Ecological Informatics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Species classification from hyperspectral leaf information using machine learning approaches

Guangman Song, Quan Wang *

Faculty of Agriculture, Shizuoka University, Shizuoka 422-8529, Japan

ARTICLE INFO

Keywords:
Hyperspectral reflectance
Machine learning
Species classification
Wavelength selection

ABSTRACT

Information on plant species is fundamental to forest ecosystems, in the context of biodiversity monitoring and forest management. Traditional methods for plant species inventories are generally inefficient, in terms of cost and performance, and there is a high demand for a quick and feasible approach to be developed. Of the various attempts, remote sensing has emerged as an active approach for plant species classification, but most studies have concentrated on image processing and only a few of them ever use hyperspectral information, despite the wealth of information it contains. In this study, plant species are classified from hyperspectral leaf information using different machine learning models, coupled with feature reduction and selection methods, and their performance is optimized through Bayesian optimization. The results show that including feature selection and Bayesian optimization increases the classification accuracy of machine learning models. Among these, the Bayesian optimization-based support vector machine (SVM) model, combined with the recursive feature elimination (RFE) feature selection method, yields the best output, with an overall accuracy of 86% and a kappa coefficient of 0.85. Furthermore, the confusion matrix revealed that the number of samples correlates with classification accuracy. The support vector machine with informative bands after Bayesian optimization outperformed in classing plant species. The results of this study facilitate a better understanding of spectral (phenotype) information with plant species (genotype) and help to bridge hyperspectral information with ecosystem functions.

1. Introduction

Information about plant species is fundamental to our knowledge of forest biodiversity and ecosystems and is an inevitable requirement for natural ecosystem conservation and sustainable management (Dalponte et al., 2012; Santos et al., 2019). The ability to identify and classify plant species is, therefore, essential for the automatic mapping of vegetation composition, distribution, and forest dynamics. A traditional plant species survey approach, however, involves a lot of laborious, costly, and time-consuming fieldwork, which is commonly localized and can rarely be extended to cover large scales (Cho et al., 2009; Ribeiro da Luz, 2006). As a result, effective and accurate techniques for comprehensively classifying plant species are in high demand.

Among various attempts, the remote sensing technique offers a practical approach for classifying plant species, especially at large scales (Alonzo et al., 2014; Liu et al., 2021; Mäyrä et al., 2021). Several types of remotely sensed data have been reported in previous studies for plant species classification, ranging from multispectral, hyperspectral, light detection and ranging to their combinations (Aviña-Hernández et al., 2023; Fassnacht et al., 2016; Heinzel and Koch, 2012; Zhang et al., 2020b). Among the diverse remote sensing information, hyperspectral leaf reflectance contains a number of continuous narrow spectral bands for the finer discrimination of spectral properties and should have the potential to classify plant species (Cavender-Bares et al., 2016; Hennessy et al., 2020; Omeer and Deshmukh, 2021; Prospere et al., 2014). Differences between the spectra relate closely to differences in biochemical composition, pigments, and water content, as well as structures (Curran, 1989; Nakaji et al., 2019), contributing to the ability to distinguish plant species. What is more, studies have demonstrated that spectral diversity can be used as a surrogate for functional or phylogenetic diversity (Frye et al., 2021; Schweiger et al., 2018), which provides a solid foundation for plant species classification using spectroscopy. So far, leaf spectral differences in different species have been confirmed in previous studies (Castro-Esau et al., 2006; Jin et al., 2020), implying the possibility of classifying species based on leaf hyperspectral signatures. In truth, several previous studies have already demonstrated that there is a great

* Corresponding author.
E-mail address: wang.quan@shizuoka.ac.jp (Q. Wang).

https://doi.org/10.1016/j.ecoinf.2023.102141
Received 13 February 2023; Received in revised form 24 April 2023; Accepted 21 May 2023
Available online 24 May 2023

potential for spectrally discriminating plant species (Hycza et al., 2018; Ullah et al., 2021).

Classifying plant species from hyperspectral information relies on a number of classification approaches, including both parametric and non-parametric models. The most commonly and continuously used method is discriminant analysis, e.g. linear discriminant analysis (LDA) (Fisher, 1936), which is a parametric classifier. There is an assumption of a normal distribution and it yields good results for species classification in various ecosystems, particularly in tropical forests (Clark et al., 2005; Féret and Asner, 2011). However, the use of non-parametric classifiers, such as the k-nearest neighbor algorithm (KNN), random forest (RF), and support vector machine (SVM), has gradually increased to classify plant species in recent years (Khan et al., 2022; Maxwell et al., 2018; Omeer and Deshmukh, 2021). These non-parametric classifiers are effective for classifying complex and high-dimensional data and have achieved good classification accuracies in previously reported studies (Ferreira et al., 2016; Grabska et al., 2020). To the best of our knowledge, however, few studies have ever conducted a direct comparison of different approaches, with most previous reports focused only on each specific classification method. As a result, the potential generalizability of classification models has yet to be fully evaluated.

On the other hand, it is well known that hyperspectral data has the disadvantage of high data dimensionality and multicollinearity and thus contains redundant information, resulting in overfitting and decreasing the performance of the classification models (Fassnacht et al., 2014; Zhang et al., 2020a). Consequently, using feature reduction and selection algorithms to select optimal latent factors for summarizing and the optimal subset from the original hyperspectral data, such as principal component analysis and recursive feature elimination (Demarchi et al., 2020; Kalacska et al., 2007), is a critical step to addressing such shortcomings. Nevertheless, limited knowledge of how feature selection impacts the accuracy of classification methods greatly confined our understanding. Appropriate feature selection strategies for deriving critical hyperspectral bands for accurate species classification should be examined.

Importantly, machine learning models are highly sensitive to the choice and values of the hyperparameters involved; ways of optimizing these hyperparameters are considered to be one of the most critical steps in machine learning models (Agrawal, 2021; Bischl et al., 2023). Commonly applied hyperparameter optimization techniques, such as grid search and random search, typically involve discretizing the parameter space and implementing an iterative search procedure to approximate the optimal hyperparameter (Bergstra and Bengio, 2012; Yang and Shami, 2020). Recently, the Bayesian optimization (BO) algorithm has gradually captured the attention of many fields and emerged as an efficient method for hyperparameter tuning because of its superiority to conventional methods (Snoek et al., 2012; Yang and Shami, 2020). This method views the hyperparameter tuning process as the optimization of a black-box function and determines the next hyperparameter value based on the previously-obtained results, avoiding numerous unnecessary evaluations (Eggensperger et al., 2013; Malu et al., 2021). Therefore, it may be worthwhile trying to use the algorithm to optimize hyperparameters for machine learning models on plant species classification from hyperspectral leaf information.

Furthermore, most previously reported studies on plant species classification only explored specific species grown in limited environments. Very few studies have ever been conducted to encompass a wide range of species belonging to different plant functional types, growing in varied environments or ecosystems. As reported in previous studies, large variations between and within the species could influence the performance of classification models (Clark and Roberts, 2012; Hesketh and Sánchez-Azofeifa, 2012), hence a comprehensive representation of plant species is necessary. In addition, no consensus has yet been reached on which classification models should be applied in discriminating plant species.

This study, therefore, aims to investigate different classification methods to classify plant species using hyperspectral leaf information, using a relatively comprehensive dataset compiled from several publicly available datasets covering different species from different locations. Specific objectives include: 1) examine the feasibility of classifying plant species using machine learning models; 2) assess the impact of feature selection algorithms and hyperparameter optimization on classification accuracy; and 3) evaluate how the classification accuracy for specific species is sensitive to sample size.

Table 1
Descriptions of the datasets used in this study.

Dataset	Location	Date	Spectroradiometer	Spectral range (nm)
ANGERS	Angers, France	2003	ASD FieldSpec	400–2450
LOPEX	Ispra	1993	Perkin Elmer Lambda 19	400–2500
FAB	Cedar Creek LTER, USA	2018	psr + 3500	350–2500
NEON	Eastern USA	2017	ASD FieldSpec/psr+	350–2500
BHI	Blackhawk Island, WI	2018	psr + 3500	350–2500
NAKA	Nakakawane, Japan	2014–2019, 2021	ASD FieldSpec	350–2500

2. Materials and methods

2.1. Data used in this study

The data used in this study were compiled from a number of independent datasets covering various species and plant functional types from different locations, each of which contained hyperspectral leaf reflectance. In detail, the publicly available datasets included those from: ANGERS (National Institute for Agricultural Research) (Jacquemound et al., 2003), LOPEX (Joint Research Centre) (Hosgood et al., 1993), FAB (Cedar Creek LTER, East Bethel, MN, USA) (Kothari et al., 2018), NEON (University of Wisconsin Environmental Spectroscopy Laboratory) (Wang, 2017a, 2017b), and BHI (Blackhawk Island) (Chlus and Townsend, 2018). In addition, another dataset from Nakakawane Forest in Japan (NAKA) (138°06′E, 35°04′N) was also used in this study, in which samples of *Acer nipponicum*, *Acer shirasawanum*, *Betula grossa*, *Fagus crenata*, *Pterostyrax hispidus*, and *Stewartia pseudocamellia* (collected from 2014 to 2019 and then 2021) were considered. The detailed descriptions of these datasets are presented in Table 1.

For each dataset, spectral outliers were excluded by using principal component space with a ratio statistic (Dangal et al., 2019) before they were combined together. Since under-representative data samples have difficulty meeting the requirement for separation into training and test data sets in the machine learning classification methods, those species with <50 spectral samples were also excluded, finally resulting in a total of 52 species (a total of 8340 spectral samples) for further analysis (Table 2). Furthermore, as the reflectance spectra of these datasets were measured using either the ASD Field Spec (Analytical Spectral Devices, Boulder, CO), Spectral Evolution PSR+, or Perkin–Elmer Lambda 19 spectroradiometer, which cover different wavelength ranges, the spectral range was uniformed to the domain from 400 to 2400 nm in this study. The distributions of leaf spectra in each combined dataset are illustrated in Fig. 1.

2.2. Data analysis

The flowchart describing the steps for classifying plant species based on spectral leaf information and using different machine learning models, is illustrated in Fig. 2. Hyperspectral data within the domain of 400 to 2400 nm were smoothed using the Savitzky-Golay method first. Then, four different feature reduction and selection methods were

Table 2
Plant species list and the number of samples for each species used in this study.

No	Species	Species Code	Sample Number	No	Species	Species Code	Sample Number
1	*Acalypha neomexicana*	ACANEO	72	27	*Magnolia acuminata*	MAGACU	64
2	*Acacia saligna*	ACASAL	147	28	*Phacelia austromontana*	PHAAUS	100
3	*Acer nipponicum*	ACENIP	52	29	*Platanus occidentalis*	PLAOCC	156
4	*Acer pseudoplatanus*	ACEPSE	191	30	*Poa tracyi*	POATRA	72
5	*Acer rubrum*	ACERUB	709	31	*Polycarpon depressum*	POLDEP	90
6	*Acer saccharinum*	ACESAC	627	32	*Populus grandidentata*	POPGRA	160
7	*Acer shirasawanum*	ACESHI	218	33	*Prenanthes serpentaria*	PRESER	60
8	*Ailanthus altissima*	AILALT	58	34	*Pterostyrax hispidus*	PTEHIS	56
9	*Andropogon gerardii*	ANDGER	181	35	*Quercus ×macnabiana*	QUE × MA	90
10	*Besseya alpina*	BESALP	78	36	*Quercus ×palaeolithicola*	QUE × PA	58
11	*Betula alleghaniensis*	BETALL	80	37	*Quercus alba*	QUEALB	665
12	*Betula grossa*	BETGRO	88	38	*Quercus falcata*	QUEFAL	80
13	*Betula nigra*	BETNIG	151	39	*Quercus mohriana*	QUEMOH	86
14	*Betula papyrifera*	BETPAP	120	40	*Quercus nigra*	QUENIG	86
15	*Calliandra conferta*	CALCON	64	41	*Quercus phellos*	QUEPHE	56
16	*Calochortus tolmiei*	CALTOL	134	42	*Quercus rubra*	QUERUB	772
17	*Carya cordiformis*	CARCOR	214	43	*Quercus velutina*	QUEVEL	86
18	*Celtis occidentalis*	CELOCC	87	44	*Quercus virginiana*	QUEVIR	77
19	*Digitaria villosa*	DIGVIL	66	45	*Robinia pseudoacacia*	ROBPSE	74
20	*Fagus crenata*	FAGCRE	320	46	*Sophora nuttalliana*	SOPNUT	68
21	*Fagus grandifolia*	FAGGRA	78	47	*Spartina pectinata*	SPAPEC	81
22	*Fraxinus nigra*	FRANIG	66	48	*Stewartia pseudocamellia*	STEPSE	97
23	*Fraxinus pennsylvanica*	FRAPEN	105	49	*Tilia americana*	TILAME	694
24	*Juglans nigra*	JUGNIG	108	50	*Trichomanes davallioides*	TRIDAV	68
25	*Linum striatum*	LINSTR	122	51	*Typha latifolia*	TYPLAT	68
26	*Liriodendron tulipifera*	LIRTUL	128	52	*Ulmus americana*	ULMAME	112

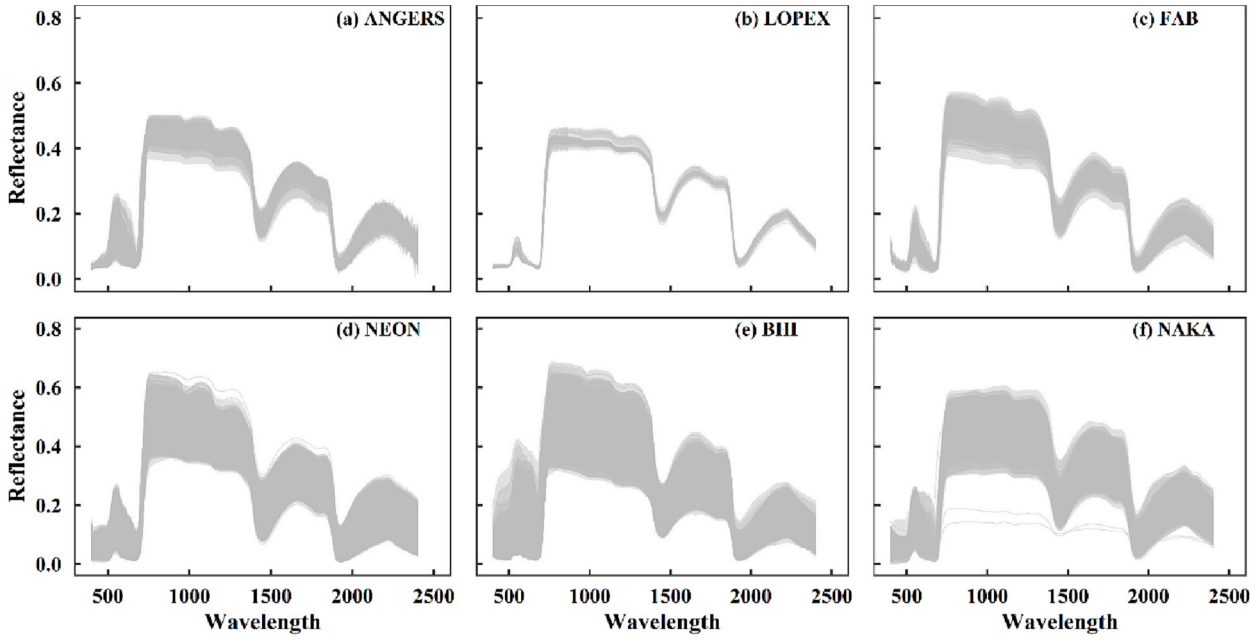

Fig. 1. Hyperspectral leaf reflectance of different datasets.

employed to summarize and extract potential informative spectral bands for later plant species classification using different classification models, using full bands or the selected informative spectral bands. In the meantime, the Bayesian optimization algorithm was incorporated into the machine learning models.

2.2.1. Feature reduction and selection methods

Several popularly applied feature reduction and selection methods were used in this study, including principal components analysis (PCA), recursive feature elimination (RFE), least absolute shrinkage and selection operator (LASSO), and a genetic algorithm for the identification of a robust subset (GARS). Their brief summaries are presented in Table 3. These methods have been successfully used together with many classification algorithms to build robust classification models (Chiesa et al., 2020; Demarchi et al., 2020; Kalacska et al., 2007). In addition, the available functions of each method are also shown in Table 3. As for this study, the scree plot of variance, explained by each component in PCA, and the cumulative spectral variance were examined to determine the number of principal components (PCs) in subsequent analyses. The correlations of selected PCs with each wavelength were then calculated to evaluate the spectral information they represented. The number of features in the RFE and GARS methods were retained in the range of 20 to 200. The criteria in the LASSO to select variables was based on the penalty factor (λ). Ten-fold validation was applied in all RFE, LASSO, and GARS selection methods.

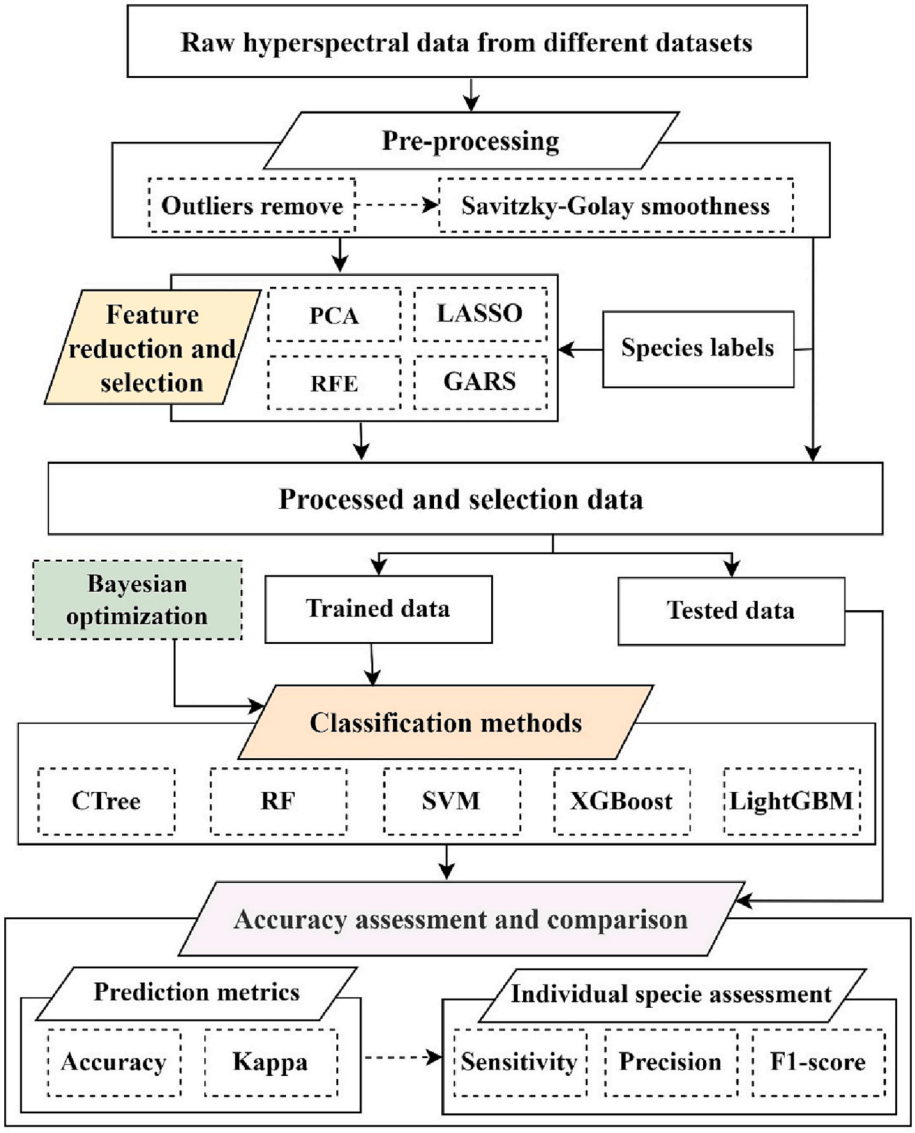

Fig. 2. Flowchart of the plant species classification procedure in this study.

Table 3
Feature extraction and selection methods, as well as the functions that implement these methods in plant species classification in this study.

Selection method	Type	Description	Package: Function	Parameter
Principal components analysis (PCA)	Feature extraction	Transform the data into a set of orthogonal components and maximize the variance to summarize the spectral information (Lever et al., 2017).	stats: prcomp	–
Recursive feature elimination (RFE)	Feature selection	Aim at estimating the features that are most helpful for the classes of interest and recursively eliminate the unimportant features (Guyon et al., 2002).	caret: rfe	feature_range (20–200)
Least absolute shrinkage and selection operator (LASSO)	Feature selection	A popular high-dimensional data analysis method that compresses non-relevant variables to be exactly zero (Tibshirani, 1996).	glmnet: glmnet	lambda (λ, 0.001–0.1)
Genetic algorithm for the identification of a robust subset (GARS)	Feature selection	An innovative implementation of genetic algorithms for fast and accurate identification of informative features in multi-class and high-dimensional datasets (Chiesa et al., 2020).	Bioconductor: GARS_GA	feature_range (20–200)

2.2.2. Classification approaches

In this study, the composited data were separated into training (80%) and testing (20%) datasets following a stratified sampling method based on species to maintain the same class distribution as the original dataset. This splitting was conducted using the 'createDataPartition' function in the caret package (Kuhn, 2008). Well-known types of machine learning classification approaches were considered in this study, such as conditional inference trees (CTree), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), and gradient boosting (XGBoost: extreme gradient boosting; LightGBM: light gradient boosting machine). These classification methods were extensively documented in previous studies classifying plant species and have demonstrated moderate to good performance (Georganos et al., 2018; Sabat-Tomala et al., 2020; Venkatasubramaniam et al., 2017). The description, implemented packages, and functions for these classification methods are shown in Table 4. In this study, the Bayesian optimization (BO) method, as

Table 4
Description of different classification methods for plant species classification and the packages and functions used to implement these methods in this study.

Classification method	Description	Package: Function	Parameter
Conditional inference trees (CTree)	It determines the variable to split on based on a measure of association between each covariate and target, and then calculates the best split point for that variable (Hothorn et al., 2006).	partykit: ctree	minsplit (10–40), minbucket (5–15), maxsurrogate (0–5)
Random forest (RF)	It is a congregation of decision trees that are created with subsets which are selected on a random basis with replacement (Breiman, 2001).	randomForest: randomForest	mtry (1-(ncol(feature)-1), ntree (200−1000), nodesize (1-(sqrt(nrow (traindata)))
Support vector machine (SVM)	It aims to obtain optimal hyperplanes through the selection of the points that have the largest gaps between different classes (Cortes and Vapnik, 1995).	caret: svm	cost (0.01–100), gamma (0.001–0.1), kernel (radial)
K-nearest neighbor (KNN)	It searches k samples that are nearest to the point to be classified based on the distance metrics between different data points (Steinbach and Tan, 2009).	class: knn	n_neighbors (3–50)
Extreme gradient boosting (XGBoost)	It is a scalable machine learning method for tree boosting that performs classification based on obtained feature importance (Chen and Guestrin, 2016).	xgboost: xgb.train	eta (0.1–1), max_depth (4–6), subsample (0.1–1), bytree (0.4–1), nrounds (50–200)
Light gradient boosting machine (LightGBM)	It is a fast and efficient gradient boosting method to perform classification using tree-based learning algorithms (Ke et al., 2017).	lightgbm: lgb.train	learning_rate (0.01–1), max_depth (4–6), bytree (0.4–1), subsample (0.1–1), min_data (20), nrounds (50–200)

implemented in the rBayesianOptimization package, was employed to obtain the optimal values of these hyperparameters (Table 4).

2.3. Classification performance evaluation

All analyses were performed in the R environment on a Windows 10 desktop with 32 GB of RAM, and an NVIDIA GeForce RTX 3060 Ti graphics card with 8 GB of RAM. The performance of machine learning models on plant species classification was evaluated using overall accuracy and kappa coefficient (Cohen, 1960), by generating confusion matrices (Theissler et al., 2022), which provide a comparison between actual and predicted labels of species and have been a very commonly used measure for solving classification problems. The correspondence of specific species was assessed in terms of sensitivity, precision, and F1-score in the confusion matrices. Sensitivity and precision are the fractions of correctly predicted actual positive classes and predictive positive classes, respectively. In addition, the higher F1-score (the harmonic mean of sensitivity and precision) showed a better performance. The specific formulas for calculating these criteria are given below:

$$Sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (3)$$

where TP, FN, and FP are true positive, false negative, and false positive, respectively, based on the comparison of the actual and predicted species labels.

3. Results

3.1. Informative hyperspectral bands identified by different feature reduction and selection methods

PCA analysis of spectral leaf data revealed that the first three principal components (PCs) explained 94% of the cumulative variance, according to the screen plot of variance. Correlations between the determined PCs and each wavelength are presented in Fig. 3a, clearly showing that each PC carries information from specific wavelengths. In detail, the strongest correlations between the first PC and wavelengths were found in the shortwave infrared (SWIR) regions with a spectral domain longer than 1600 nm. The second and third PCs had the strongest correlations, with the near-infrared (NIR) domain (800–1100 nm) and the visible (VIS) domain (500–700 nm), respectively.

The informative hyperspectral bands selected for plant species classification using RFE, LASSO, and GARS selection methods are illustrated in Fig. 3b. In detail, 72, 57, and 48 informative bands were selected by RFE, LASSO, and GARS, respectively. Obviously, much fewer bands were involved in the GARS selection method and the selected bands were spread over the whole spectral regions from 400 to 2400 nm. In comparison, the LASSO selection method selected the bands in specific spectral regions. Commonly, spectral bands selected by the three selection methods most include the visible region of the spectrum (400–700 nm); furthermore, the wavelengths around 1900 nm were consistently selected, irrespective of the selection method.

3.2. Classification models with or without feature reduction and selections

The performance (overall accuracy and kappa coefficient) of each machine learning classification model (including CTree, KNN, RF, SVM, XGBoost, and LightGBM) on plant species classification coupling with or without feature reduction and selections was depicted in Fig. 4 and Table 5. The Bayesian optimization (BO) algorithm was used for hyperparameter optimization.

In detail, the SVM model achieved the highest mean accuracy (84%) and kappa coefficient (0.83) among all the classification models, when it was without coupling with any feature reduction or selection methods. This was followed by the LightGBM and XGBoost models, which had the mean accuracy and kappa coefficient of 74% and 0.71, and 72% and 0.68, respectively. The CTree model performed worst of all and only yielded a mean accuracy of 41% and a mean kappa coefficient of 0.38.

The accuracy and kappa coefficient of the classification models with the fewer informative features selected by PCA and LASSO were lower than the performance of the models without feature reduction and selection methods. However, the performance of classification models based on the selected features of RFE and GARS was comparable and even better in comparison to using all features. The SVM model also outperformed the other models, based on the features selected by PCA, RFE, LASSO, and GARS methods, with the accuracy and kappa coefficient ranging from 61% to 86% and 0.55 to 0.85, respectively. More specifically, after hyperparameter optimization, the SVM model combined with the RFE selection method obtained an accuracy and kappa coefficient of 86% and 0.85, respectively, which is higher than the model combined with the GARS (82% and 0.81), LASSO (63% and 0.61), and PCA (61% and 0.55) methods. Furthermore, the other classification models combined with the RFE band selection method exhibited a higher accuracy and kappa coefficient than the PCA, GARS, and LASSO selection methods. On the other hand, however, the GARS involved much fewer spectral bands (48) than when based on the RFE selection method (72). Overall, the SVM model had the highest accuracy and

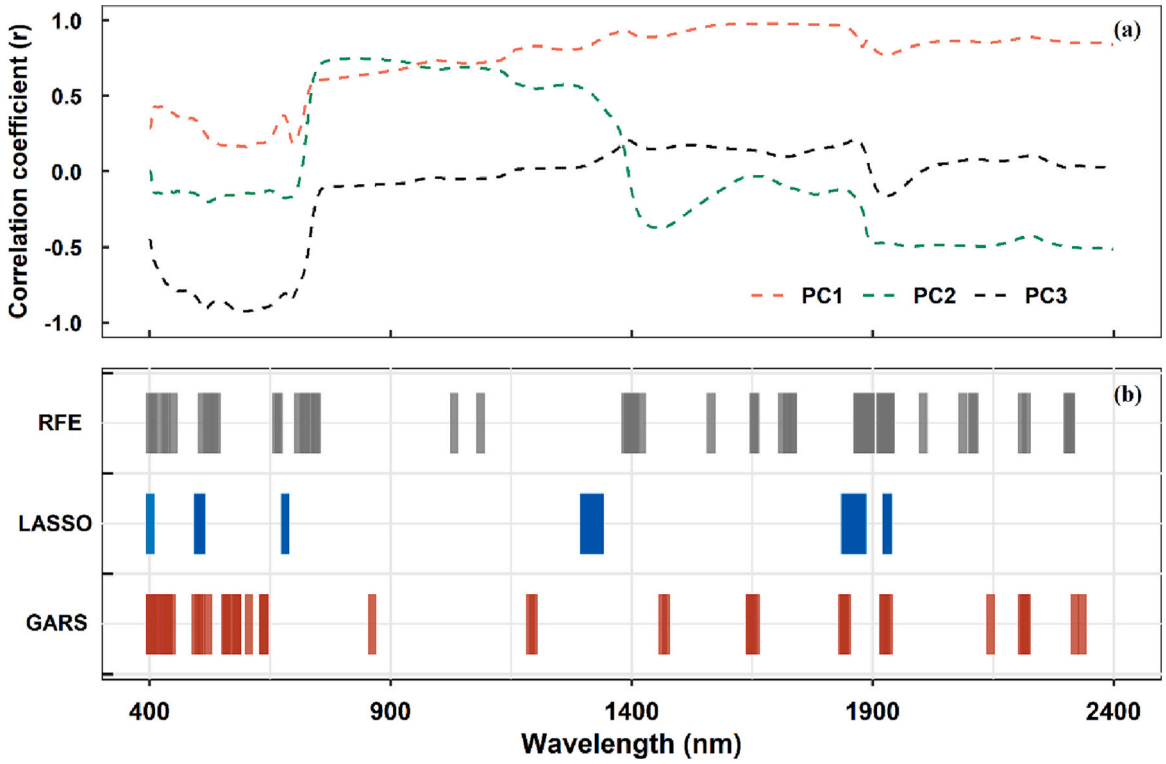

Fig. 3. Correlation coefficients between the three principal components (PC1: orange, PC2: green, and PC3: black) and each wavelength (a) and the distribution of the bands selected by different band-selection methods (RFE: grey, LASSO: blue, and GARS: red) (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

kappa coefficient among the classification models which underwent the Bayesian optimization for hyperparameters in all cases.

3.3. Confusion matrix for individual species classification

The confusion matrix generated from the best model, namely the SVM model in combination with RFE after Bayesian optimization, was further evaluated in terms of sensitivity, precision, and F1-score (Fig. 5). The model very accurately classified ACEPSE (*Acer pseudoplatanus*), CARCOR (*Carya cordiformis*), FAGCRE (*Fagus crenata*), PHAAUS (*Phacelia austromontana*), QUERUB (*Quercus rubra*), and POPGRA (*Populus grandidentata*) with high sensitivity, precision, and F1 (all >0.95). The ACERUB (*Acer rubrum*), ACESAC (*Acer saccharinum*), ACESHI (*Acer shirasawanum*), ANDGER (*Andropogon gerardii*), QUEALB (*Quercus alba*), SPAPEC (*Spartina pectinate*), and TILAME (*Tilia americana*) were next best, with sensitivity, precision, and F1 >0.90. The model was less accurate in the prediction of PRESER (*Prenanthes serpentaria*) with sensitivity, precision, and F1 <0.50.

Furthermore, the sensitivity, precision, and F1-score were found to be non-linearly related to the number of samples (Fig. 6). Sensitivity, precision, and F1-score increased with the number of samples until they reached about 150 and then they remained relatively constant. Specifically, the sensitivity, precision, and F1-score were all higher than 0.82 in each species with samples >150, mostly within the range of 0.85 to 0.98. The sensitivity, precision, and F1-score in the species with <150 samples mostly ranged from 0.50 to 0.85, except for a few individual species.

4. Discussion

4.1. Classification approaches

In this study, different non-parametric classification models were examined, with regard to their potential to classify plant species from hyperspectral leaf information. The results show that hyperspectral leaf reflectance has the ability to accurately classify plant species, as has been reported in previous studies (Clark et al., 2005; Omeer and Deshmukh, 2021; Prospere et al., 2014). It has been reported that spectral leaf reflectance varies considerably due to a variety of factors, such as seasonality, canopy variations, and environmental variability (Asner et al., 2014; Hesketh and Sánchez-Azofeifa, 2012; Jin et al., 2020), and this can restrict the separability of plant species. In our study, the species were collected from different datasets in various environments, with seasonal and canopy variations contained in some individual datasets. Several robust machine learning classification methods were applied to the plant species classification in this study and obtained moderate to good results in this study, suggesting a strong potential for classifying plant species using machine learning models based on leaf-level hyperspectral information.

Of all the machine learning classification models available, the best accuracy and kappa coefficient for plant species classification was achieved by the SVM model after Bayesian optimization, regardless of whether feature reduction was used or which selection method was used. The results are consistent with previous studies which reported that the SVM model obtained superior performance in plant species classification across the remote sensing fields, owing to the advantage of handling high-dimensional feature spaces (Cervantes et al., 2020; Mountrakis et al., 2011). For example, Grabska et al. (2020) demonstrated that the SVM model outperformed the random forest and extreme gradient boosting for forest species mapping using a combination of Sentinel-2 imagery and environmental data. Furthermore, the SVM model based on informative bands, selected by RFE selection methods after the Bayesian optimization, obtained the highest accuracy (86%) and kappa coefficient (0.85) throughout this study. The predictive ability, as shown in this study, is higher than those reported in previous studies (Cavender-Bares et al., 2016; Hesketh and Sánchez-

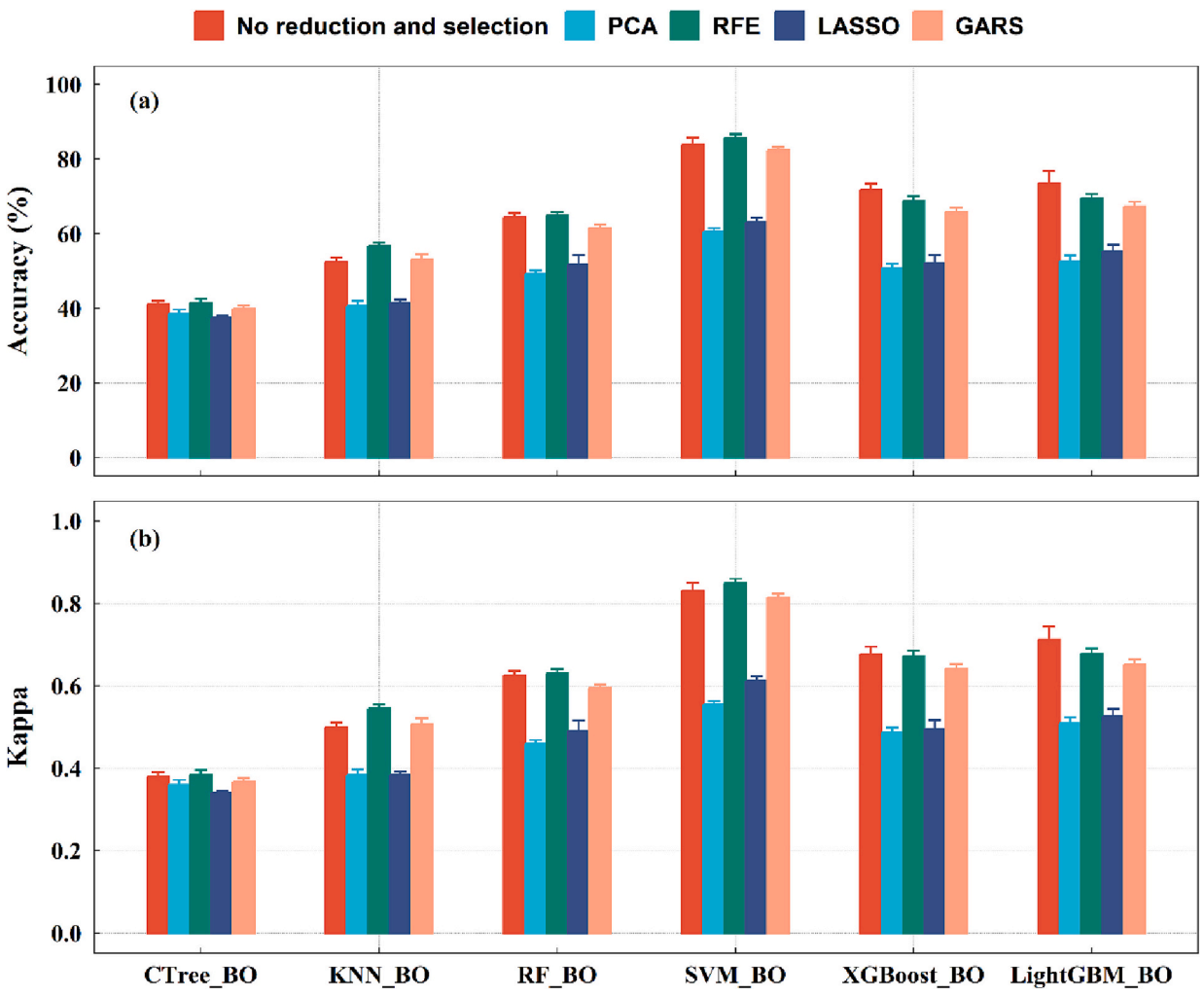

Fig. 4. Accuracy and kappa coefficient of the different classification methods with and without feature reduction and selection methods after Bayesian optimization (BO) for the plant species classification.

Azofeifa, 2012; Prospere et al., 2014), albeit the dataset used in this study contains the species with different environmental, seasonal, and canopy variations. The performance was similar to the accuracy of 11 tree species classification (85.6%) using the combination of SVM and RFE from spectral imagery (Grabska et al., 2020), but was superior to the most popular multi-layer perception neural network (MLP) (Rawat and Wang, 2017; Sumsion et al., 2019), which had the accuracy and kappa coefficient of 0.81 and 0.80, respectively, when the hidden layer consisted of 32 neurons. The results clearly revealed the effectiveness of feature selection and hyperparameter optimization in machine learning models for plant species classification.

4.2. *Importance of feature selections for plant species classifications*

Hyperspectral data contains narrow bands that are correlated with one another and commonly involve feature reduction and selection algorithms to isolate the most important bands when performing plant species classification. In this study, with the RFE feature selection, the classification of these plant species with the Bayesian optimization-based SVM model resulted in an overall accuracy of 86% and a kappa coefficient of 0.85, which were relatively higher than other feature reduction and selection methods, as well as full bands. In comparison to full bands, the RFE selection method uses much fewer spectral bands. On the other hand, the PCA, LASSO, and GARS methods reduced the number of spectral bands during classification, while losing part of the inherent spectral information, resulting in lower classification performance. The results are in line with previous studies, which reported the effectiveness of the feature selection method in the machine learning models for species classification based on hyperspectral signatures (Demarchi et al., 2020; Hennessy et al., 2020).

Within the feature selection methods, the relatively unique spectral information was further identified for plant species classification. The spectral regions selected by the RFE selection method were mostly concentrated in visible light (400–700 nm) and shortwave infrared (SWIR). The results are consistent with previous studies, highlighting the classification ability and relatively unique spectral information contributed by pigment absorption related to the xanthophyll cycle and chlorophyll (Alonzo et al., 2014). The bands around 1900 nm were selected consistently in the feature selection methods, which are related to water absorption and largely influence leaf reflectance (Das et al., 2021; Skoneczny et al., 2020). These two spectral regions are also closely related to the phylogenetic structure and signatures, and thus contribute to species classification (Diniz et al., 2021; McManus et al., 2016). Nevertheless, fewer near-infrared spectral regions were selected in this study, in contrast to the study by Clark et al. (2005), which found that near-infrared (NIR: 700–1327 nm) bands were important spectral regions for tropical tree species discrimination. The results are in accordance with the study by Dalponte et al. (2012), which found that the NIR region was inappropriate for tree species classification, based on the fusion of spectral images and LiDAR data, due to high within-class

Table 5
Accuracy and kappa coefficient for species classification of each machine learning approach without and with feature reduction and selection methods under Bayesian optimization.

	Feature reduction and selection	Accuracy (%)	Kappa
	Full band	41	0.38
	PCA	39	0.36
CTree	RFE	41	0.38
	LASSO	38	0.34
	GARS	40	0.37
	Full band	52	0.50
	PCA	41	0.38
KNN	RFE	57	0.55
	LASSO	41	0.38
	GARS	53	0.51
	Full band	64	0.62
	PCA	49	0.46
RF	RFE	65	0.63
	LASSO	52	0.49
	GARS	62	0.59
	Full band	84	0.83
	PCA	61	0.55
SVM	RFE	**86**	**0.85**
	LASSO	63	0.61
	GARS	82	0.81
	Full band	72	0.68
	PCA	51	0.49
XGBoost	RFE	69	0.67
	LASSO	52	0.50
	GARS	66	0.64
	Full band	74	0.71
	PCA	53	0.51
LightGBM	RFE	70	0.68
	LASSO	55	0.53
	GARS	67	0.65

diversity. The selection of particular spectral domains improved classification accuracy and computation efficiency, compared to the retention of whole hyperspectral wavelengths, thereby better classifying plant species, especially in large datasets.

4.3. Impact of Bayesian optimization on classification accuracy

This study aimed to seek a robust classification model that continuously delivers high classification performance when classifying plant species in datasets with large variations. Thereby, the Bayesian optimization method was applied to determine the optimal hyperparameters for these machine learning models. So far, the Bayesian optimization method has captured a lot of attention as an efficient tool for hyperparameter tuning (Agrawal, 2021; Malu et al., 2021; Yang and Shami, 2020) and it is essential for models involving black-box functions and very little prior information. Accordingly, the analysis revealed that a combination of the Bayesian optimization yielded an overall accuracy range of 41–84% when the classification models were built without considering feature reduction and selection methods. The ranges of accuracy in different classification models ranged from 38 to 86% using the Bayesian optimization method including feature reduction and selection. This result is in line with the previous studies in various fields (Sameen et al., 2020; Wang et al., 2021), which showed superior performance in the use of Bayesian optimization in machine learning models for classification problems.

We also explored the performance of classification models using the grid search optimization method, which is one of the most commonly employed hyperparameter optimization methods (Injadat et al., 2020; Yang and Shami, 2020). Using the grid search optimization method yielded a relatively weak performance, with accuracies ranging from 37% to 70% without considering or including feature reduction and selection (Fig. 7). Obviously, the hyperparameter tuning with the Bayesian optimization method in the machine learning models produced an improved classification of the plant species in all cases, in terms of accuracy and kappa coefficient, highlighting the importance of hyperparameter optimization using the Bayesian optimization method. Interestingly, we note that, with the grid search optimization method, the feature reduction and selection is no longer a critical step, as we obtained similar or, even, slightly better results (accuracies varied from 37% to 69%, Fig. 7), especially for the LightGBM and XGBoost models.

4.4. Classification performance of individual species and future studies

We divided the species into different plant types, based on xylem properties, and found that the sensitivity, precision, and F1-score of herbaceous species (0.88, 0.89, and 0.88) were higher than woody species (0.79, 0.82, and 0.80). Further separation of woody species into evergreen broadleaf and deciduous broadleaf revealed that deciduous species have higher sensitivity, precision, and F1-score than the evergreens, the evergreen broadleaf shrub having the lowest sensitivity (0.63), precision (0.72), and F1-score (0.67). In addition, in terms of individual species, deciduous broadleaf trees (except *P. austromontana*) had the highest sensitivity, precision, and F1-score. For the species *P. austromontana,* the villous on the leaf of this species might have affected the hyperspectral signatures. Overall, our results indicate that hyperspectral leaf reflectance captured functional differences. Furthermore, we found the classification performance of most individual species in the classification models to correlate with the number of samples, more or less. When the sample numbers are large (>200, see Fig. 6), their classification accuracies are very high, although some plant species with small samples also achieved high accuracies.

Relatively high classification accuracy was obtained (for a total of 52 species) in this study when using leaf-level spectra on species classification. Even so, it was realized that it may not currently be possible to simultaneously map all species in biodiverse forests. Expanding the classification feature space with non-spectral data may be required in the future, since Castro-Esau et al. (2004) demonstrated that incorporating leaf chlorophyll content as ancillary data improved the classification of lianas and trees, suggesting the potential to include other leaf parameters into plant species classification. In addition, several studies have demonstrated that LiDAR data are valuable for identifying different species with different biophysical characteristics, thereby the classification accuracy could be further improved in combination with them (Cao et al., 2021; Dalponte et al., 2012; Liu et al., 2017). As a result, additional plant species information, such as plants' biophysical and biochemical parameters, should help to further differentiate plant species and improve the interpretability of hyperspectral signatures.

5. Conclusion

Plant species classification has been attempted using machine learning models from hyperspectral leaf information with a composite dataset covering a broad range of species and plant functional types, grown in various environments. The classification accuracy was considerably influenced by the feature selection and hyperparameter optimization approaches. The support vector machine, based on informative bands after Bayesian optimization, performed well in classing plant species. The results obtained in this study demonstrated the potential of using hyperspectral leaf information to classify plant species, which can provide a valuable basis for remote species mapping in a wider variety of ecosystems, even though leaf spectral characteristics may be influenced by environmental factors. While it remains a major challenge to collect representative ground data over large and difficult-to-entry areas, the integration of leaf hyperspectral data with other types of remote sensing data may improve classification accuracy and the ability to detect plant species in complex environments. In addition, future studies should progressively attempt to automate plant species classification using advanced machine learning and artificial intelligence techniques.

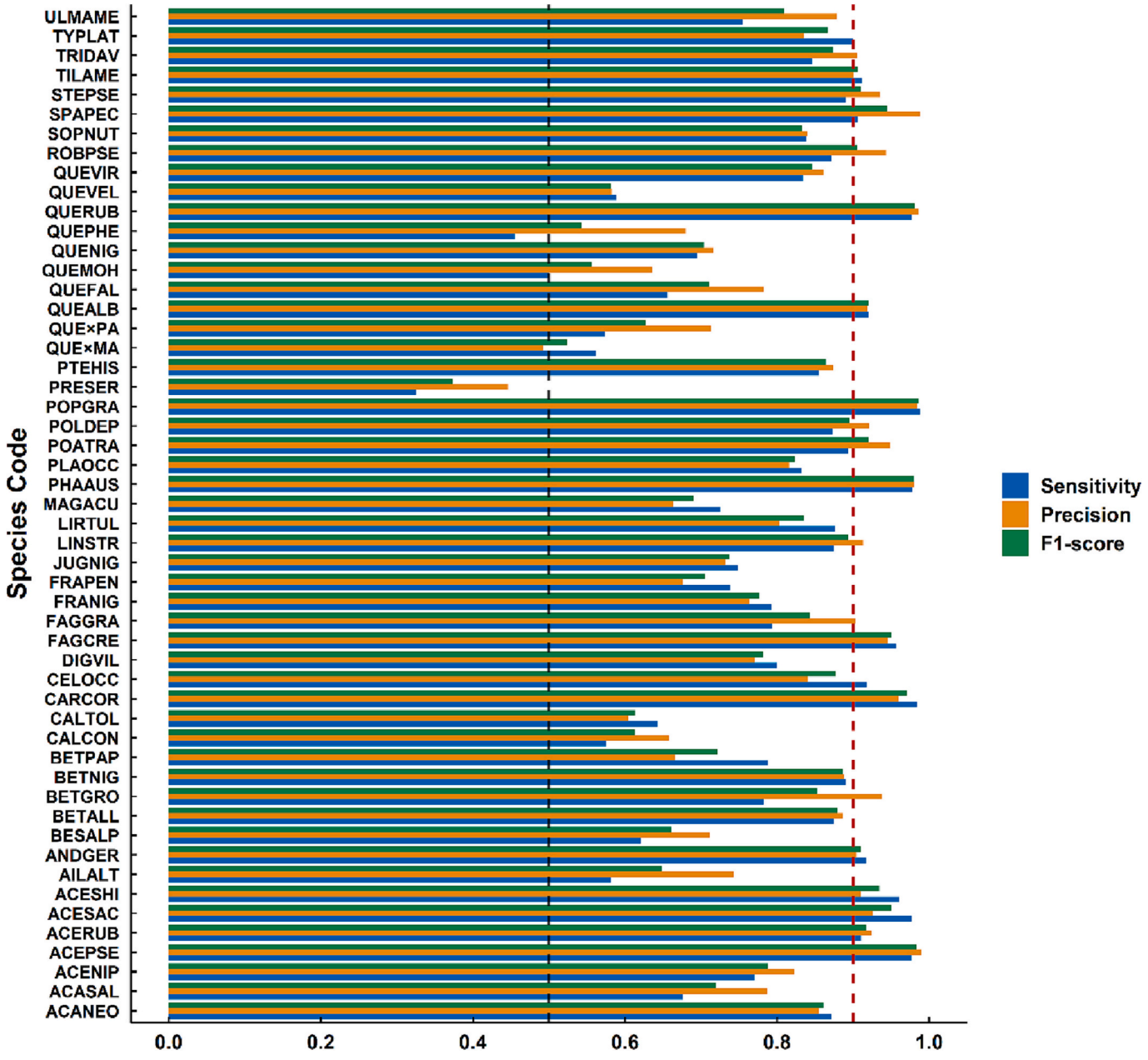

Fig. 5. Sensitivity, precision, and F1-score for each species in the Bayesian optimization-based SVM model with the RFE feature selection method.

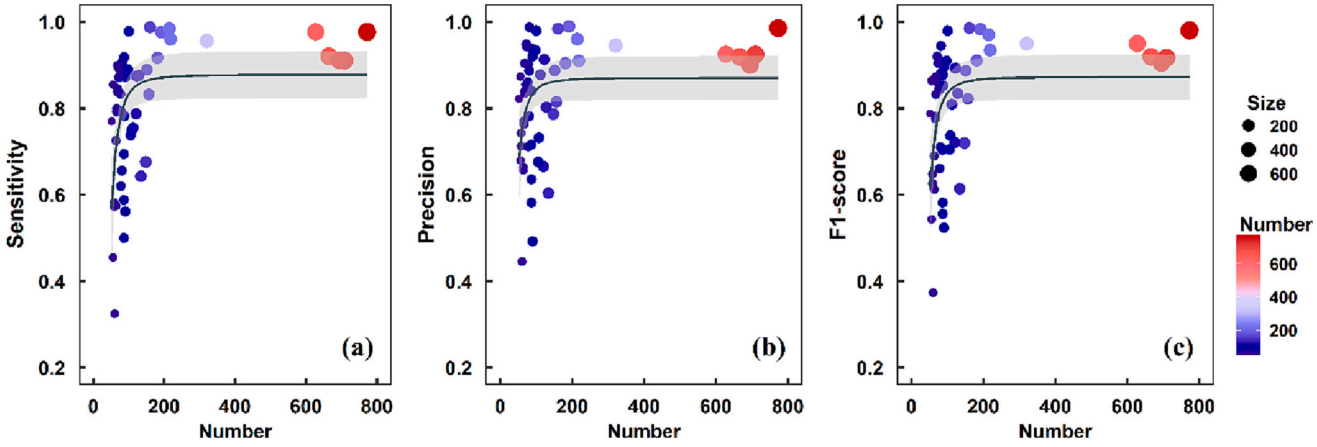

Fig. 6. Sensitivity, precision, and F1 of each species in the Bayesian optimization-based SVM model with the RFE feature selection method along with the number of samples.

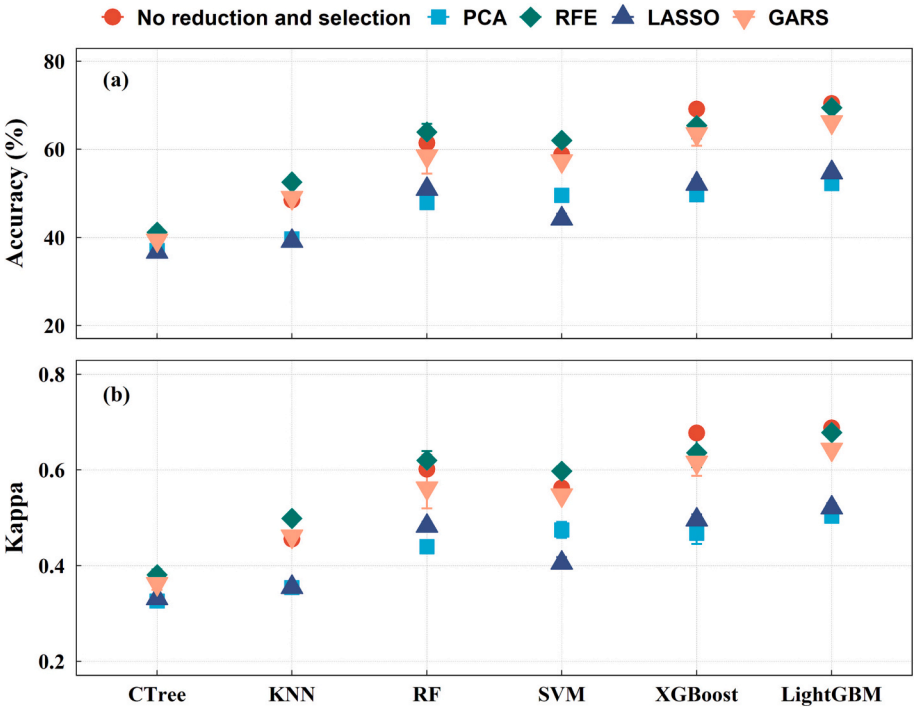

Fig. 7. Accuracy and kappa coefficient of different classification models for plant species classification with and without feature reduction and selection methods based on the grid search optimization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by the Japan Society for the Promotion of Science (JSPS) (Grant No. 21H02230). We thank the members of the Laboratory of Macroecology and the Institute of Silviculture, Shizuoka University, for providing support in conducting both fieldwork and laboratory analyses. We also gratefully acknowledge the data provided by the Ecological Spectral Information System public datasets.

References

Agrawal, T., 2021. Make your machine learning and deep learning models more efficient. In: Hyperparameter Optimization in Machine Learning. Apress Berkeley, CA, pp. 1–166.

Alonzo, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sens. Environ. 148, 70–83.

Asner, G.P., Martin, R.E., Carranza-Jiménez, L., Sinca, F., Tupayachi, R., Anderson, C.B., Martinez, P., 2014. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. New Phytol. 204, 127–139.

Aviña-Hernández, J., Ramírez-Vargas, M., Roque-Sosa, F., Martínez-Rincón, R.O., 2023. Predictive performance of random forest on the identification of mangrove species in arid environments. Ecol. Inform. 75, 102040.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.L., Deng, D., Lindauer, M., 2023. Hyperparameter optimization: foundations, algorithms, best practices, and open challenges. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 13, e1484.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cao, J., Liu, K., Zhuo, L., Liu, L., Zhu, Y., Peng, L., 2021. Combining UAV-based hyperspectral and LiDAR data for mangrove species classification using the rotation forest algorithm. Int. J. Appl. Earth Obs. Geoinf. 102, 102414.

Castro-Esau, K.L., Sánchez-Azofeifa, G.A., Caelli, T., 2004. Discrimination of lianas and trees with leaf-level hyperspectral data. Remote Sens. Environ. 90, 353–372.

Castro-Esau, K.L., Sánchez-Azofeifa, G.A., Rivard, B., Wright, S.J., Quesada, M., 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. Am. J. Bot. 93, 517–530.

Cavender-Bares, J., Meireles, J.E., Couture, J.J., Kaproth, M.A., Kingdon, C.C., Singh, A., Serbin, S.P., Center, A., Zuniga, E., Pilz, G., Townsend, P.A., 2016. Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity. Remote Sens. 8, 221.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408, 189–215.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chiesa, M., Maioli, G., Colombo, G.I., Piacentini, L., 2020. GARS: genetic algorithm for the identification of a robust subset of features in high-dimensional datasets. BMC Bioinform. 21, 54.

Chlus, A., Townsend, P.A., 2018. Seasonal fresh leaf spectra and traits, Blackhawk Island, WI. Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst.

Cho, M.A., Skidmore, A.K., Sobhan, I., 2009. Mapping beech (Fagus sylvatica L.) forest structure with airborne hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. 11, 201–211.

Clark, M.L., Roberts, D.A., 2012. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. Remote Sens. 4, 1820–1855.

Clark, M.L., Roberts, D.A., Clark, D.B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sens. Environ. 96, 375–398.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Leam. 20, 273–297.

Curran, P.J., 1989. Remote sensing of foliar chemistry. Remote Sens. Environ. 30, 271–278.

Dalponte, M., Bruzzone, L., Gianelle, D., 2012. Tree species classification in the southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sens. Environ. 123, 258–270.

Dangal, S.R.S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Syst. 3, 11.

Das, B., Sahoo, R.N., Pargal, S., Krishna, G., Verma, R., Viswanathan, C., Sehgal, V.K., Gupta, V.K., 2021. Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near

infrared spectroscopy. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 247, 119104.
Demarchi, L., Kania, A., Ciezkowski, W., Piórkowski, H., Oświecimska-Piasko, Z., Chormański, J., 2020. Recursive feature elimination and random forest classification of natura 2000 grasslands in lowland river valleys of Poland based on airborne hyperspectral and LiDAR data fusion. Remote Sens. 12, 1842.
Diniz, É.S., Amaral, C.H., Sardinha, S.T., Thiele, J., Meira-Neto, J.A.A., 2021. Phylogenetic signatures in reflected foliar spectra of regenerating plants in Neotropical forest gaps. Remote Sens. Environ. 253, 112172.
Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H.H., Leyton-Brown, K., 2013. Towards an empirical foundation for assessing Bayesian optimization of hyperparameters. NIPS Work. Bayesian Optim. Theory Pract. 10, 1–5.
Fassnacht, F.E., Neumann, C., Forster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central european test sites. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 2547–2561.
Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87.
Féret, J.B., Asner, G.P., 2011. Spectroscopic classification of tropical forest species using radiative transfer modeling. Remote Sens. Environ. 115, 2415–2422.
Ferreira, M.P., Zortea, M., Zanotta, D.C., Shimabukuro, Y.E., de Souza Filho, C.R., 2016. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sens. Environ. 179, 66–78.
Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179–188.
Frye, H.A., Aiello-Lammens, M.E., Euston-Brown, D., Jones, C.S., Kilroy Mollmann, H., Merow, C., Slingsby, J.A., van der Merwe, H., Wilson, A.M., Silander, J.A., 2021. Plant spectral diversity as a surrogate for species, functional and phylogenetic diversity across a hyper-diverse biogeographic region. Glob. Ecol. Biogeogr. 30, 1403–1417.
Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., Wolff, E., 2018. Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GISci. Remote Sens. 55, 221–242.
Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sens. Environ. 251, 112103.
Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.
Heinzel, J., Koch, B., 2012. Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. Int. J. Appl. Earth Obs. Geoinf. 18, 101–110.
Hennessy, A., Clarke, K., Lewis, M., 2020. Hyperspectral classification of plants: a review of waveband selection generalisability. Remote Sens. 12, 113.
Hesketh, M., Sánchez-Azofeifa, G.A., 2012. The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. Remote Sens. Environ. 118, 73–82.
Hosgood, B., Jacquemound, S., Andreeoli, G., Verdebout, J., Pedrini, A., Schmuck, G., 1993. Leaf Optical Properties Experiment Database (LOPEX93). Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst.
Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674.
Hycza, T., Stereńczak, K., Bałazy, R., 2018. Potential use of hyperspectral data to classify forest tree species. New Zeal. J. For. Sci. 48, 18.
Injadat, M.N., Moubayed, A., Nassif, A.B., Shami, A., 2020. Systematic ensemble model selection approach for educational data mining. Knowledge-Based Syst. 200, 105992.
Jacquemound, S., Bidel, L., Francois, C., Pavan, G., 2003. ANGERS Leaf Optical Properties Database (2003). Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst.
Jin, J., Pratama, B.A., Wang, Q., 2020. Tracing leaf photosynthetic parameters using hyperspectral indices in an Alpine deciduous forest. Remote Sens. 12, 1124.
Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G.A., Castro-Esau, K., Caelli, T., 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: comparative data reduction approaches at the leaf and canopy levels. Remote Sens. Environ. 109, 406–415.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision ree. Adv. Neural Inf. Proces. Syst. 30.
Khan, A., Vibhute, A.D., Mali, S., Patil, C.H., 2022. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. Ecol. Inform. 69, 101678.
Kothari, S., Montgomery, R., Cavender-Bares, J., 2018. FAB Leaf Spectra Across a Light Gradient at Cedar Creek LTER. Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst. https://doi.org/10.21232/FR7US97g.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.
Lever, J., Krzywinski, M., Altman, N., 2017. Points of significance: principal component analysis. Nat. Methods 14, 641–642.
Liu, L., Coops, N.C., Aven, N.W., Pang, Y., 2017. Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. Remote Sens. Environ. 200, 170–182.
Liu, E., Zhao, H., Zhang, S., He, J., Yang, X., Xiao, X., 2021. Identification of plant species in an alpine steppe of Northern Tibet using close-range hyperspectral imagery. Ecol. Inform. 61, 101213.
Malu, M., Dasarathy, G., Spanias, A., 2021. Bayesian optimization in high-dimensional spaces: a brief survey. In: 2021 12th Int. Conf. Information, Intell. Syst. Appl. (IISA). IEEE 1–8.
Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817.
Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. Remote Sens. Environ. 256, 112322.
McManus, K.M., Asner, G.P., Martin, R.E., Dexter, K.G., Kress, W.J., Field, C.B., 2016. Phylogenetic structure of foliar spectral traits in tropical forest canopies. Remote Sens. 8, 196.
Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.
Nakaji, T., Oguma, H., Nakamura, M., Kachina, P., Asanok, L., Marod, D., Aiba, M., Kurokawa, H., Kosugi, Y., Kassim, A.R., Hiura, T., 2019. Estimation of six leaf traits of East Asian forest tree species by leaf spectroscopy and partial least square regression. Remote Sens. Environ. 233, 111381.
Omeer, A.A., Deshmukh, R.R., 2021. Improving the classification of invasive plant species by using continuous wavelet analysis and feature reduction techniques. Ecol. Inform. 61, 101181.
Prospere, K., McLaren, K., Wilson, B., 2014. Plant species discrimination in a tropical wetland using in situ hyperspectral data. Remote Sens. 6, 8494–8523.
Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29, 2352–2449.
Ribeiro da Luz, B., 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. New Phytol. 172, 305–318.
Sabat-Tomala, A., Raczko, E., Zagajewski, B., 2020. Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data. Remote Sens. 12, 516.
Sameen, M.I., Pradhan, B., Lee, S., 2020. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. Catena 186, 104249.
Santos, F., Meneses, P., Hostert, P., 2019. Monitoring long-term forest dynamics with scarce data: a multi-date classification implementation in the Ecuadorian Amazon. Eur. J. Remote Sens. 52, 62–78.
Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Wang, R., Tilman, D., Gamon, J.A., 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. Nat. Ecol. Evol. 2, 976–982.
Skoneczny, H., Kubiak, K., Spiralski, M., Kotlarz, J., Mikiciński, A., Puławska, J., 2020. Fire blight disease detection for apple trees: hyperspectral analysis of healthy, infected and dry leaves. Remote Sens. 12, 2101.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. Adv. Neural Inf. Proces. Syst. 25, 1–9.
Steinbach, M., Tan, P.N., 2009. kNN: k-nearest neighbors. The Top Ten Algorithms in Data Mining. Chapman and Hall/CRC, Boca Raton, pp. 151–161.
Sumsion, G.R., Bradshaw, M.S., Hill, K.T., Pinto, L.D.G., Piccolo, S.R., 2019. Remote sensing tree classification with a multilayer perceptron. PeerJ 7, e6101.
Theissler, A., Thomas, M., Burch, M., Gerschner, F., 2022. ConfusionVis: comparative evaluation and selection of multi-class classifiers based on confusion matrices. Knowledge-Based Syst. 247, 108651.
Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288.
Ullah, S., Shakir, M., Iqbal, M.S., Iqbal, A., Ali, M., Shafique, M., Rehman, A., Godwin, J., 2021. Identifying optimal waveband positions for discriminating Parthenium hysterophorus using hyperspectral data. Ecol. Inform. 64, 101362.
Venkatasubramaniam, A., Wolfson, J., Mitchell, N., Barnes, T., Jaka, M., French, S., 2017. Decision trees in epidemiological research. Emerg. Themes Epidemiol. 14, 11.
Wang, Z., 2017a. Fresh Leaf Spectra to Estimate Foliar Functional Traits over NEON domains in eastern United States. Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst. https://doi.org/10.21232/gx9f-5546.
Wang, Z., 2017b. Fresh Leaf Spectra to Estimate LMA over NEON domains in eastern United States. Data set. Available on-line [http//ecosis.org] from Ecol. Spectr. Inf. Syst. https://doi.org/10.21232/9831-rq60.
Wang, Y., Wang, H., Peng, Z., 2021. Rice diseases detection and classification using attention based neural network and Bayesian optimization. Expert Syst. Appl. 178, 114770.
Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316.
Zhang, J., Wang, C., Yuan, L., Liu, P., Zhang, Y., Wu, K., 2020a. Construction of a plant spectral library based on an optimised feature selection method. Biosyst. Eng. 195, 1–16.
Zhang, B., Zhao, L., Zhang, X., 2020b. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. Remote Sens. Environ. 247, 111938.