IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 8, 2018, accepted September 18, 2018, date of publication October 2, 2018, date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872801

Ensemble Feature Selection for Plant Phenotyping: A Journey From Hyperspectral to Multispectral Imaging

ALI MOGHIMI, CE YANG, AND PETER M. MARCHETTO, (Member, IEEE)
Department of Bioproducts and Biosystems Engineering, University of Minnesota Twin Cities, Saint Paul, MN 55108, USA

Corresponding author: Ce Yang (ceyang@umn.edu)

This work was supported by the Minnesota's Discovery, Research, and InnoVation Economy (MnDRIVE) program through the research area of Robotics, Sensors, and Advanced Manufacturing.

ABSTRACT Hyperspectral imaging is becoming an increasingly popular tool for high-throughput plant phenotyping, because it provides remarkable insights about the health status of plants. Feature selection is a key component in a hyperspectral image analysis, largely because a significant portion of spectral features are redundant and/or irrelevant, depending on the desired application. This paper presents an ensemble feature selection method to identify the most informative spectral features for practical applications in plant phenotyping. The hyperspectral data set contained the images of four wheat lines, each with a control and a salt (NaCl) treatment. To rank spectral features, six feature selection methods were used as the base for the ensemble: correlation-based feature selection, ReliefF, sequential feature selection, support vector machine-recursive feature elimination (SVM-RFE), LASSO logistic regression, and random forest. The best results were achieved by the ensemble of ReliefF, SVM-RFE, and random forest, which drastically reduced the dimension of the hyperspectral data set from 215 to 15 features, while improving the accuracy in classifying the salt-treated vegetation pixels from the control pixels by 8.5%. To transform the hyperspectral data set into a multispectral data set, six wavelengths as the center of broad multispectral bands around the most prominent features were determined by a clustering algorithm. The result of salt tolerance assessment of the four wheat lines using the derived multispectral data set was similar to that of the hyperspectral data set. This demonstrates that the proposed feature selection pipeline can be utilized for determining the most informative features and can be a valuable tool in the development of tailored multispectral cameras.

INDEX TERMS Band selection, classification, ensemble feature selection, hyperspectral imaging, machine learning, multispectral imging, plant phenotyping, salt stress, wheat.

I. INTRODUCTION

Hyperspectral imaging (HSI) integrates imaging and high resolution spectroscopy whereby a continuous spectrum is scanned for each pixel, typically across the visible and near infrared range of electromagnetic spectrum. Because of this unique aspect, HSI has been utilized as a fascinating and intriguing research tool in various domains including agriculture [1], [2], environmental surveillance [3], [4], mineralogy [5], [6], biotechnology [7], [8], medical diagnoses [9]–[11], and pharmaceuticals [12], [13]. HSI has recently drawn substantial attention in plant biology because it enables extensive investigation on the internal activities, physiological dynamics, and cell structure of plants [14]–[16]. Recently, hyperspectral sensors have become more compact, more lightweight, and more available, and this has introduced new opportunities for field applications because hyperspectral images can be captured via unmanned aerial vehicles.

Despite impressive progress of hyperspectral sensor development and autonomous platforms, there are still several challenges in the post-processing and analysis portion of the workflow, hindering the comprehensive application of HSI. Analysis of hyperspectral images is oftentimes challenging, mainly due to the complexity and high-dimensionality of the hyperspectral data. To find the correlation between the spectral response of plants and the desired traits, the complex and high-dimensional hyperspectral images must be interpreted appropriately. Among the hundreds of wavelengths scanned by the hyperspectral imager, a small set of wavelengths may be related to the desired traits, and the remaining are oftentimes either irrelevant or redundant. These redundant and irrelevant features make interpretation complex, and increase

risk of overfitting, computational cost, and required storage space. In addition, when applying classification algorithms to a dataset with a high dimension of features a larger number of annotated samples (e.g., pixels) are required for each class (referred to as 'the curse of dimensionality'). Another drawback of high-dimensional feature space is that the performance of subsequent modeling (e.g., classification, clustering, regression, etc.) might be deteriorated due to the presence of redundant and irrelevant features that merely expand the dimension of feature space without offering new information. Therefore, dimensionality reduction is a beneficial preprocessing step in implementing machine learning algorithms for hyperspectral image analysis.

Two main approaches exist for dimensionality reduction of a high-dimensional dataset: feature extraction and feature selection. In the feature extraction approach, all original features are combined in a transformation process to map from a high-dimension to a low-dimension feature space. For example, principal component analysis and linear discriminant analysis (two popular feature extraction methods) require all features to project to a lower-dimensional feature space. Alternatively, in feature selection, a subset containing discriminative features is selected such that it captures the most valuable information with minimum redundancy. The original semantics of features do not change using the feature selection approach, and this is advantageous because the subset is interpretable by a domain expert [17].

There are two main reasons that feature selection is preferred over feature extraction for hyperspectral image analysis. First, feature selection maintains the physical meaning associated with the measured reflectance per each wavelength as they are the functions of the object characteristics such as plants' internal/physiological activities, whereas with feature extraction, the physical information is lost through transforming to a new feature space. Secondly, all features must still be measured in feature extraction approach because each feature in the new space is a combination of all the original features. Alternatively, if a subset of the original wavelengths can be identified for a specific type of application using the feature selection approach, then a multispectral camera can be designed for that particular application to leverage the advantages incorporated with multispectral imaging. Multispectral cameras are less expensive and complex compared to hyperspectral cameras, and thus enable wider application of spectral imaging for trait analysis among plant breeders and plant scientists.

Feature selection algorithms can be classified in three general subsets: filter, wrapper, and embedded methods, each of which has advantages and drawbacks [17].

Filter methods rank features based on a statistical criteria, such as distance, correlation, and information, without involving classification/regression models (i.e., no learning is involved). Filter methods are either *univariate* or *multivariate*. In univariate methods, each single feature is ranked independently, whereas multivariate methods incorporate feature dependency [17], inferring the ability of multivariate methods in handling feature redundancy [18]. Filter methods are fast and computationally efficient, and are therefore scalable to large datasets. The major drawback of filter methods is that the selected feature subset may not be the best feature subset for classification purposes since they were selected regardless to their effects on the performance of classifiers [19]. Some of the widely used filter methods are correlation-based feature selection [20], ReliefF [20], and information gain [21].

Wrapper methods include a feature subsets search algorithm wrapped around a classification algorithm [22]. The search algorithm explores the space of all possible feature subsets and generates various feature subsets. The performance of the classifier is evaluated using the generated feature subsets, and subsequently an optimal feature subset is identified based on the classification performance. Wrapper methods tend to offer improved results compared to filter methods because the best feature subset is selected by incorporating the performance of a predefined classifier algorithm as the selection criteria. Moreover, wrappers have the ability to handle feature dependencies. These advantages of wrapper methods are achieved at the cost of losing computational efficiency and being biased towards the classifier algorithm [23]. Examples of wrappers methods include sequential feature selection [24], genetic algorithms [25], [26], and support vector machine-recursive feature elimination (SVM-RFE) [27].

In embedded methods, selecting an optimal feature subset is embedded in the construction of a classifier. Embedded methods offer a high accuracy due to the search process being guided by the learning process. However, in contrast to wrapper methods, embedded methods tend to have less computational cost because feature selection is conducted during the classifier construction (i.e., no need to iteratively run the classifier for all feature subsets). Similar to wrappers, feature selection in embedded methods inevitably depends on the classifier that was utilized, meaning that the selected feature subset may not necessarily be effective for other classifiers [28]. Random forest [29] and LASSO [30] are among the most representative algorithms of embedded methods.

Each of the feature selection approaches has advantages and disadvantages, and tend to be tailored for specific applications. Despite the availability of several feature selection approaches and a substantial number of feature selection algorithms, defining an efficient benchmark to select predominant spectral features has remained a challenge in hyperspectral image analysis [31].

This study was motivated by the need to reduce the complexity and high-dimensionality of hyperspectral images in high-throughput plant phenotyping where HSI has recently drawn a substantial attention and introduced new opportunities. The first objective of this research was to develop an ensemble feature selection pipeline to aggregate the benefits of multiple feature selection approaches, therefore increasing the stability and accuracy of selecting predominant spectral features from hyperspectral images. The second objective was to rank the spectral features based on their ability to discriminate salt-stressed wheat plants from healthy plants at

TABLE 1. Hyperspectral image dataset of four wheat lines to assess salt stress [32].

Wheat lines	Abbreviation	Number of spectral features	Number of pixels	class size	
				C0 (control)	C1 (salt)
Chinese Spring	CS	215	12896	7060	5836
speltoides aucheri KU2201B (CS)	sp(CS)	215	11665	5590	6075
Aegilops columnaris KU11-2 (CS)	co(CS)	215	5228	1878	3350
Kharchia	-	215	14652	9013	5639

the earliest stages of stress. The purpose of ranking spectral features was to form six broad spectral bands around the most prominent features to aid in development of a multispectral camera. The third objective was to use these multispectral bands to assess the salt tolerance of four wheat lines in the context of phenotyping and evaluate results based on findings in previous studies.

The remainder of the paper is organized as follows. Section II describes the hyperspectral dataset containing the images of four wheat lines with various levels of salt tolerance. Section III introduces the ensemble feature selection method proposed in this study for spectral feature selection. Section IV reports the ranking of spectral features and discusses the experimental results of salt stress phenotyping in wheat using the selected multispectral bands. Section V draws the conclusion and presents the idea of developing a digital spectral library for plant diseases and stresses in future work.

II. HYPERSPECTRAL DATASET

A hyperspectral dataset from a phenotyping study [14], freely accessible to the public at [32], was used in this study. They investigated the salt stress tolerance in four wheat lines using HSI. The hyperspectral dataset for each wheat line contained the spectral reflectance of leaves for a null control and a salt treatment (as sodium chloride; Table 1). The spatial resolution was about 1 mm and the wavelengths range was from 400 nm to 900 nm with a spectral resolution of about 2.1 nm. In this study, the dataset of the most susceptible line (Chinese Spring) was selected among the four wheat lines to find sensitive spectral features for salt stress phenotyping of wheat. Once the spectral features were identified, all four hyperspectral datasets of wheat lines were utilized to evaluate the feasibility of the selected bands in ranking the wheat lines based on their tolerance to salt stress. The matrix of the data was initially normalized to make each spectral feature have zero-mean and unit-variance.

III. PROPOSED ENSEMBLE FEATURE SELECTION METHOD

A. FEATURE SELECTION METHODS

The hyperspectral dataset of CS was split into training (70% of dataset) and test (30% of dataset) datasets (Fig. 1-step I). Training dataset was used by feature selection techniques, referred to as rankers in this study, and the test dataset was employed to evaluate all potential ensembles of rankers. The following six rankers were the individual base rankers and used to construct an ensemble feature selection pipeline (Fig. 1-step II).

1) CORRELATION-BASED FEATURE SELECTION (CFS)

CFS is a univariate filter method that measures the Pearson's correlation coefficient between features and the output variable [33]. The spectral features are ranked based on their correlation to the class. In this study, the number of components in the feature subset was considered to be one to identify the rank of each individual feature. Therefore, CFS ignored features redundancy and correlation between features since the features were treated independently.

2) RELIEFF

ReliefF is also a univariate filter method that assigns weight to features through an iterative process [34], [35]. In each iteration, an instance is randomly selected and the weight of features is updated based on the distance between the randomly selected instance and the k closest instances from each class. The weight of a given feature increases at each iteration if it makes the selected instance be more similar to its neighbors in the same class and be more distinct from its neighbors from the other class. In this study, the iteration was performed for all samples and the number of neighbors, k, was set to 10.

3) SEQUENTIAL FEATURE SELECTION - FORWARD SELECTION (SFS-FORWARD)

SFS-forward is a wrapper method to find an optimal feature subset using a greedy search algorithm. It starts with an empty subset and adds one feature through sequential iteration to the subset based on the performance of a predefined classifier. Quadratic discriminant analysis (QDA) was used in this study as the classifier. The order in which all features added to the subset denoted the ranking of features.

4) SUPPORT VECTOR MACHINE-RECURSIVE FEATURE ELIMINATION (SVM-RFE)

SVM-RFE is another wrapper method in which support vector machine is used as the classifier [36]. Liner SVM classifier is trained using the training dataset containing all features in the first iteration. Then SVM-RFE eliminates one single feature with the lowest weight among all weights required to

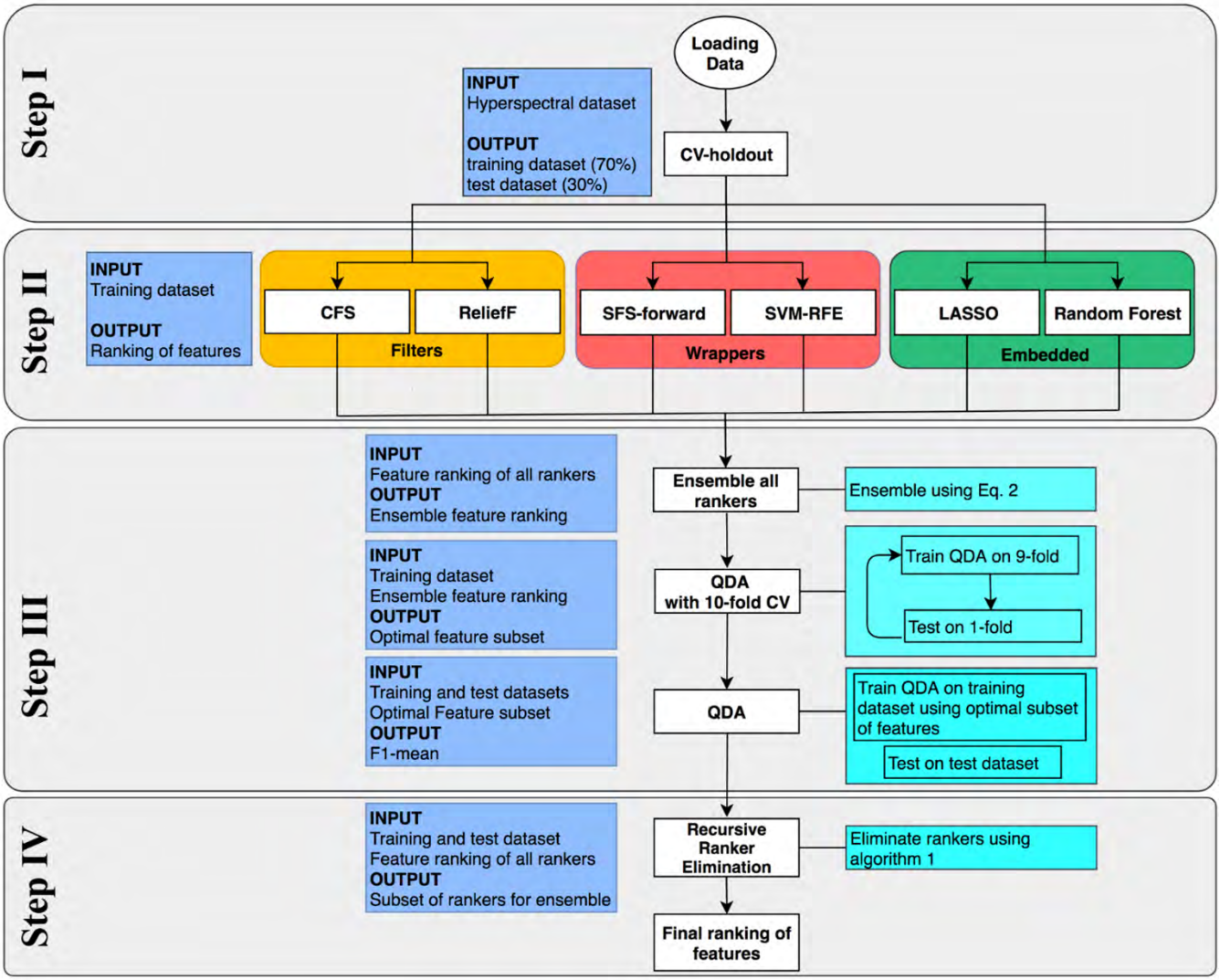

FIGURE 1. Flowchart of the proposed pipeline for ensemble feature selection. Step I) Splitting the data into training set (70%) and test set (30%) using cross validation – holdout method. Step II) Ranking the spectral features using six base rankers: correlation-based feature selection, ReliefF, sequential feature selection - forward selection, support vector machine-recursive feature elimination, and random forest. Step III) Ensemble of all six rankers. In this step, the optimal feature subset size and F1-mean for the ensemble of all six rankers are calculated. Step IV) Eliminating rankers through a recursive process presented in algorithm 1 to identify the best ensemble of the rankers.

define the best hyperplane in separating the classes. These weights are assigned by the SVM classifier to features. This procedure was recursively executed until only one feature remained in the feature dataset to rank all features based on their elimination order through the recursive process.

5) LASSO LOGISTIC REGRESSION

The idea of adding L1 regularization into the objective function of linear least-squares regression, so-called *LASSO*, was first proposed by [37] to shrink coefficients and control the model complexity. The regularization term defines an upper bound on the sum of the absolute values of the model coefficients to prevent overfitting and make the coefficient of irrelevant features equal to zero. Therefore, the coefficient vector of features becomes sparse. For classification purpose, the LASSO estimator can be added to Logistic Regression model, and coefficients (β) can be obtained as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\|Y - \beta X\|_2^2 + \lambda \|\beta\|_1 \right) \tag{1}$$

where λ is the tuning parameter which controls the balance between bias and variance. The larger it is the larger the number of coefficients are penalized to be equal to zero, and hence the variance decreases while bias increases. A number of 100 different values (between 0 and 1) for λ were tested to obtain the optimized values for coefficients ($\hat{\beta}$). Since the spectral features were first centered and scaled, the absolute value of non-zero coefficients implies the importance of features, while irrelevant features have coefficients of zero. LASSO is an embedded method for feature selection.

6) RANDOM FOREST

Random forest is another embedded method building an ensemble of de-correlated decision trees [38], each trained on a training set created by bootstrap aggregation in which samples are randomly drawn from initial training set with replacement. To split each node in each decision tree, a randomly selected subset of features is evaluated. Based on the defined split-criterion, which was entropy in this study, a feature with the best strength of classification is selected among the randomly selected feature for each node to divide the training samples into smaller and more pure subsets. The importance of features is calculated based on how much the splitting feature can decrease the weighted impurity at that node, and it is then averaged over all the trees in the forest. Number of trees was 1000 in this study.

B. ENSEMBLE FEATURE SELECTION METHOD

Each ranker has its own drawback such that the final selected feature subset may not be the optimal feature subsets in the feature space. To aggregate the benefits of each feature selection approach and increase the robustness and accuracy of the final selected feature subset, ensemble feature selection approach was used (originally proposed by [19]). Ensemble feature selection method reduce the risk of capturing local optima where a single ranker may be trapped. Various methods have been used to combine the feature ranking obtained by each base ranker [39], [40]. In this study, the complete linear aggregation method was used to average the ranking f assigned to feature i by ranker j as presented in (2), and the ranking values of features was then sorted to list features based on their importance.

$$Rank_i = \frac{1}{n}\sum_{j=1}^{n} f_{ij} \quad i = 1, 2, \ldots, 215, \text{ and n} = 6 \qquad (2)$$

In this equation, the rankers had the same weight, which is equal to one over the number of rankers. To ensure the optimal feature subset is achieved, all base rankers used for aggregation should be accurate and diverse [41]. However, based on the dataset and the nature of the problem, a given ranker may not be accurate or may contribute to redundancy in the ensemble process. Therefore, a more intelligent approach was designed to evaluate which combination of these six rankers would result in the selection of the best feature subset based on the classification of vegetation pixels into control and salt classes.

In this study, the decision on exclusion of a ranker was made through a recursive elimination process. In this greedy search, a ranker was removed at each step if its elimination resulted in the best classification performance among all choices for elimination in that step. For instance, at step 1, there were six iterations, at each of which one ranker was eliminated. At the end of step 1, the ranker whose elimination resulted in the best classification performance among the six options was eliminated for the next step. This process continued until only one ranker remained. Algorithm 1 presents a summary of the recursive elimination of rankers. Besides the ensemble of all six rankers (Fig. 1-step III), there were five other ensembles of rankers each of which was obtained at each step (Fig. 1-step IV). The combination of rankers that offered the best classification performance was selected among these six sets. This approach assures that the presence of a ranker would not deteriorate the classification result.

Two criteria related to the classification performance were defined to compare all potential ensembles of rankers at each step.

The first criterion was the size of feature subset to attain the minimum classification error of a quadratic discriminant analysis (QDA) classifier. This process was performed on the training dataset, containing 70% of the whole dataset, with 10-fold cross validation. To identify the best feature subset, the process of training and testing with QDA was conducted in an iterative process in which the ensemble feature list, obtained at iteration k by (2), was imported to QDA using a stepwise approach until all features were exhausted. Therefore, at the ith iteration, QDA was trained and tested 10 times using only the first ith features. For instance, at the first iteration, only the first feature at the top of the ensemble list was used, and all 215 features were employed at the 215th iteration.

The mean of cross validation error (CV error) and standard deviation of error over 10 folds were recorded for each iteration. The standard deviation of error indicates the error rate uncertainty; thus, one-standard error rule was used as a conservative metric to determine feature subset size [42]. Using this approach, the most parsimonious feature subset within one standard error (1SE) of the minimum error was selected as the best feature subset. In other words, the mean of error for this parsimonious subset should not be more than 1SE above the minimum error.

The second criteria was the classification accuracy of a QDA algorithm evaluated on the test dataset (the remaining 30% of the dataset), which has not been used during any of the feature selection and parameter tuning process of rankers. For this purpose, the QDA algorithm was trained on the whole training dataset using only the features obtained by the 1SE rule. The performance of the trained QDA was evaluated on the test dataset. The metric used for classification accuracy was F1-mean which was defined as a mean between F1 measures of both control ($F1_{C0}$) and salt ($F1_{C1}$) classes.

The procedure of eliminating a ranker per iteration was considered as a multiple criteria decision-making problem in which the performance of feasible ensemble of rankers, obtained at iteration k, was evaluated based on the defined criteria: feature subset size and F1-mean. To allocate weight to the alternatives, we utilized the technique for order preference by similarity to an ideal solution (TOPSIS) proposed by [43] as a simple and useful method for multiple criteria decision making problems. TOPSIS dispensed the weight among feasible ensembles at each iteration k based on the Euclidian distance of each ensemble to the negative ideal solution divided by the summation of distance to the

Algorithm 1 Recursive Ranker Elimination to Identify the Best Ensemble of the Rankers

1. Let $X_{train} = \{(x_{ij}, y_i)|i = 1, 2, \ldots, n_{trian}, j = 1, 2, \ldots, 215\}$ be the training dataset
2. Let $X_{test} = \{(x_{ij}, y_i)|i = 1, 2, \ldots, n_{test}, j = 1, 2, \ldots, 215\}$ be the test dataset
3. Let $\mathcal{F} = \{[f_{ij}]|i = 1, \ldots, 215, j = 1, 2, \ldots, n$, a matrix containing ranking of features for all n ranker
4. **For step** = 1 to (n − 1) **do**
5. **For** k = 1 to (number of columns in $\mathcal{F}$) **do**
6. $\hat{F}_k = \{[f_{ij}]|i = 1, \ldots, 215, j = \{1, 2, \ldots, 6\}k$, eliminate the kth column of F
7. E: ensemble $\hat{\mathcal{F}}_k$using Eq 2
8. R: sort features based on their ensemble ranking E
9. **For** i = 1 to 215 **do**
10. r = R[1:i], the first i features at the top of the list
11. Train and test QDA on X_{train} with 10 folds cross validation using features $\in r$
12. e_i: mean of error for 10 folds
13. std_i: standard deviation of e_i
14. **End For**
15. Err = min(e_i)
16. $SE = \frac{std_{argmin(e_i)}}{\sqrt{10}}$
17. m: Find the size of the most parsimonious feature subset whose error is less than (Err + SE)
18. Train QDA on X_{train} using only the first m feature in R
19. Test the trained QDA classifier on X_{test}
20. Calculate $F1$ measure for both C0 and C1 classes
21. F1-mean: mean of $F1_{C1}$ and $F1_{C0}$
22. $C_k = [mF1]$, two criteria used by TOPSIS to allocate weights
23. **End For**
24. W_k: weights assigned by TOPSIS to each ensemble
25. $ind = argmax(W_k)$, denotes the index of the rankers to be eliminated
26. $M_{step} = C_{ind}$, the performance criteria of the best ensemble obtained at this iteration
27. F [:, ind] = [], eliminate the ranking list whose elimination resulted in maximum TOPSIS weight
28. **End For**
29. **Return** M_{step}, presented in Table 3

positive and negative ideal solutions. The positive ideal solution was a solution capturing the maximum benefit criteria (i.e., F1-mean) and the minimum cost criteria (i.e., minimum required number of features or feature subset size), and vice versa for the negative ideal solution. The best ensemble is the one that has the maximum weight, indicating it has the maximum normalized distance to the negative ideal solutions. TOPSIS facilitates decision-making on eliminating rankers and ultimately determining the best ensemble of rankers. Fig. 1 shows the flowchart of the proposed pipeline for ensemble feature selection.

C. CLUSTERING OF TOP-RANKED FEATURES TO DEVELOP BROADER MULTISPECTRAL BANDS

After determining the optimal ensemble and the subsequent ranking of features, algorithm 2 was developed to cluster the most informative features. The clustering algorithm was designed to form six broad spectral bands around the most prominent features to mimic the spectral bands of a typical multispectral camera (six bands were chosen because multispectral cameras rarely possess more than six spectral channels). The six most informative bands are composed of multiple nearby spectral features at the top of the ensemble feature list. The clustering algorithm was initialized with a set of six cluster centers with zero values, then it clustered the features through an iterative loop starting from the first feature at the top of the ensemble ranking list. At each iteration, one feature was assigned to a cluster if it is within 10 nm of the center of cluster, otherwise was allocated to the next empty cluster with zero center. At the end of each iteration, the cluster centers were updated. This process of clustering continued until the next feature could not be assigned to any of the six clusters. The input of algorithm 2 was the ranking of features obtained by algorithm 1, and the output was the center of broad bands to select an appropriate set of optical band-pass filters for developing a multispectral sensor.

D. PHENOTYPING THE SALT TOLERANCE OF WHEAT LINES USING SELECTED MULTISPECTRAL BANDS

The ultimate goal of clustering spectral features was to identify the most discriminative broad spectral bands for the salt stress phenotyping of wheat. To test the potential of the selected broad bands for salt stress phenotyping in wheat lines, the hyperspectral dataset for each of the four wheat lines was transformed to multispectral dataset. Each multispectral band was obtained by averaging the adjacent

Algorithm 2 Clustering of the Top Features to Develop Broad Multispectral Bands

1. Let $\mathcal{F} = \{f_i | i = 1, 2, \ldots, 215\}$ be the features ranking obtained by algorithm 1
2. Let $\mathcal{C} = \{C_j = 0 | j = 1, 2, \ldots, m_c$ be the cluster centers initialized with zero values
3. Let m_c denote the number of desired clusters, here $m_c = 6$
4. Let n_c denote the number of filled clusters, initially $n_c = 0$
5. **for** $i = 1 : 215$ **do**
6. $D_j = |f_i - C_j|$, compute the distance of $f_i \in F$ to cluster centers
7. $M_d = \min(D_j)$, find the minimum of distance
8. **if** $M_d < 10$ nm **then**
9. assign f_i to $C_j, j = \text{argmin}(D_j)$
10. **else**
11. $n_c = n_c + 1$
12. **if** $n_c <= m_c$ **then**
13. assign f_i to the first empty cluster
14. **Else**
15. break
16. **end if**
17. **end if**
18. update the cluster center $C = \left\{ C_j = \frac{1}{n} \sum_{k=1}^{n} f_k | f_k \in C_j, n = |C_j| \right\}$
19. **end for**
20. **Return** $\mathcal{C}$, containing six centers of clustered features

Algorithm 3 Ranking of Six Broad Multispectral Bands

1. Let $\chi_L = \{(x_{ij}, y_i) | i = 1, 2, \ldots, N, j = 1, 2, \ldots, m, m = 6\}$ be the multispectral dataset for the Lth wheat line
2. n_bands = 6, initial number of bands before the elimination process
3. **for** step = 1: (m − 1) **do**
4. **for** L = 1 : 4 **do**, for all four wheat lines
5. **for** b = 1:n_bands **do**
6. $\hat{\chi}_L = X_L$
7. $\hat{\chi}_L = \{(x_{ij}, y_i) | i = 1, 2, \ldots, N, j = \{1, 2, \ldots, n\_bands\}\ b\}$, eliminate the bth column from dataset
8. Train and test QDA on $\hat{\chi}_L$ with 10-fold CV
9. Record mean of error and standard deviation of error for 10-folds CV
10. $e_{Lb} = mean(error) + std(error)$, error obtained for Lth wheat line by removing bth band
11. **end for**
12. **end for**
13. $ME_b = \frac{1}{4} \sum_{L=1}^{4} e_{Lb}$, b = 1, 2, . . . , n_bands; mean e_{Lb} across four wheat lines
14. $ind = argmin(ME_b)$
15. X_L [:, ind] = [], update the dataset of all four wheat lines by removing the band at column ind
16. n_bands = n_bands – 1
17. **end for**
18. **Return** ranking of bands based on the order in which they were removed

hyperspectral bands within ±5 nm of the cluster centers calculated by algorithm 2. To identify how many of these six bands might be sufficient for salt tolerance assessment, the multispectral bands were ranked in a recursive elimination process described in algorithm 3 (similar approach to algorithm 1). In algorithm 3, a band is removed at each iteration if its elimination led to the least classification error averaged across all wheat lines (Algorithm 3). The classification error referred to the summation of the mean and standard deviation of error rate obtained for 10-fold CV. During the last iteration, Algorithm 3 saved the most discriminative band that had the lowest classification error at that iteration across all four lines. In essence, algorithm 3 ranked the six bands based on the mean of classification error across four wheat lines obtained by removing a band at each iteration.

After determining and ranking the most informative multispectral bands, then the tolerance of wheat lines to salt stress were assessed using techniques proposed by [14]. These techniques included the minimum difference of pair assignments (MDPA) and Bayesian inference, the latter of

which returned the posterior probability (i.e., P(salt|x)) of belonging to salt class given an observation x, which was defined as the similarity of pixel to the salt endmember. To compare differences between the values obtained by using multispectral and hyperspectral datasets for each of these techniques, the normalized root mean square error (NRMSE) was calculated as follows:

$$NRMSE = \frac{\sqrt{(\sum_{i=1}^{n} (\hat{y}_i - y_i)^2)/n}}{(y_{max} - y_{min})} \quad (3)$$

where y denotes the value calculated for one of the techniques (posterior probability or MDPA) using hyperspectral dataset, $\hat{y}$ refers to the value obtained for the same technique by using multispectral dataset, and n is equal to the number of wheat lines.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. BAND PAIR CORRELATION

Adjacent spectral bands tend to be correlated in hyperspectral imaging. Fig. 2 shows the correlation between all pairs of bands, which is a symmetric matrix represented as a colormap image. Among highly correlated spectral bands (∼505 to ∼645 nm), a few bands (567, 569, 589, 590, and 593 nm) had less correlation with their adjacent bands. These un-correlated bands are responsible for the two stripes in Fig. 2 denoted by the arrows. In general, bands between ∼505 to ∼645 nm were highly correlated to each other and also correlated to bands between ∼690 to ∼715nm, indicating that some of these spectral features are redundant and do not contribute valuable information.

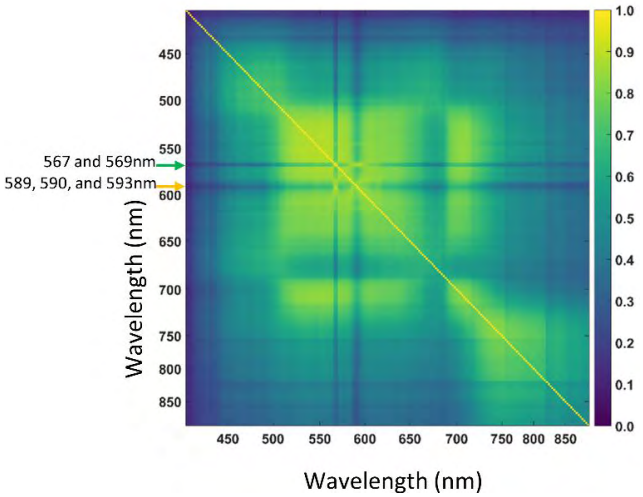

FIGURE 2. Correlation between all pairs of spectral features (i.e. 215 spectral bands from 400 nm to 900 nm) presented as a colormap image.

B. RANKING OF SPECTRAL FEATURES BY SIX RANKERS

The high correlation between features emphasized the importance of feature selection to eliminate the redundant features while retaining the relevant information. Each ranker utilized in this study ranked the spectral features based on their importance (Fig. 3). Four of the six rankers (i.e., random forest, LASSO, ReliefF, and SVM-RFE) identified 589 nm as the most informative spectral feature among all 215 bands. Intriguingly, 589 nm is among the Fraunhofer lines (i.e. spectral absorption lines of elements) and associated with sodium (*Na*). It is a promising result considering sodium chloride (NaCl) was the applied salt treatment to the wheat plants. This result suggests that these four rankers were able to capture the fundamental physical meaning concealed in the spectral response of wheat plants treated with NaCl. It is remarkable that the hyperspectral images were captured only one day after the salt treatment was applied. SFS-forward ranked 589 nm as the third most important feature and ranked 751 nm and 532 nm as the first two important spectral features. CFS ranked 589 nm as the 210th most informative feature and 583 nm as the first most informative feature despite the fairly close correlation between these two bands (i.e., 0.77). Except SFS-forward, all rankers had their first two dominant features between 581 nm and 589 nm, indicating the importance of this region of the electromagnetic spectrum for determining salt stress in wheat.

The other inference from Fig. 3 is that the spectral features in blue (around 450 nm) and red (650 nm) regions, known as the chlorophyll absorption bands, had the least importance (i.e., largest ranking value) among all the spectral features.

Feature subset size and F1-mean were computed for each ranker to compare the rankers with each other and further compare them with the results obtained by the ensemble of rankers (Table 2). The QDA classifier required a different number of features from the feature set ranked by each ranker to attain a CV error with 1SE of the minimum error. The lowest CV error on the training dataset and the maximum F1-mean on the test dataset were both achieved when QDA used the feature set ranked by SFS-forward. However, this performance was attained at the cost of larger feature size (28 features) and higher standard deviation of CV error. Alternatively, the lowest feature subset size and standard deviation of error was achieved using random forest. The F1-mean for random forest dropped only 1.63% compared to that of SFS-forward, while the feature subset size for random forest (12 features) was significantly less than SFS-forward (28 features). Among all rankers, CFS resulted in the lowest classification accuracy as indicated by the lowest $F1_{C0}$ and $F1_{C1}$ values.

C. ENSEMBLE OF FEATURE SELECTION TECHNIQUES

The ranking assigned to the features by the six rankers were aggregated together (Fig. 4 (a)) using (Fig. 1) (Fig. 1-Step III). As expected from the ranking results of the base rankers, spectral features near the 589 nm band were dominant at the top of the ranking list of features.

From the ensemble ranking list, a subset of features was selected by QDA based on the minimum CV error and 1SE rule (Fig. 4 (b)). Fig. 4 (b) illustrates the mean and standard deviation of the 10-fold CV error for all 215 feature

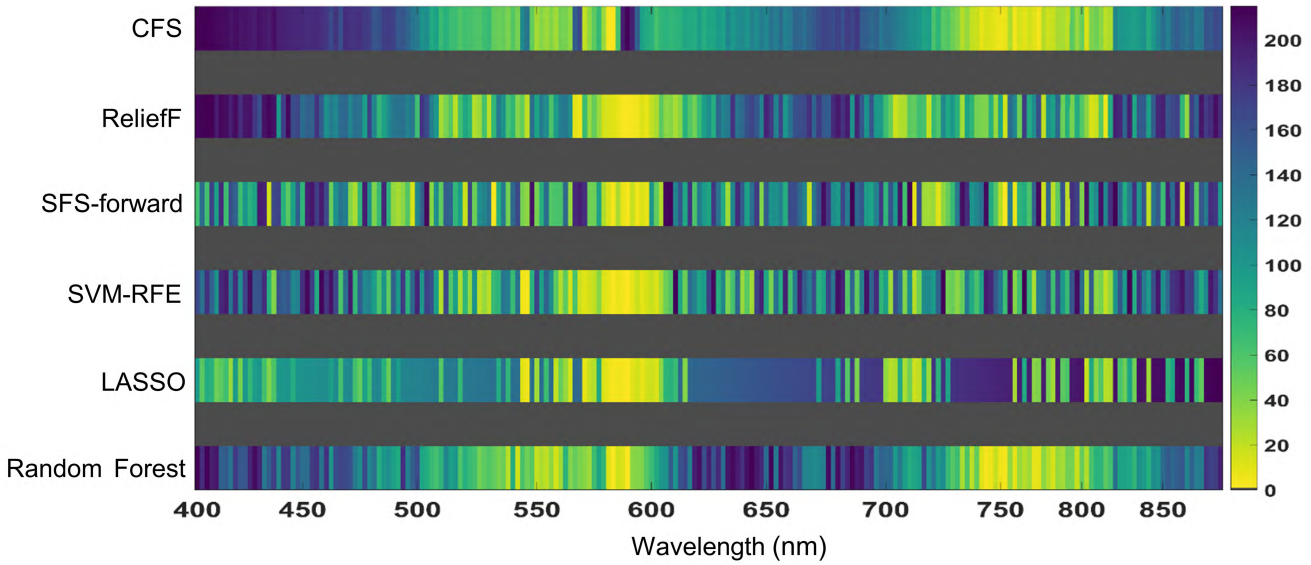

FIGURE 3. Ranking of the 215 spectral features for each of the six rankers. Top-ranked spectral features are represented with lighter color and low-ranked features are represented with darker color. Dark blue refers to the least important features with large ranking values and light yellow refers to the most important features with low ranking values. Gray color (zero value in colorbar) was used to separate the results of rankers.

TABLE 2. Classification performance of QDA on the training and test datasets for individual rankers (for training dataset: mean and standard deviation of cross validation error and feature subset size obtained by the one standard error rule; for test dataset: F1 measure of control (C0) and salt (C1) classes obtained by training QDA on the training dataset using only features in feature subset and testing on unseen dataset).

	Training dataset with 10-fold CV			Test dataset		
	Feature subset size	Error (%)	StD (%)	$F1_{C0}$ (%)	$F1_{C1}$ (%)	F1-mean (%)
CFS	21	27.33	1.30	71.74	71.76	71.75
ReliefF	19	21.57	1.47	78.41	75.85	77.13
SFS-forward	28	20.36	1.78	78.87	76.55	77.71
SVM-RFE	19	21.86	1.55	78.27	74.91	76.59
LASSO	18	21.39	1.68	78.23	75.20	76.71
Random Forest	16	22.58	1.29	77.29	74.87	76.08

subsets, in which the leftmost subset contains only the top ranked feature and the rightmost subset contains all features (the model complexity increases across the horizontal axis). CV error decreased with an increasing size of feature subset until feature subset contained 23 features whereby CV error began to increase, which underscore the need for feature selection. Because of the uncertainty around CV error of each feature subset (shown in Fig. 4 (b)), it was appropriate to determine the most parsimonious feature subset within one standard error of the minimum CV error for determining the optimum number of features. QDA classifier needed the first 18 features at top of the feature list ranked by the ensemble of all rankers to achieve an error rate within 1SE of the minimum CV error. To achieve a similar CV error obtained by using all the 215 features (CV error $=$ 0.30), QDA required the top seven features from the feature list.

D. RECURSIVE ELIMINATION OF RANKERS

A base ranker was disregarded from the ensemble process if its elimination resulted in an improved classification performance among all options at a given step. The F1-mean and number of features in the subset were used by TOPSIS to determine which rankers should be eliminated from the ensemble process at each step (results are presented in Table 3). The highest F1-mean was achieved at step 1 using 19 features obtained by the ensemble of all rankers except CFS. However, elimination of LASSO and SFS-forward at the next two iterations resulted in four less features while only reducing F1-mean by less than 0.5% compared to step 1 that offered the largest F1-mean. Although the results obtained at step 2 and 3 were similar, the ensemble of three rankers at step 3 is preferred due to less complexity and less computation compared to the ensemble of four rankers at step 2. Therefore, the final ensemble feature subset was obtained by aggregating the ranking of ReliefF, SVM-RFE, and random forest, each of which belongs to a different feature selection category: filter, wrapper, and embedded, respectively.

The size of feature subset obtained by the ensemble of these three rankers (Table 3; Fig. 5) dropped to 15 features, three features less than the ensemble of all six rankers.

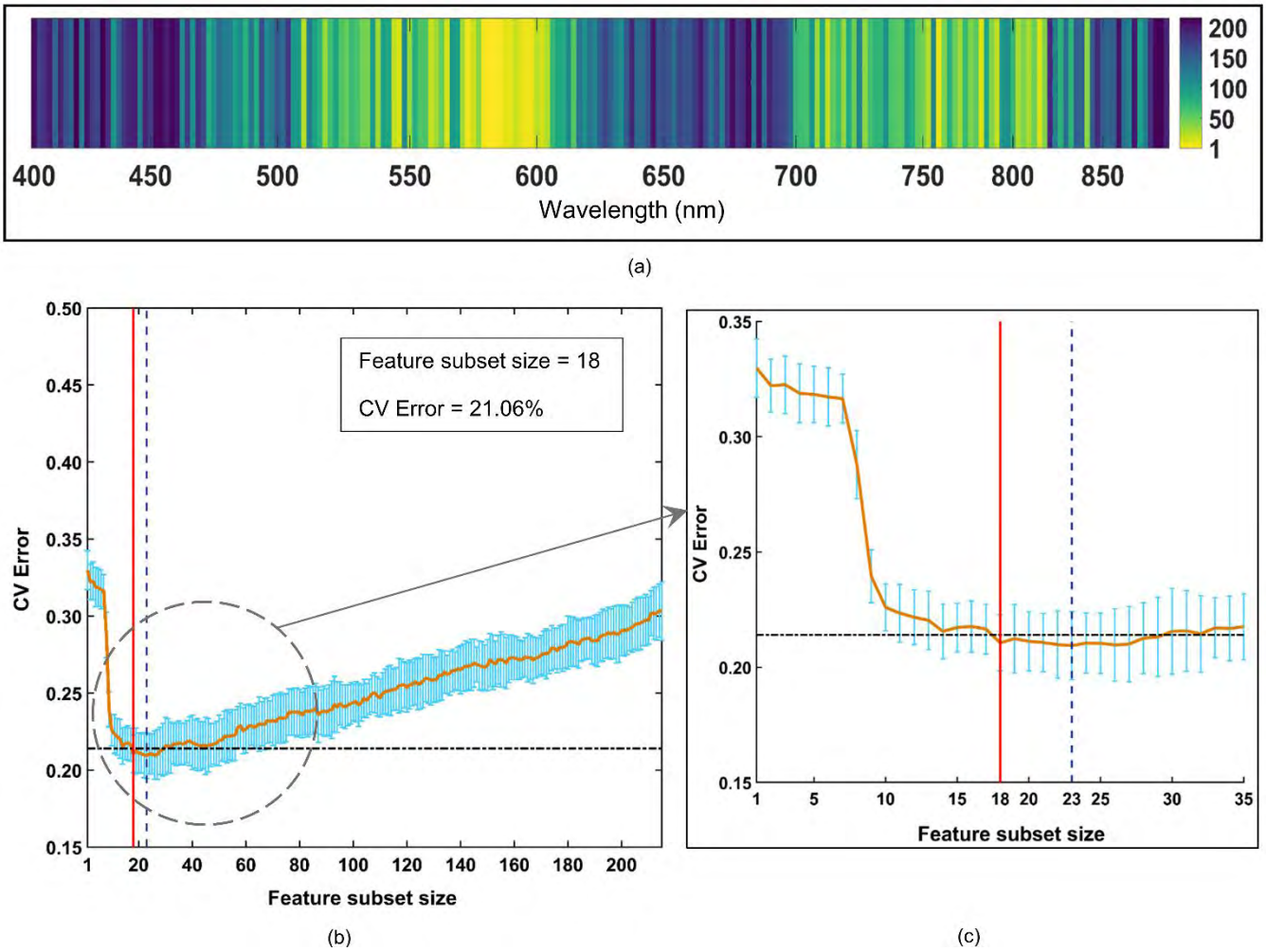

FIGURE 4. (a) Ranking of features obtained by the ensemble of all six rankers: correlation-based feature selection, ReliefF, sequential feature selection - forward selection, support vector machine-recursive feature elimination, and random forest. (b) Cross validation error (CV error) of QDA on the unseen fold of the 10-fold cross validation for each feature subset size. The feature subset size required to obtain the minimum CV error is shown by the vertical dashed line, the lowest error plus one standard error is presented by the horizontal dashed line, and the feature subset size of the most parsimonious model whose error was within 1SE of the minimum error is illustrated by the vertical solid line. (c) CV error for the first few feature subset size to show with more detail how the most parsimonious feature subset size is selected.

TABLE 3. Feature subset size and F1-mean after each step of the recursive elimination of rankers (Algorithm 3).

Recursive step	Removed technique	Feature subset size	F1-mean (%)
Full dimensional hyperspectral dataset	-	215	68.49
Step 0- Ensemble 6 techniques	-	18	77.32
Step 1- Ensemble 5 techniques	CFS	19	77.47
Step 2- Ensemble 4 techniques	LASSO	15	77.09
Step 3- Ensemble 3 techniques	SFS-forward	15	77.07
Step 4- Ensemble 2 techniques	ReliefF	17	76.89
Step 5- 1 technique	SVM-RFE	16	76.08

Furthermore, the difference between classification accuracy of the ensemble of three rankers (F1-mean = 77.07) and the maximum accuracy of individual rankers achieved by SFS-forward (F1-mean = 77.71) was less than 1%, while the model complexity of SFS-forward (28 features) was approximately twice as that of the three rankers ensemble (15 features). This demonstrates the benefit of using ensemble feature selection.

Fig. 5 depicts the feature ranking achieved by the ensemble of ReliefF, SVM-RFE, and random forest, and Fig. 5 (b)

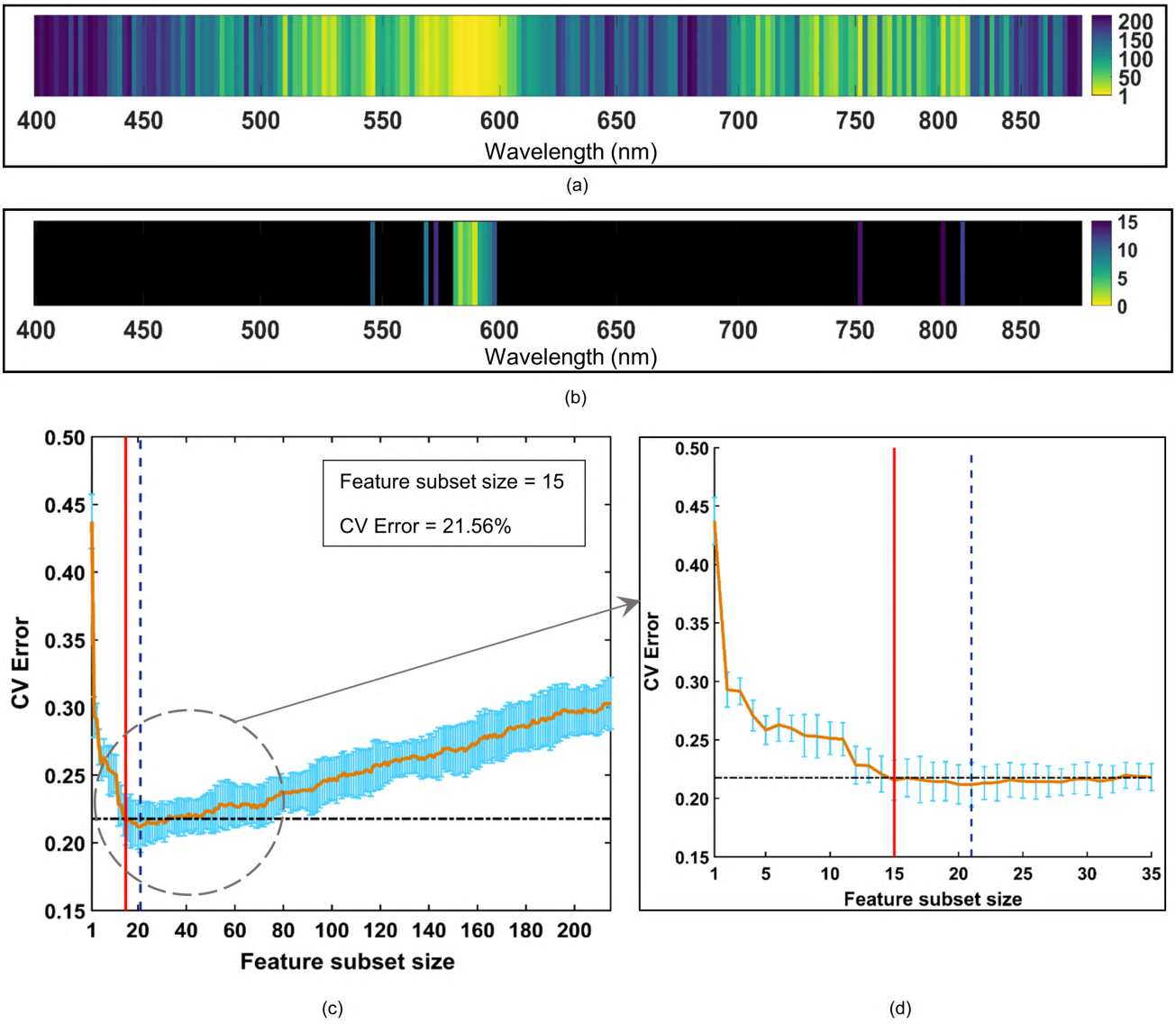

FIGURE 5. (a) Ranking of features obtained by the ensemble of selected rankers: ReliefF, SVM-RFE, and random forest. (b) The 15 features selected among the 215 spectral features scanned from 400 nm to 900 nm. (c) Cross validation error (CV error) of QDA on the unseen fold of the 10-fold cross validation for each feature subset size. The feature subset size required to obtain the minimum CV error is shown by the vertical dashed line, the lowest error plus one standard error is presented by the horizontal dashed line, and the feature subset size of the most parsimonious model whose error was within 1SE of the minimum error is illustrated by the vertical solid line. (d) CV error for the first few feature subset size to show with more detail how the most parsimonious feature subset size is selected.

demonstrates the ranking and distribution of the 15 selected features over the scanned electromagnetic wavelengths. The band nearest to the region of sodium absorption (589 nm) and the adjacent bands were among the most informative and were consequently the top-ranked features.

Similar to the ensemble of all six rankers, the ensemble of the top three rankers resulted in a decreasing CV error of QDA to a point (21 features), followed by an increasing CV error as feature subset size increased. However, the CV error obtained by the ensemble of three rankers had a sharper decrease within the first few feature subsets compared to the ensemble of six rankers. The two top-ranked features (589 nm and 583 nm) from the ensemble of three rankers were able to achieve a similar CV error obtained by using all 215 features (CV error = 0.30), whereas the top seven features from the ensemble of six rankers were required to achieve a comparable CV error.

E. CLUSTERING OF TOP-RANKED FEATURES

The ensemble feature list derived from the top three techniques was utilized in algorithm 2 to find the first six clusters of spectral features whose mean values were at least 10 nm apart. The first 18 features were needed to form the six

TABLE 4. Clustering of features ranked by the ensemble of three rankers: ReliefF, SVM-RFE, and random forest. Algorithm 1 was used for clustering (only integer part of wavelengths are presented; superscripts denote feature rank).

	Cluster center	Members of each cluster								
Cluster 1	589	589^{1}	583^{2}	587^{3}	585^{4}	581^{5}	591^{6}	593^{7}	595^{8}	597^{11}
Cluster 2	573	569^{9}	573^{13}	579^{17}						
Cluster 3	546	546^{10}								
Cluster 4	805	809^{12}	801^{15}							
Cluster 5	751	751^{14}								
Cluster 6	528	525^{16}	530^{18}							

clusters (Table 4). Cluster 1 had the largest size (nine members), whereas cluster 3 and 5 had the smallest size (one member).

F. IMPORTANCE OF MULTISPECTRAL BANDS

The spectral bands within ±5 nm of the six cluster centers (Table 4) from the hyperspectral dataset were averaged to imitate a multispectral dataset for each of the four wheat lines (Table 1). Algorithm 3 ranked these six bands via a recursive elimination process as

$$\mathcal{R} = \{\mathbf{528}, \mathbf{805}, \mathbf{589}, \mathbf{573}, \mathbf{751}, \mathbf{546}\}$$

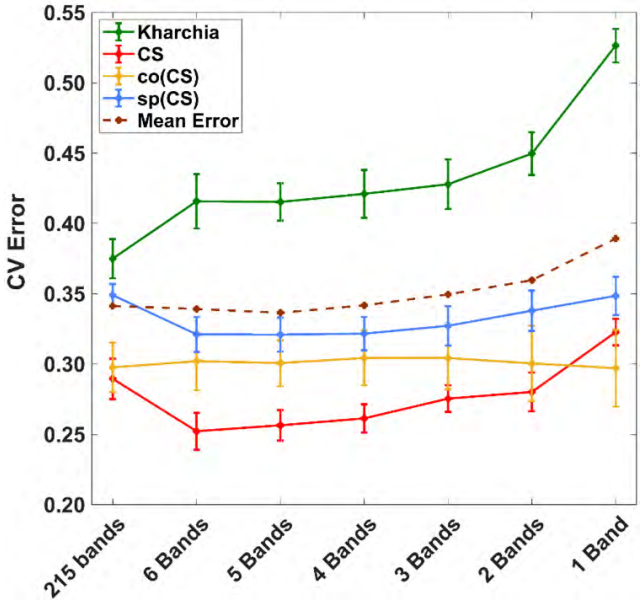

FIGURE 6. Mean of error across four wheat lines (dash line) and CV error of four wheat lines for hyperspectral ($n = 215$) and multispectral datasets using 10-fold cross validation. Bands were selected from the $\mathcal{R}$ set in order (i.e., *5 Bands* contains the first five bands in $\mathcal{R}$ and *1 Band* contains only the first band in $\mathcal{R}$).

where 528 nm is the center of the most informative band that persisted after the last iteration and 546 nm is the center of the least important band eliminated during the first iteration. Among all wheat lines, Kharchia had the largest CV error across the hyperspectral and all multispectral datasets (Fig. 6). This result can be explained by the fact that Kharchia is the most salt-tolerant line among the four lines [14], indicating the spectral response of plants under the salt treatment were less likely to be affected by the imposed stress. Consequently, the classification algorithm failed to accurately discriminate between pixels that represented plants under each of the control and salt treatments for Kharchia. Alternatively, the lowest CV error was found for CS, which was the most susceptible line among the others. The lowest CV error for CS was achieved by using the six multispectral broad bands and it increased as the number of bands decreased.

G. SALT STRESS PHENOTYPING OF WHEAT LINES USING MULTISPECTRAL BANDS

The performance of multispectral bands in ranking wheat lines based on their salt tolerance was determined using methods used by [14] (i.e., MDPA and posterior probability), and results were compared with those derived from the hyperspectral dataset (Fig. 7). Using hyperspectral dataset and MDPA and Bayesian approaches, the ranking of salt tolerance for the evaluated wheat lines (from most tolerant to most susceptible) is: 1) Kharchia, 2) co(CS), 3) sp(CS), and 4) CS, which is consistent with the results of conventional phenotyping and historical evidence [14], [44], [45]. The results presented here show that the same ranking of wheat lines is obtained using six multispectral bands developed in this study (Fig. 7). Bands were selected from the $\mathcal{R}$ set, which contains the bands in the order of their importance.

For both MDPA and posterior probability, lower values indicate a higher tolerance to salt stress. Use of only the top two multispectral bands in $\mathcal{R}$ set (i.e., 528 nm and 805 nm) resulted in the same ranking of wheat lines as was achieved using more multispectral bands or the complete hyperspectral dataset. However, using fewer bands may lead to more uncertainty because as the number of bands decreases, the values deviate more from the base values obtained by using the hyperspectral bands (Fig. 7). To quantify the difference between the values achieved by using various multispectral band combinations with base values obtained by all bands, NRMSE was computed for both techniques (IV-H). NRMSE could also be used to compare the performance of MDPA and Bayesian inference in ranking wheat lines as the number of multispectral bands decreased, since NRMSE calculated

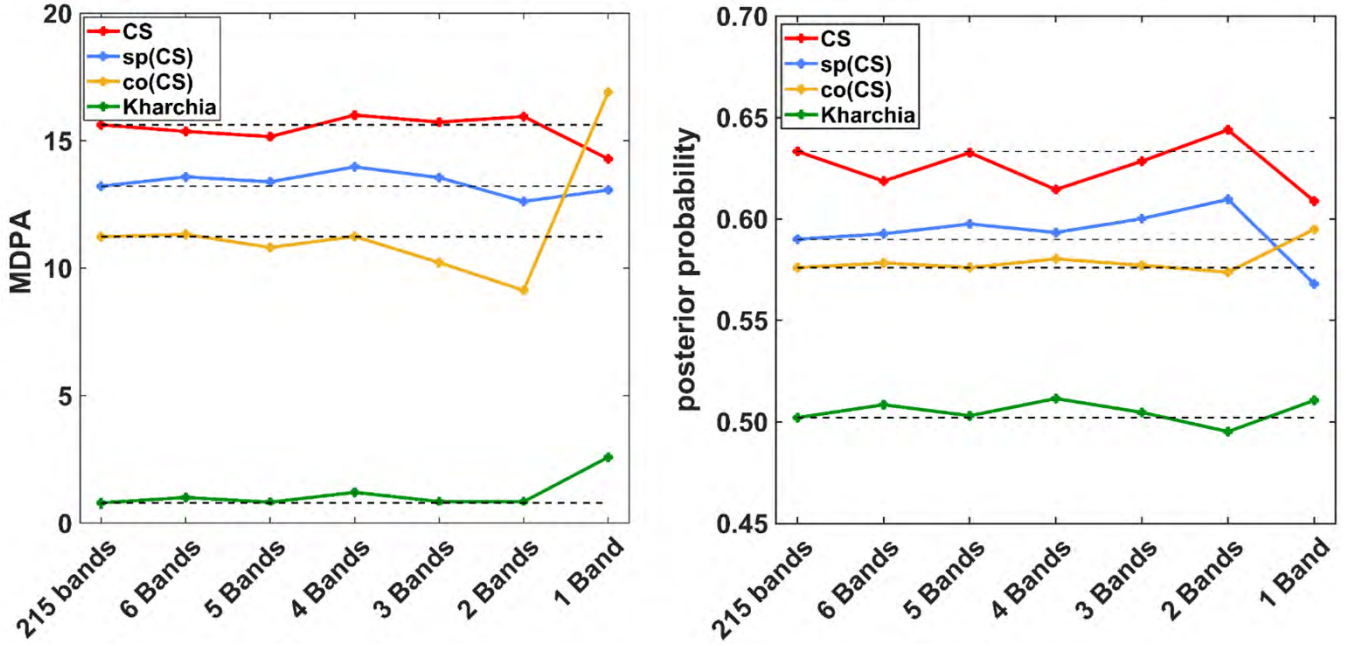

FIGURE 7. Performance comparison between the selected multispectral bands and the full hyperspectral dataset (reported by [14]) for ranking wheat lines based on their salt stress tolerance using (a) Minimum difference of pair assignments (MDPA) and (b) the Posterior probability P(salt|x). For both MDPA and posterior probability, lower values indicate a higher tolerance to salt stress.

TABLE 5. Normalized root mean square error (NRMSE) of minimum difference of pair assignments (MDPA) and Bayesian inference approach calculated for the selected six multispectral bands.

	NRMSE (%)					
	6 Bands	5 Bands	4 Bands	3 Bands	2 Bands	1 Band
MDPA	1.71	2.20	3.18	3.63	7.46	20.55
Bayesian inference	6.23	2.90	8.29	4.42	8.94	14.83

for both techniques has the same scale due to normalization based on the range of base values for each technique.

By decreasing the number of multispectral bands, NRMSE fluctuated for the Bayesian inference approach, whereas it consistently increased for the MDPA approach. By using only the first three bands in $\mathcal{R}$ (528, 805, and 589 nm), NRMSE for MDPA was less than 4%, indicating a multispectral camera with three broad bands centered at these bands would be sufficient for salt stress phenotyping.

H. DISCUSSION

The present study proposed a pipeline to ensemble multiple rankers to rank spectral features of a high-dimensional hyperspectral dataset. The proposed pipeline offered a novel ensemble of base rankers to aggregate the benefits associated with the base rankers while disregarding those whose presence reduced the performance of classification. The algorithm was able to determine the inclusion/exclusion of a ranker for the final ensemble based on two criteria: feature subset size and F1-mean.

The ensemble feature selection drastically reduced the dimension of the hyperspectral dataset (from 215 to 15 spectral features). While 200 features were eliminated, the F1-mean improved by 8.5% from 68.49% to 77.09% using only 15 features of the feature ranking list obtained by the ReliefF, SVM-RFE, and random forest rankers. This reveals that a small subset of spectral features are able to capture a large portion of the most valuable information, while the remaining features tend to be redundant and/or contribute to noise. Moreover, ranking 589 nm, a known *Na* absorption band, as the top-ranked feature indicated that the proposed ensemble method was able to discover a physical meaning concealed in the high-dimensional spectral data of NaCl-treated wheat plants under salt stress.

In addition to the ensemble feature selection pipeline, the other contribution of this study was finding the centers of broad multispectral bands and ordering them based on their importance in discriminating salt stress leaves from healthy ones. This discovery can be a foundation for developing a custom-designed multispectral camera that aims to leverage the advantages of a multispectral versus hyperspectral sensing (e.g., simplicity, reducing cost, avoiding complications associated with line scanning hyperspectral cameras, etc.). Furthermore, by using the multispectral bands identified in this study, the salt stress tolerance of four wheat lines was assessed in a more interpretable and efficient manner. The ranking of wheat lines using the multispectral dataset was

consistent with the analysis of the hyperspectral dataset evaluated by [14].

In this study, we initiated the idea of establishing a library for spectral features correlated to the plant diseases and stresses, each of which adversely affecting the plant physiology. Consequently, the spectral response of plant would change based on the type of disease and stress. This underscores the importance of hyperspectral imaging as a promising research tool to find meaningful wavelengths for each type of applications at the first place. Once important features are identified, then multispectral sensors can be designed accordingly for that specific application. In general, the benefits of identifying appropriate multispectral bands to reduce the dimension of a hyperspectral dataset can be summarized as (i) reducing the complexity of interpretation as complexity is a function of input data dimensions, (ii) reducing the risk of overfitting by selecting a less complex model, (iii) reducing the running time, computational cost, and storage required for processing and storing large hyperspectral datasets, (iv) reducing the required number of samples/pixels, and (v) reducing the cost of data collection (e.g. a multispectral sensor versus a hyperspectral sensor). Furthermore, a simpler model with less redundant/irrelevant features is more robust.

V. CONCLUSION

The proposed ensemble feature selection pipeline ranked the spectral features of a hyperspectral dataset from the most to the least important based on the accuracy of a classifier in distinguishing between healthy and stressed vegetation pixels. The top 15 features in the feature ranking list achieved the best classification results, and the top two features were able to achieve a similar CV error as that of all 215 spectral features. The findings of this study suggest that *feature selection* should be embedded as one of the main pre-processing steps in hyperspectral image analysis. Feature selection mitigates the challenges in post-processing (e.g., complex interpretation and expensive computational cost) while improving classification accuracy. We were able to determine six multispectral bands that can substantially contribute in the development of a custom-designed multispectral camera for wheat salt stress phenotyping. A multispectral camera designed for a specific application can benefit scientists, farmers, agricultural practitioners, and field consultants across a wide range of use cases. Some benefits include avoiding: (i) the high cost of hyperspectral sensors, (ii) difficulties with data collection, especially as it relates to properly using a line-scanning camera, and (iii) the required expertise for analyzing high-dimensional hyperspectral images. The proposed feature selection pipeline can be utilized for feature selection and, subsequently, dimensionality reduction of other types of high-dimensional datasets (e.g., microarray datasets). In future work, we will employ this feature selection pipeline to select informative spectral features in phenotyping for other abiotic/biotic stress tolerance in plants, and work to extend the library of spectral features associated with various plant diseases and stresses.

REFERENCES

[1] M. S. M. Asaari *et al.*, "Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 121–138, Apr. 2018.

[2] S. Thomas *et al.*, "Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform," *Plant Methods*, vol. 14, p. 45, Jun. 2018.

[3] T. Schmid *et al.*, "Characterization of soil erosion indicators using hyperspectral data from a mediterranean rainfed cultivated region," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 845–860, Feb. 2016.

[4] S. Veraverbeke *et al.*, "Hyperspectral remote sensing of fire: State-of-the-art and future perspectives," *Remote Sens. Environ.*, vol. 216, pp. 105–121, Oct. 2018.

[5] E. Bedini and T. M. Rasmussen, "Use of airborne hyperspectral and gamma-ray spectroscopy data for mineral exploration at the Sarfartoq carbonatite complex, southern West Greenland," *Geosci. J.*, vol. 22, no. 4, pp. 641–651, 2018.

[6] Z. Aslett, J. V. Taranik, and D. N. Riley, "Mapping rock forming minerals at boundary Canyon, death valey national park, California, using aerial SEBASS thermal infrared hyperspectral image data," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 64, pp. 326–339, Feb. 2018.

[7] X. Feng *et al.*, "Nondestructive and rapid determination of lignocellulose components of biofuel pellet using online hyperspectral imaging system," *Biotechnol. Biofuels*, vol. 11, p. 88, Apr. 2018.

[8] A. Moghimi, C. Yang, M. E. Miller, S. Kianian, and P. Marchetto, "Hyperspectral imaging to identify salt-tolerant wheat lines," *Proc. SPIE*, vol. 10218, May 2017. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10218/1021805/Hyperspectral-imaging-to-identify-salt-tolerant-wheat-lines/10.1117/12.2262388.full?SSO=1&tab=ArticleLink

[9] Z. Liu, H. Wang, and Q. Li, "Tongue tumor detection in medical hyperspectral images," *Sensors*, vol. 12, no. 1, pp. 162–174, 2012.

[10] R. J. Beaulieu, S. D. Goldstein, J. Singh, B. Safar, A. Banerjee, and N. Ahuja, "Automated diagnosis of colon cancer using hyperspectral sensing," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 3, p. e1897, 2018.

[11] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, p. 010901, 2014.

[12] F. Masia, I. Pope, P. Watson, W. Langbein, and P. Borri, "Bessel-beam hyperspectral CARS microscopy with sparse sampling: Enabling high-content high-throughput label-free quantitative chemical imaging," *Anal. Chem.*, vol. 90, no. 6, pp. 3775–3785, 2018.

[13] N. Pavurala, X. Xu, and Y. S. R. Krishnaiah, "Hyperspectral imaging using near infrared spectroscopy to monitor coat thickness uniformity in the manufacture of a transdermal drug delivery system," *Int. J. Pharmaceutics*, vol. 523, no. 1, pp. 281–290, 2017.

[14] A. Moghimi, C. Yang, M. E. Miller, S. F. Kianian, and P. M. Marchetto, "A novel approach to assess salt stress tolerance in wheat using hyperspectral imaging," *Frontiers Plant Sci.*, vol. 9, p. 1182, Aug. 2018.

[15] M. Wahabzada, A.-K. Mahlein, C. Bauckhage, U. Steiner, E.-C. Oerke, and K. Kersting, "Plant phenotyping using probabilistic topic models: Uncovering the hyperspectral language of plants," *Sci. Rep.*, vol. 6, Mar. 2016, Art. no. 22482.

[16] A.-K. Mahlein, E.-C. Oerke, U. Steiner, and H.-W. Dehne, "Recent advances in sensing plant diseases for precision crop protection," *Eur. J. Plant Pathol.*, vol. 133, no. 1, pp. 197–209, 2012.

[17] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[18] J. Tang, S. Alelyani, and H. Liu, *Feature Selection for Classification: A Review*. Boca Raton, FL, USA: CRC Press, 2014.

[19] Y. Saeys, T. Abeel, and Y. van de Peer, "Robust feature selection using ensemble feature selection techniques," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2008, pp. 313–325.

[20] D. Chutia, D. K. Bhattacharyya, J. Sarma, and P. N. L. Raju, "An effective ensemble classification framework using random forests and a correlation based feature selection technique," *Trans. GIS*, vol. 21, no. 6, pp. 1165–1178, 2017.

[21] K. Koonsanit, C. Jaruskulchai, and A. Eiumnoh, "Band selection for dimension reduction in hyper spectral image using integrated information gain and principal components analysis technique," *Int. J. Mach. Learn. Comput.*, vol. 2, no. 3, pp. 248–251, Jun. 2012.
[22] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.
[23] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2015, pp. 1200–1205.
[24] A. Rady, N. Ekramirad, A. A. Adedeji, M. Li, and R. Alimardani, "Hyperspectral imaging for detection of codling moth infestation in GoldRush apples," *Postharvest Biol. Technol.*, vol. 129, pp. 37–44, Jul. 2017.
[25] K. Nagasubramanian, S. Jones, S. Sarkar, A. K. Singh, A. Singh, and B. Ganapathysubramanian. (2017). "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean." [Online]. Available: https://arxiv.org/abs/1710.04681
[26] K. R. Thorp, G. Wang, K. F. Bronson, M. Badaruddin, and J. Mon, "Hyperspectral data mining to identify relevant canopy spectral features for estimating durum wheat growth, nitrogen status, and grain yield," *Comput. Electron. Agricult.*, vol. 136, pp. 1–12, Apr. 2017.
[27] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, Oct. 2007.
[28] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, May 2015, Art. no. 198363.
[29] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.
[30] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.
[31] B. B. Damodaran, N. Courty, and S. Lefèvre, "Sparse Hilbert Schmidt independence criterion and surrogate-kernel-based feature selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2385–2398, Apr. 2017.
[32] A. Moghimi and C. Yang, "Hyperspectral image dataset for salt stress phenotyping of wheat," Univ. Minnesota, Minnesota, MN, USA, Tech. Rep., 2018, doi: 10.13020/D69Q3K.
[33] M. A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 359–366.
[34] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. 19th Int. Workshop Mach. Learn.*, 1992, pp. 249–256.
[35] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Appl. Intell.*, vol. 7, no. 1, pp. 39–55, Jan. 1997.
[36] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.
[37] R. Tibshirani, "Regression selection and shrinkage via the lasso," *J. Roy. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[39] A. Ben Brahim and M. Limam, "Ensemble feature selection for high dimensional data: A new method and a comparative study," in *Advances in Data Analysis and Classification*. Berlin, Germany: Springer, 2017, pp. 1–16, doi: 10.1007/s11634-017-0285-y.
[40] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," *Bioinformatics*, vol. 26, no. 3, pp. 392–398, 2010.
[41] L. I. Kuncheva, "Ensemble feature selection," in *Combining Pattern Classifiers: Methods and Algorithms*, 2nd ed. Hoboken, NJ, USA: Wiley, 2014, pp. 290–325.
[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer-Verlag, 2009.
[43] C.-L. Hwang and K. Yoon, *Methods for Multiple Attribute Decision Making*. Berlin, Germany: Springer-Verlag, 1981.
[44] Y. Zhang *et al.*, "Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (Triticum aestivum L.)," *Sci. Rep.*, vol. 6, 2016, Art. no. 21476.
[45] R. Munns, R. A. James, and A. Läuchli, "Approaches to increasing the salt tolerance of wheat and other cereals," *J. Exp. Botany*, vol. 57, no. 5, pp. 1025–1043, 2006.

ALI MOGHIMI received the B.S. degree in agricultural machinery engineering from the Bahonar University of Kerman, Iran, in 2004, the M.S. degree in biosystems engineering from the Ferdowsi University of Mashhad, Iran, in 2008, and the Ph.D. degree in computer science from the College of Science and Engineering, University of Minnesota, Twin Cities, MN, USA, where he is currently pursuing the Ph.D. degree in biosystems engineering.

His research interests include remote sensing applications for precision agriculture, hyperspectral and multispectral imaging, computer vision, and machine learning. He has been a member of the American Society of Agricultural and Biological Engineering since 2015.

Mr. Moghimi's awards include the MnDRIVE Global Food Ventures Fellowship in 2017, the BBE Graduate Fellowship in 2018, and the Runner-up for the Best Paper Award from the SPIE Conference on Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping in 2017.

CE YANG received the B.S. degree in electrical engineering and the M.S. degree in agricultural engineering from China Agricultural University in 2007 and 2009, respectively, and the M.S. degrees in agricultural engineering and computer science from the University of Florida in 2013 and 2014, respectively. Since 2014, she has been an Assistant Professor with the University of Minnesota, Twin Cities, MN, USA.

Her research interests span remote sensing, deep learning, and machine learning to solve agricultural problems, such as plant stress detection, crop yield detection, and nitrogen management using advanced tools and techniques.

Dr. Yang has been a member of the American Society of Agricultural and Biological Engineering for eight years. From 2009 to 2014, she received the Graduate School Fellowship that supported her Ph.D. research.

PETER M. MARCHETTO (M'13) was born in Teaneck, NJ, USA, in 1984. He received the B.S. degree in physics from the Ramapo College of New Jersey (RCNJ) and the M.S. and Ph.D. degrees in biological and environmental engineering from Cornell University, with a focus on bioinstrumentation. His background is primarily in physics and its applications to sensing systems.

His previous relevant experience begins with biomedical instrumentation work at the New Jersey Institute of Technology, magnetic materials characterization at RCNJ, calibration work at Calyx, Inc., materials work on piezoelectric polymers at PennState, and a staff position as an Engineering Physicist for the Bioacoustics Research Program at Cornell University, followed by a post-doctoral appointment at the Cornell Soil and Water Lab. He is currently an Assistant Professor with the Department of Bioproducts and Biosystems Engineering, University of Minnesota, where he is also the Co-Director of the Instrumentation and Robotics Group.

Dr. Marchetto is a member of the American Geophysical Union, the American Physical Society, the Acoustical Society of America, the American Society for Engineering Education, and Sigma Xi.