HEC) Pakistan and the German Academic Exchange Service (DAAD) for funding this study under the Overseas Scholarship Program. The authors sincerely thank the reviewers for their valuable feedback, which improved the quality of the manuscript.

Funding The authors are also thankful to the InoCottonGrow project funded by the German Ministry for Education and Research (BMBF, funding code: 02WGR1422E) for providing the funds to carry out the field visits for the accomplishment of this study.

Author Contribution Conceptualization: T. Mahmood, C. Conrad; methodology: T. Mahmood, C. Conrad; Data collection and analysis: T. Mahmood, M. Usman; writing—original draft preparation: T. Mahmood; visualization: T. Mahmood; writing—review and editing: T. Mahmood, C. Conrad, M. Usman; Supervision and resources: C. Conrad.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Sentinel-1 and 2 images are publicly and freely available. The ground data collected during the field visit of the study area is available from the authors upon reasonable request.

Conflict of interest T. Mahmood, M. Usman and C. Conrad declare that they have no competing interests.

References

Aduvukha GR, Abdel-Rahman EM, Sichangi AW et al (2021) Cropping pattern mapping in an agro-natural heterogeneous landscape using sentinel-2 and sentinel-1 satellite datasets. Agriculture 11:1–22. https://doi.org/10.3390/agriculture11060530

Awan UK, Ismaeel A (2020) A new technique to map groundwater recharge in irrigated areas using a SWAT model under changing climate. J Hydrol 519:1368–1382. https://doi.org/10.1016/j.jhydrol.2014.08.049

Basha SM, Rajput DS (2018) Evaluating the importance of each feature in the classification task. In: Proc—2018 8th Int Conf Commun Syst Netw Technol CSNT 2018, pp 151–155 https://doi.org/10.1109/CSNT.2018.8820216

Belgiu M, Drăgu L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Breiman L et al (1984) Classification and regression trees. Taylor & Francis, New York

Bueno IT, Mcdermid GJ, Silveira EMO et al (2020) Spatial agreement among vegetation disturbance maps in tropical domains using landsat time series. Remote Sens 12:2948. https://doi.org/10.3390/rs12182948

Burke M, Lobell DB (2017) Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc Natl Acad Sci U S A 114:2189–2194. https://doi.org/10.1073/pnas.1616919114

Cheema MJM, Bastiaanssen WGM (2010) Land use and land cover classification in the irrigated Indus Basin using growth phenology information from satellite data to support water management analysis. Agric Water Manag. https://doi.org/10.1016/j.agwat.2010.05.009

Chen Q, Meng Z, Liu X et al (2018) Decision variants for the automatic determination of optimal feature subset in RF-RFE. Genes. https://doi.org/10.3390/genes9060301

Chen S, Useya J, Mugiyo H (2020) Decision-level fusion of Sentinel-1 SAR and Landsat 8 OLI texture features for crop discrimination and classification: case of Masvingo, Zimbabwe. Heliyon 6:e5358. https://doi.org/10.1016/j.heliyon.2020.e05358

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46. https://doi.org/10.1177/001316446002000104

Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data: principles and practices, 2nd edn. CRC Press

Conrad C, Rahmann M, Machwitz M et al (2013) Satellite based calculation of spatially distributed crop water requirements for cotton and wheat cultivation in Fergana Valley, Uzbekistan. Glob Planet Change 110:88–98. https://doi.org/10.1016/j.gloplacha.2013.08.002

Conrad C, Dech S, Dubovyk O et al (2014) Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. Comput Electron Agric 103:63–74. https://doi.org/10.1016/j.compag.2014.02.003

Conrad C, Schönbrodt-stitt S, Löw F et al (2016) Cropping intensity in the Aral sea basin and its dependency from the runoff formation 2000–2012. Remote Sens. https://doi.org/10.3390/rs8080630

Cui J, Zhang X, Wang W, Wang L (2020) Integration of optical and sar remote sensing images for crop-type mapping based on a novel object-oriented feature selection method. Int J Agric Biol Eng 13:178–190. https://doi.org/10.25165/j.ijabe.20201301.5285

Demarez V, Helen F, Marais-Sicre C, Baup F (2019) In-season mapping of irrigated crops using landsat 8 and sentinel-1 time series. Remote Sens 11:118. https://doi.org/10.3390/RS11020118

Foerster S, Kaden K, Foerster M, Itzerott S (2012) Crop type mapping using spectral-temporal profiles and phenological information. Comput Electron Agric 89:30–40. https://doi.org/10.1016/j.compag.2012.07.015

Foody GM (2010) Assessing the accuracy of land cover change with imperfect ground reference data. Remote Sens Environ 114:2271–2285. https://doi.org/10.1016/j.rse.2010.05.003

Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sens 42:1335–1343. https://doi.org/10.1109/TGRS.2004.827257

Forkuor G, Conrad C, Thiel M et al (2014) Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern Benin, West Africa. Remote Sens 6:6472–6499. https://doi.org/10.3390/rs6076472

Gascon F, Bouzinac C, Thépaut O et al (2017) Copernicus Sentinel-2A calibration and products validation status. Remote Sens. https://doi.org/10.3390/rs9060584

Georganos S, Grippa T, Vanhuysse S et al (2018) Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban appli-

cation. GIscience Remote Sens 55:221–242. https://doi.org/10.1080/15481603.2017.1408892
Ghazaryan G, Dubovyk O, Löw F et al (2018) A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. Eur J Remote Sens 51:511–524. https://doi.org/10.1080/22797254.2018.1455540
Gilbertson JK, van Niekerk A (2017) Value of dimensionality reduction for crop differentiation with multi-temporal imagery and machine learning. Comput Electron Agric 142:50–58. https://doi.org/10.1016/j.compag.2017.08.024
Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random forests for land cover classification. Pattern Recognit Lett 27:294–300. https://doi.org/10.1016/j.patrec.2005.08.011
Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182. https://doi.org/10.1162/153244303322753616
Hamzeh S, Mokarram M, Haratian A et al (2016) Feature selection as a time and cost-saving approach for land suitability classification (Case study of Shavur Plain, Iran). Agriculture. https://doi.org/10.3390/agriculture6040052
Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Syst Man Cybern 3:610–621. https://doi.org/10.1109/TSMC.1973.4309314
Hariharan S, Mandal D, Tirodkar S et al (2018) A novel Phenology based feature subset selection technique using random forest for Multitemporal PolSAR crop classification. IEEE J Sel Top Appl Earth Obs Remote Sens 11:4244–4258. https://doi.org/10.1109/JSTARS.2018.2866407
He S, Peng P, Chen Y, Wang X (2022) Multi-crop classification using feature selection-coupled machine learning classifiers based on spectral, textural and environmental features. Remote Sens 14:1–17. https://doi.org/10.3390/rs14133153
Heupel K, Spengler D, Itzerott S (2018) A progressive crop-type classification using multitemporal remote sensing data and phenological information. Remote Sens 86:53–69. https://doi.org/10.1007/s41064-018-0050-7
Heydari SS, Mountrakis G (2018) Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. Remote Sens Environ 204:648–658. https://doi.org/10.1016/j.rse.2017.09.035
Holtgrave AK, Röder N, Ackermann A et al (2020) Comparing sentinel-1 and -2 data and indices for agricultural land use monitoring. Remote Sens. https://doi.org/10.3390/RS12182919
Htitiou A, Boudhar A, Lebrini Y et al (2022) A comparative analysis of different phenological information retrieved from Sentinel-2 time series images to improve crop classification: a machine learning approach. Geocarto Int 37:1426–1449. https://doi.org/10.1080/10106049.2020.1768593
Hu Q, Sulla-Menashe D, Xu B et al (2019) A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. Int J Appl Earth Obs Geoinf 80:218–229. https://doi.org/10.1016/j.jag.2019.04.014
Ibrahim ES, Rufin P, Nill L et al (2021) Mapping crop types and cropping systems in Nigeria with sentinel-2 imagery. Remote Sens 13:3523. https://doi.org/10.3390/rs13173523
Immitzer M, Atzberger C, Koukal T (2012) Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data. Remote Sens 4:2661–2693. https://doi.org/10.3390/rs4092661
Immitzer M, Vuolo F, Atzberger C (2016) First experience with sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens. https://doi.org/10.3390/rs8030166
Inglada J, Vincent A, Arias M et al (2017) Operational high resolution land cover map production at the country scale using satellite image time series. Remote Sens 9:95. https://doi.org/10.3390/rs9010095
Křížek P, Kittler J, Hlaváč V (2007) Improving stability of feature selection methods BT—computer analysis of images and patterns. In: Kropatsch WG, Kampel M, Hanbury A (eds) Springer, Berlin, Heidelberg, pp 929–936
Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621. https://doi.org/10.1080/01621459.1952.10483441
Kumar P, Prasad R, Choudhary A et al (2017) A statistical significance of differences in classification accuracy of crop types using different classification algorithms. Geocarto Int 32:206–224. https://doi.org/10.1080/10106049.2015.1132483
Lemoine G, Léo O (2015) Crop mapping applications at scale: using Google earth engine to enable global crop area and status monitoring using free and open data sources. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp 1496–1499. https://doi.org/10.1109/IGARSS.2015.7326063
Löw F, Duveiller G (2014) Defining the spatial resolution requirements for crop identification using optical remote sensing. Remote Sens 6:9034–9063. https://doi.org/10.3390/rs6099034
Löw F, Michel U, Dech S, Conrad C (2013) Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. ISPRS J Photogramm Remote Sens 85:102–119. https://doi.org/10.1016/j.isprsjprs.2013.08.007
Löw F, Fliemann E, Abdullaev I et al (2015a) Mapping abandoned agricultural land in Kyzyl-Orda, Kazakhstan using satellite remote sensing. Appl Geogr 62:377–390. https://doi.org/10.1016/j.apgeog.2015.05.009
Löw F, Knöfel P, Conrad C (2015b) Analysis of uncertainty in multitemporal object-based classification. ISPRS J Photogramm Remote Sens 105:91–106. https://doi.org/10.1016/j.isprsjprs.2015.03.004
Mazzia V, Khaliq A, Chiaberge M (2020) Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-Convolutional Neural Network (R-CNN). Appl Sci 10:1–23. https://doi.org/10.3390/app10010238
McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12:153–157. https://doi.org/10.1007/BF02295996
Meng S, Zhong Y, Luo C et al (2020) Optimal temporal window selection for winter wheat and rapeseed mapping with sentinel-2 Images: A case study of Zhongxiang in China. Remote Sens. https://doi.org/10.3390/rs12020226
Meyer H, Reudenbach C, Wöllauer S, Nauss T (2019) Importance of spatial predictor variable selection in machine learning applications—moving from data reproduction to spatial prediction. Ecol Modell 411:108815. https://doi.org/10.1016/j.ecolmodel.2019.108815
Millard K, Richardson M (2015) On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sens 7:8489–8515. https://doi.org/10.3390/rs70708489
Miyazaki H, Iwao K, Shibasaki R (2011) Development of a new ground truth database for global urban area mapping from a gazetteer. Remote Sens 3:1177–1187. https://doi.org/10.3390/rs3061177
de Moraes MB, Gradvohl ALS (2021) A comparative study of feature selection methods for binary text streams classification. Evol Syst 12:997–1013. https://doi.org/10.1007/s12530-020-09357-y
Mujtaba A, Nabi G, Masood M et al (2022) Impact of cropping pattern and climatic parameters in lower chenab canal system—case study from Punjab Pakistan. Agric. https://doi.org/10.3390/agriculture12050708

Muthukumarasamy I, Shanmugam R, Usha T (2019) Incorporation of textural information with SAR and optical imagery for improved land cover mapping. Environ Earth Sci 78:643. https://doi.org/10.1007/s12665-019-8654-9

Nosratabadi S, Ardabili S, Lakner Z et al (2021) Prediction of food production using machine learning algorithms of multilayer perceptron and anfis. Agric 11:1–13. https://doi.org/10.3390/agriculture11050408

Olofsson P, Foody GM, Herold M et al (2014) Good practices for estimating area and assessing accuracy of land change. Remote Sens Environ 148:42–57. https://doi.org/10.1016/j.rse.2014.02.015

van Oort PAJ, Bregt AK, de Bruin S et al (2004) Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database. Int J Geogr Inf Sci 18:611–626. https://doi.org/10.1080/13658810410001701969

Orynbaikyzy A, Gessner U, Conrad C (2019) Crop type classification using a combination of optical and radar remote sensing data: a review. Int J Remote Sens 40:6553–6595. https://doi.org/10.1080/01431161.2019.1569791

Orynbaikyzy A, Gessner U, Mack B, Conrad C (2020) Crop type classification using fusion of sentinel-1 and sentinel-2 data: assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. Remote Sens 12:2779. https://doi.org/10.3390/rs12172779

Orynbaikyzy A, Gessner U, Conrad C (2022) Spatial transferability of random forest models for crop type classification using sentinel-1 and sentinel-2. Remote Sens. https://doi.org/10.3390/rs14061493

Ramezan CA, Warner TA, Maxwell AE (2019) Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sens. https://doi.org/10.3390/rs11020185

Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23:2507–2517. https://doi.org/10.1093/bioinformatics/btm344

Sheykhmousa M, Mahdianpari M, Ghanbari H et al (2020) Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review. IEEE J Sel Top Appl Earth Obs Remote Sens 13:6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724

Shi S, Ye Y, Xiao R (2022) Evaluation of food security based on remote sensing data—taking Egypt as an example. Remote Sens. https://doi.org/10.3390/rs14122876

da Silva FG, Ramos LP, Palm BG, Machado R (2022) Assessment of machine learning techniques for oil rig classification in C-band SAR images. Remote Sens 14:1–24. https://doi.org/10.3390/rs14132966

Sishodia RP, Ray RL, Singh SK (2020) Applications of remote sensing in precision agriculture: a review. Remote Sens 12:1–31. https://doi.org/10.3390/rs12193136

Sitokonstantinou V, Papoutsis I, Kontoes C et al (2018) Scalable parcel-based crop identification scheme using Sentinel-2 data time-series for the monitoring of the common agricultural policy. Remote Sens. https://doi.org/10.3390/rs10060911

Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 134:93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Stehman SV, Foody GM (2019) Key issues in rigorous accuracy assessment of land cover products. Remote Sens Environ 231:111199. https://doi.org/10.1016/j.rse.2019.05.018

Steinhausen MJ, Wagner PD, Narasimhan B, Waske B (2018) Combining sentinel-1 and sentinel-2 data for improved land use and land cover mapping of monsoon regions. Int J Appl Earth Obs Geoinf 73:595–604. https://doi.org/10.1016/j.jag.2018.08.011

Treitz PM, Howarth PJ, Filho OR (2014) Agricultural crop classification using SAR tone and texture statistics. Can J Remote Sens 26:18–29. https://doi.org/10.1080/07038992.2000.10874751

Ulaby FT, Moore RK, Fung AK (1982) Radar remote sens surf scatt emiss theory. Microwave remote sensing: active and passive, vol II

Usman M, Liedl R, Shahid MA, Abbas A (2015) Land use/land cover classification and its change detection using multi-temporal MODIS NDVI data. J Geogr Sci 25:1479–1506. https://doi.org/10.1007/s11442-015-1247-y

Usman M, Talha M, Conrad C (2018) Evaluation of MODIS data for mapping of major crop types in semi-arid Punjab of Pakistan. Global Food Security and Food Safety: Role of Universities, Ghent

Van Tricht K, Gobin A, Gilliams S, Piccard I (2018) Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: a case study for Belgium. Remote Sens 10:1642. https://doi.org/10.3390/rs10101642

Veloso A, Mermoz S, Bouvet A et al (2017) Understanding the temporal behavior of crops using sentinel-1 and sentinel-2-like data for agricultural applications. Remote Sens Environ 199:415–426. https://doi.org/10.1016/j.rse.2017.07.015

Vergara JR, Estévez PA (2014) A review of feature selection methods based on mutual information. Neural Comput Appl 24:175–186. https://doi.org/10.1007/s00521-013-1368-0

Wright MN, Ziegler A (2017) ranger: a fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw 77:1–17. https://doi.org/10.18637/jss.v077.i01

Wu M, Huang W, Niu Z et al (2017) Fine crop mapping by combining high spectral and high spatial resolution remote sensing data in complex heterogeneous areas. Comput Electron Agric 139:1–9. https://doi.org/10.1016/j.compag.2017.05.003

Yang S, Gu L, Li X et al (2020) Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. Remote Sens 12:1–23. https://doi.org/10.3390/RS12193119

Yi Z, Jia L, Chen Q (2020) Crop classification using multi-temporal sentinel-2 data in the Shiyang river basin of China. Remote Sens 12:1–21. https://doi.org/10.3390/rs12244052

Yin L, You N, Zhang G et al (2020) Optimizing Feature Selection of Individual Crop Types for Improved Crop Mapping. Remote Sens 12:162. https://doi.org/10.3390/rs12010162