Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review

Mohammadreza Sheykhmousa , Masoud Mahdianpari , *Member, IEEE*, Hamid Ghanbari ,
Fariba Mohammadimanesh, Pedram Ghamisi , *Senior Member, IEEE*,
and Saeid Homayouni , *Senior Member, IEEE*

Abstract—Several machine-learning algorithms have been proposed for remote sensing image classification during the past two decades. Among these machine learning algorithms, Random Forest (RF) and Support Vector Machines (SVM) have drawn attention to image classification in several remote sensing applications. This article reviews RF and SVM concepts relevant to remote sensing image classification and applies a meta-analysis of 251 peer-reviewed journal papers. A database with more than 40 quantitative and qualitative fields was constructed from these reviewed papers. The meta-analysis mainly focuses on 1) the analysis regarding the general characteristics of the studies, such as geographical distribution, frequency of the papers considering time, journals, application domains, and remote sensing software packages used in the case studies, and 2) a comparative analysis regarding the performances of RF and SVM classification against various parameters, such as data type, RS applications, spatial resolution, and the number of extracted features in the feature engineering step. The challenges, recommendations, and potential directions for future research are also discussed in detail. Moreover, a summary of the results is provided to aid researchers to customize their efforts in order to achieve the most accurate results based on their thematic applications.

Index Terms—Image classification, meta-analysis, random forest (RF), remote sensing (RS), support vector machine (SVM).

Mohammadreza Sheykhmousa is with OpenGeoHub, 6708 Wageningen, The Netherlands (e-mail: mohammadreza.sheykhmousa@opengeohub.org).

Masoud Mahdianpari is with C-CORE, St. John's, NL A1B 3X5, Canada, and also with the Faculty of Engineering and Applied Science, Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: m.mahdianpari@mun.ca).

Hamid Ghanbari is with the Department of Geography, Université Laval, Québec, QC G1V 0A6, Canada (e-mail: hamid.ghanbari.1@ulaval.ca).

Fariba Mohammadimanesh is with C-CORE, St. John's, NL A1B 3X5, Canada (e-mail: fariba.mohammadimanesh@c-core.ca).

Pedram Ghamisi is with the Division of "Exploration Technology," Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany (e-mail: p.ghamisi@gmail.com).

Saeid Homayouni is with the Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Quebec City, QC G1K 9A9, Canada (e-mail: saeid.homayouni@ete.inrs.ca).

Digital Object Identifier 10.1109/JSTARS.2020.3026724

I. INTRODUCTION

RECENT advances in remote sensing (RS) technologies, including platforms, sensors, and information infrastructures, have significantly increased the accessibility to the Earth Observations for geospatial analysis [1]–[5]. In addition, the availability of high-quality data, the temporal frequency, and comprehensive coverage make them advantageous for several agro-environmental applications compared to the traditional data collections approaches [6]–[8]. In particular, land use and land cover (LULC) mapping is the most common application of RS data for a variety of environmental studies, given the increased availability of RS image archives [9]–[12]. The growing applications of LULC mapping alongside the need for updating the existing maps have offered new opportunities to effectively develop innovative RS image classification techniques in various land management domains to address local, regional, and global challenges [13]–[20].

The large volume of RS data [14], the complexity of the landscape in a study area [21]–[23], as well as limited and usually imbalanced training data [24], [25], make the classification a challenging task. Efficiency and computational cost of RS image classification [26] is also influenced by different factors, such as classification algorithms [27]–[30], sensor types [31], [32], training samples [33]–[37], input features [38]–[41], pre- and postprocessing techniques [42], [43], ancillary data [44], target classes [22], [45], and the accuracy of the final product [44], [46]–[48]. Accordingly, these factors should be considered with caution for improving the accuracy of the final classification map. Carrying a simple accuracy assessment, through the overall accuracy (OA) and Kappa coefficient of agreement (K), by the inclusion of ground truth data might be the most common and reliable approach for reporting the accuracy of thematic maps. These accuracy measures make the classification algorithms comparable when independent training and validation data are incorporated into the classification scheme [28], [49]–[51].

Given the development and employment of new classification algorithms, several review articles have been published. To date, most of these reviews on RS classification algorithms have provided useful guidelines on the general characteristics of a large group of techniques and methodologies. For example, Ma *et al.* [52] represented a meta-analysis of the popular supervised
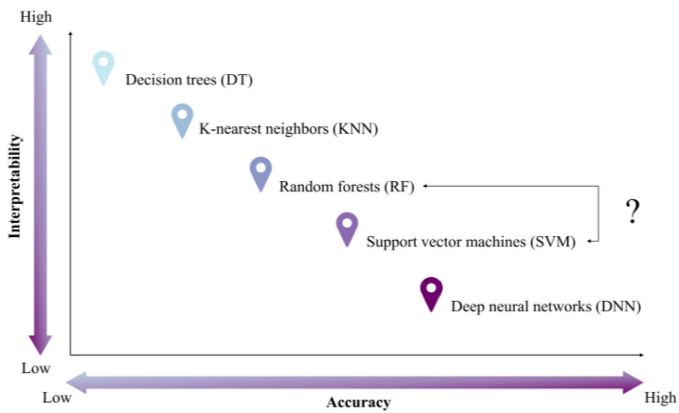
Fig. 1. Interpretability-accuracy tradeoff in machine learning classification algorithms.

object-based classifiers and reported the classification accuracies with respect to several influential factors, such as spatial resolution, sensor type, training sample size, and classification approach. Several review papers also demonstrated the performance of the classification algorithms for a specific type of application. For instance, Fassnacht *et al*. [53] summarized the significant trends in RS techniques for the classification of tree species and discussed the effectiveness of different sensors and algorithms for this application. The development in methodologies for processing a specific type of data is another dominant type of review papers in RS image classification. For example, Yan *et al*. in [54] review the usefulness of high-resolution LiDAR sensor and its application for urban land cover classification, or in [7], an algorithmic perspective review for processing of hyperspectral images is provided.

Over the past few years, deep learning algorithms have drawn attention for several RS applications [30], [55], and as such, several review articles have been published on this topic. For instance, three typical models of deep learning algorithms, namely deep belief network, convolutional neural networks, and stacked auto-encoder, were analyzed in [56]. They also discussed the most critical parameters and the optimal configuration of each model. Studies, such as [57] which compared the capability of deep learning architectures with support vector machine (SVM) for RS image classification, and [58] which focused on the classification of hyperspectral data using deep learning techniques, are other examples of RS deep learning review papers.

The commonly employed classification algorithms in the RS community include support vector machines SVMs [59]–[61], ensemble classifiers, e.g., random forest (RF) [10], and deep learning algorithms [27]. Deep learning methods have the ability to retrieve complex patterns and informative features from the satellite image data. For example, CNN has shown performance improvements over SVM and RF [62], [63]. However, one of the main problems with deep learning approaches is their hidden layers; "black box" [60] nature, which results in the loss of interpretability (see Fig. 1). Another limitation of a deep learning models is that they are highly dependent on the amount of training data, i.e.; ground truth. Moreover, implementing CNN required expert knowledge and computationally is expensive and

needs dedicated hardware to handle the process. On the other hand, recent research works show SVM and RF (i.e., relatively easily implementable methods) can handle learning tasks with a small amount of training dataset, yet demonstrate competitive results with CNNs [64]. Deep learning methods have the ability to retrieve complex patterns and informative features from the satellite imagery. For example, CNN has shown performance improvements over SVM and RF [62], [63]. However, one of the main problems with deep learning approaches is their hidden layers; "black box" nature [60], which results in the loss of interpretability (see Fig. 1). Another limitation of deep learning models is that they are highly dependent on the availability of abundant high-quality ground truth data. Moreover, implementing CNN requires expert knowledge and it is computationally expensive and needs dedicated hardware to handle the process. On the other hand, recent research works show SVM and RF (i.e., relatively easily implementable methods) can handle learning tasks with a small amount of training dataset, yet demonstrate competitive results with CNNs [64]. Although there is an ongoing shift in the application of deep learning in RS image classification, SVM and RF have still held the researchers' attention due to lower computational complexity and higher interpretability capabilities compared to deep learning models. More specifically, SVM maintenance among the top classifiers is mainly because of its ability to tackle the problems of high dimensionality and limited training samples [65], while RF holds its position due to ease of use (i.e., does not need much hyperparameter finetuning) and its ability to learn both simple and complex classification functions [66], [67]. As a result, the relatively high similar performance of SVM and RF in terms of classification accuracies makes them among the most popular machine learning classifiers within the RS community [68]. As a result, giving merit to one of them is a difficult task as past comparison-based studies, as well as some review papers, provide readers with often contradictory conclusions, which was somehow confusing. For instance, Mountrakis *et al.* in [66] reported SVMs can be considered the "best of class" algorithms for classification; however, the work in [67] and [69] suggested that RF classifiers may outperform support vector machines for RS image classification. This knowledge gap was identified in the field of bioinformatics and filled by an exclusive review of RFs versus SVMs [70]; however, no such a detailed survey is available for RS image classification.

Table I summarizes the review papers on recent classification algorithms of RS data, where a large part of the literature is devoted to RFs or is discussed as an alternative classifier. The majority of these review papers are descriptive and do not offer a quantitative assessment of the stability and suitability of RF and SVM classification algorithms. Accordingly, the general objective of this study is to fill this knowledge gap by comparing RF and SVM classification algorithms through a meta-analysis of published papers and provide RS experts with a "big picture" of the current research in this field. To the best of the authors' knowledge, this is the first study in the RS filed that provides a one to one comparison analysis for RF and SVM in various RS applications.

TABLE I
SUMMARY OF RELATED SURVEYS ON REMOTE SENSING IMAGE CLASSIFICATION (THE NUMBER OF CITATIONS IS REPORTED BY APRIL 20, 2020)

No.	Title	# papers	Year	Citation	Publication
1	A Survey of Image Classification Methods and Techniques for Improving Classification Performance	NA	2007	2365	IJRS
2	Support Vector Machines in Remote Sensing: A Review	108	2011	1881	ISPRS
3	Random Forest in Remote Sensing: A Review of Applications and Future Directions	NA	2016	1037	ISPRS
4	Advances in Hyperspectral Image Classification: Earth Monitoring with Statistical Learning Methods	NA	2013	467	IEEE SPM
5	Optical Remotely Sensed Time Series Data for Land Cover Classification	NA	2016	414	ISPRS
6	Quality Assessment of Image Classification Algorithms for Land-Cover Mapping: A Review and A Proposal for A Cost-Based Approach	NA	1999	382	IJRS
7	Urban Land Cover Classification Using Airborne Lidar Data: A Review	NA	2015	258	RSE
8	A Review of Supervised Object-Based Land-Cover Image Classification	173	2017	257	ISPRS
9	A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classification Processes: General Guidelines for Practitioners and Future Research	266	2016	216	RSE
10	A Review of Remote Sensing Image Classification Techniques: The Role of Spatio-Contextual Information	NA	2014	202	EJRS
11	Advanced Spectral Classifiers for Hyperspectral Images: A Review	NA	2017	190	IEEE GRSM
12	Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review	NA	2018	164	IJRS
13	Developments in Landsat Land Cover Classification Methods: A Review	36	2017	98	RS/MDPI
14	Multiple Kernel Learning for Hyperspectral Image Classification: A Review	NA	2017	75	IEEE TGRS
15	Meta-Discoveries from A Synthesis of Satellite-Based Land-Cover Mapping Research	1651	2014	74	IJRS
16	Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review	176	2019	70	ISPRS
17	New Frontiers in Spectral-Spatial Hyperspectral Image Classification: The Latest Advances Based on Mathematical Morphology, Markov Random Fields, Segmentation, Sparse Representation, and Deep Learning	NA	2018	57	IEEE GRSM
18	Remote Sensing Image Classification: A Survey of Support-Vector-Machine-Based Advanced Techniques	55	2017	53	IEEE GRSM
19	Meta-Analysis of Deep Neural Networks in Remote Sensing: A Comparative Study of Mono-Temporal Classification to Support Vector Machines	103	2019	5	ISPRS

To fulfill the proposed meta-analysis task, more than 250 peer-reviewed papers have been reviewed in order to construct a database of case studies that include RFs and SVMs either in one-to-one comparison or individually with other machine learning methods in the field of RS image classification.

II. OVERVIEW OF SVM AND RF CLASSIFIER

Image classification algorithms can be broadly categorized into supervised and unsupervised approaches. Supervised classifiers are preferred when sufficient amounts of training data are available. Parametric and nonparametric methods are another categorization of classification algorithms based on data distribution assumptions (see Fig. 2). Yu *et al.* [71] reviewed 1651 articles and reported that supervised parametric algorithms were the most frequently used technique for RS image classification. For example, the maximum likelihood (ML) classifier, as a supervised parametric approach, was employed in more than 32% of the reviewed studies. Supervised nonparametric algorithms, such as SVM and ensemble classifiers, obtained more accurate results, yet they were less frequently used compared to supervised parametric methods [71].

Prior to the introduction of RF, SVM has been in the spotlight for RS image classification, given its superiority compared to ML classifier, K-nearest neighborsKNN, artificial neural networks, and decision tree for RS image classification. Since the
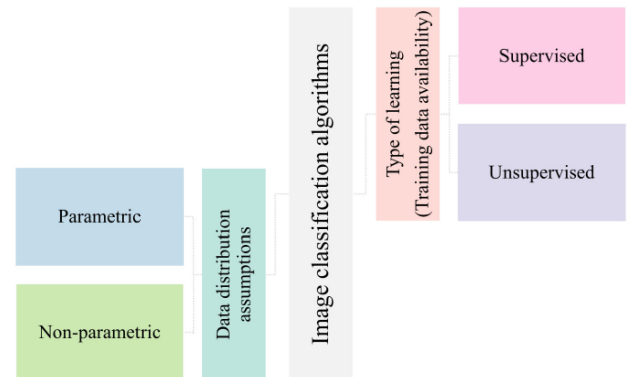

Fig. 2   Taxonomy of image classification algorithms.

introduction of RF, however, it has drawn attention within the RS community, as it produced classification accuracies comparable with those of SVM [67], [68]. An overview of SVM and RF classification algorithms has been presented in the following sections.

A. Support Vector Machine Classifier

The SVM algorithm, introduced first in the late 1970s by Vapnik and his group, is one of the most widely used kernel-based learning algorithms in a variety of machine learning
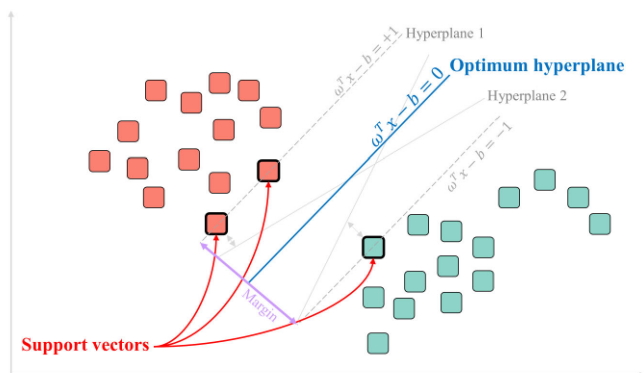
Fig. 3.   SVM example of linearly separable data.

Fig. 4.   SVM example of nonlinearly separable data with the kernel trick.

applications, and especially image classification [66]. In SVMs, the main goal is to solve a convex quadratic optimization problem for obtaining a globally optimal solution in theory and thus, overcoming the local extremum dilemma of other machine learning techniques. SVM belongs to nonparametric supervised techniques, which are insensitive to the distribution of underlying data. This is one of the advantages of SVMs compared to other statistical techniques, such as ML, wherein data distribution should be known in advance [68].

SVM, in its basic form, is a linear binary classifier, which identifies a single boundary between two classes. The linear SVM assumes the multidimensional data are linearly separable in the input space (see Fig. 3). In particular, SVMs determine an optimal hyperplane (a line in the simplest case) to separate the dataset into a discrete number of predefined classes using the training data. To maximize the separation or margin, SVMs use a portion of the training sample that lies closest in the feature space to the optimal decision boundary, acting as support vectors [72]–[75]. These samples are the most challenging data to classify and have a direct impact on the optimum location of the decision boundary [38], [75]. The optimal hyperplane, or the maximal margin, can be mathematically and geometrically defined. It refers to a decision boundary that minimizes misclassification errors attained during the training step [65], [66]. As seen in Fig. 3, a number of hyperplanes with no sample between them are selected, and then the optimal hyperplane is determined when the margin of separation is maximized [75], [76]. This iterative process of constructing a classifier with an optimal decision boundary is described as the learning process [77].

In practice, the data samples of various classes are not always linearly separable and overlap with each other (see Fig. 4). Thus, linear SVM cannot guarantee a high accuracy for classifying such data and needs some modifications. Cortes and Vapnik [78] introduced the soft margin and kernel trick methods to address the limitation of linear SVM. To deal with nonlinearly separable data, additional variables (i.e., slack variables) can be added to SVM optimization in the soft margin approach. However, the idea behind the kernel trick is to map the feature space into a higher dimension (Euclidean or Hilbert space) to improve the separability between classes [78], [79]. In other words, using the kernel trick, an input dataset is projected into
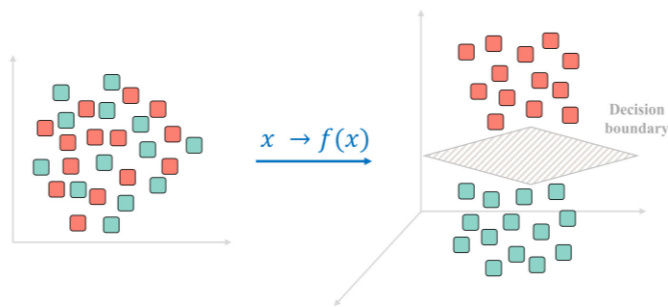
a higher-dimensional feature space where the training samples will become linearly separable.

The performance of SVM largely depends on the suitable selection of a kernel function that generates the dot products in the higher-dimensional feature space. This space can theoretically be of an infinite dimension where the linear discrimination is possible. There are several kernel models to build different SVMs, satisfying the Mercer's condition, including Sigmoid, Radial basis function, Polynomial, and Linear [80]. Commonly used kernels for remotely sensed image analysis are polynomial and the radial basis function (RBF) kernels [66]. Generally, kernels are selected by predefining a kernel's model (Gaussian, polynomial, etc.) and then adjusting the kernel parameters by tuning techniques, which could be computationally very costly. The classifier's performance, on a portion of the training sample or validation set, is the most important criteria for selecting a kernel function. However, kernel-based models can be quite sensitive to overfitting, which is possibly the main limitation of kernel-based methods, such as SVM [57]. Accordingly, innovative approaches, including automatic kernel selection [75], [79], [81], and multiple kernel learning [82], were proposed to address this problem. Notably, the task of determining the optimum kernel falls into the category of the optimization problem [83]–[87]. Optimizing several SVM parameters is very resource-intensive. So there comes a need for an alternate way of searching out the SVM parameters; genetic optimization algorithm (GA) [88]–[90] and particle swarm optimization (PSO) algorithm [91]. GA-SVM and SVM-PSO are both evolutionary techniques that exploit principles inspired from biological systems to optimize C and gamma. Compared with GA and other similar evolutionary techniques, PSO has some attractive characteristics and, in many cases, proved to be more effective [91].

The binary nature of SVMs usually involves complications on their use for multiclass scenarios, which frequently occur in RS. This requires a multiclass task to be broken down into a series of simple SVM binary classifiers, following either the one-against-one or one-against-all strategies [92]. However, the binary SVM can be extended to a one-shot multiclass classification, requiring a single optimization process. For example, a complete set of five-class classifications only requires to be optimized once for determining the kernel's parameters C (a parameter that controls the amount of penalty during the SVM optimization) and γ (spread of the RBF kernel), in contrast to the five-times for one-against-all and ten-times for the one-against-one methods,
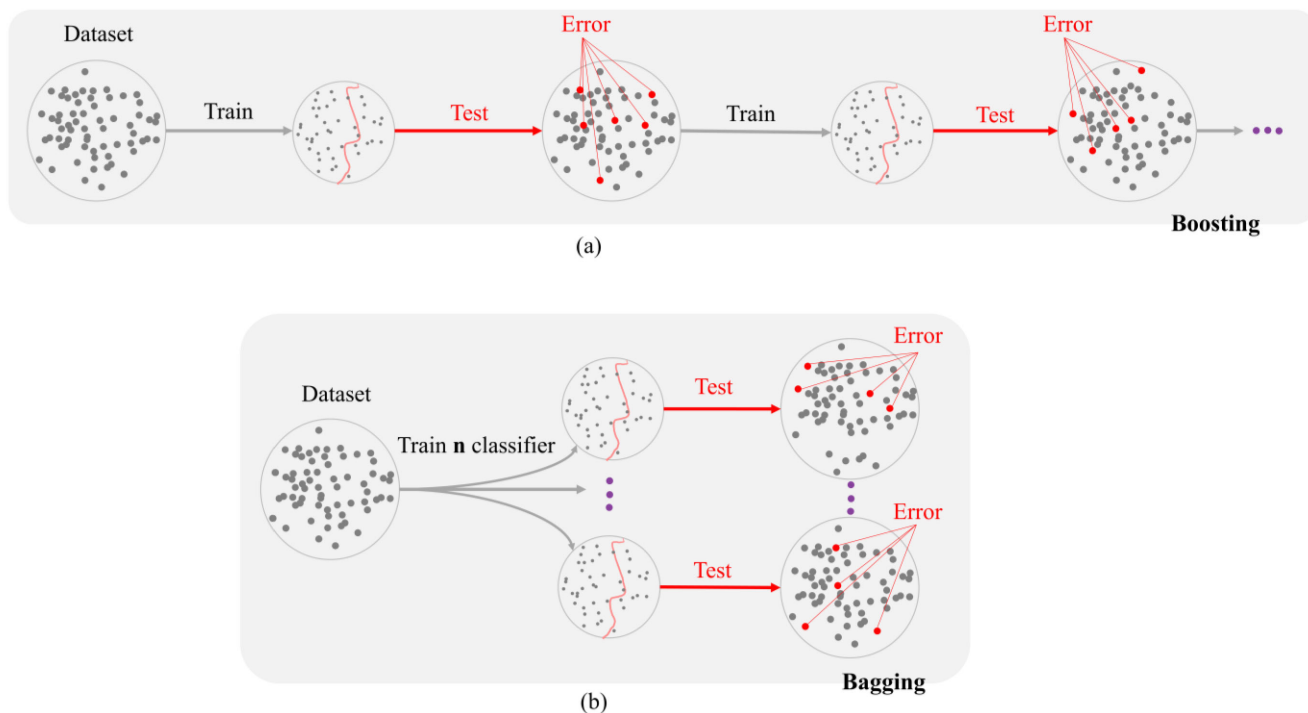
Fig. 5.　Widely used ensemble learning methods. (a) Boosting. (b) Bagging.

respectively. Furthermore, the one-shot multiclass classification uses fewer support vectors while, unlike the conventional one-against-all strategy, it guarantees to produce a complete confusion matrix [93]. SVM classifies high-dimensional big data into a limited number of support vectors, thus achieving the distinction between subgroups within a short period of time. However, the classification of big data is always computationally expensive. As such, hybrid methodologies of the SVM, e.g., Granular Support Vector Machine (GSVM), are introduced in different applications. GSVM is a novel machine learning model based on granular computing and statistical learning theory that addresses the inherently low-efficiency learning problem of the traditional SVM while obtaining satisfactory generalization performance [94].

SVMs are particularly attractive in the field of RS owing to their ability to manage small training data sets effectively and often delivering higher classification accuracy compared to the conventional methods [95]–[98]. SVM is an efficient classifier in high-dimensional spaces, which is particularly applicable to RS image analysis field where the dimensionality can be extremely large [99], [100]. In addition, the decision process of assigning new members only needs a subset of training data. As such, SVM is one of the most memory-efficient methods, since only this subset of training data needs to be stored in memory [100]. The ability to apply new kernels rather than linear boundaries also increases the flexibility of SVMs for the decision boundaries, leading to a greater classification performance. Despite these benefits, there are also some challenges, including the choice of a suitable kernel, optimum kernel parameters selection, and the relatively complex mathematics behind the SVM, especially from a nonexpert user point of view, that restricts the effective cross-disciplinary applications of SVMs [101].

B. Random Forest Classifier

RF is an ensemble learning approach, developed by Breiman [102], for solving classification and regression problems. Ensemble learning is a machine learning scheme to boost accuracy by integrating multiple models to solve the same problem. In particular, multiple classifiers participate in ensemble classification to obtain more accurate results compared to a single classifier. In other words, the integration of multiple classifiers decreases variance, especially in the case of unstable classifiers, and may produce more reliable results. Next, a voting scenario is designed to assign a label to unlabeled samples [11], [103]–[105]. The commonly used voting approach is majority voting, which assigns the label with the maximum number of votes from various classifiers to each unlabeled sample [106]. The popularity of the majority voting method is due to its simplicity and effectiveness. More advanced voting approaches, such as the veto voting method, wherein one single classifier vetoes the choice of other classifiers, can be considered as an alternative for the majority voting method [107].

The widely used ensemble learning methods are boosting and bagging. Boosting is a process of building a sequence of models, where each model attempts to correct the error of the previous one in that sequence [see Fig. 5(a)]. AdaBoost was the first successful boosting approach, which was developed for binary classification cases [108]. However, the main problem of AdaBoost is model overfitting [103]. The Bootstrap Aggregating, known as Bagging, is another type of ensemble learning methods
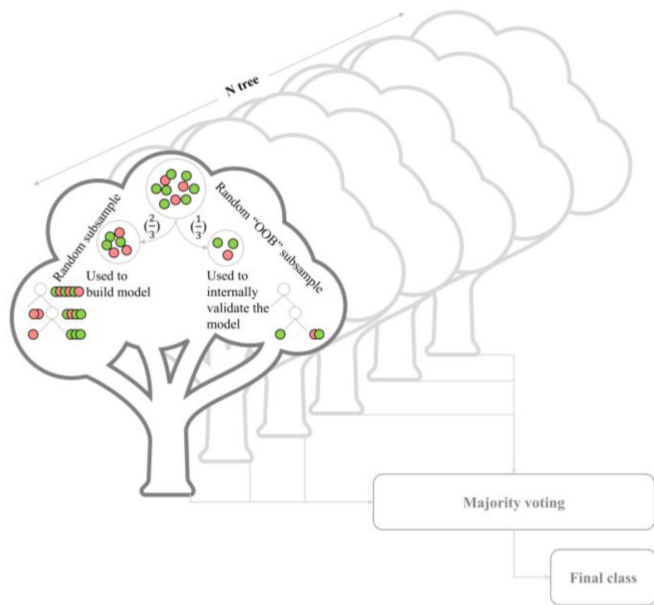
Fig. 6.   Random forest model.

[109]. Bagging is designed to improve the stability and accuracy of integrated models while reducing variance [see Fig. 5(b)]. As such, Bagging is recognized to be more robust against the overfitting problem compared to the boosting approach [102].

RF was the first successful bagging approach, which was developed based on the combination of Breiman's bagging sampling approach, random decision forests, and the random selection of features independently introduced in [110]–[112].

RFs with significantly different tree structures and splitting variables encourage different instances of overfitting and outliers among the various ensemble tree models. Therefore, the final prediction voting mitigates the overfitting in case of the classification problem, while the averaging is the solution for the regression problems. Within the generation of these individual decision trees, each time the best split in the random sample of predictors is chosen as the split candidates from the full set of predictors. A fresh sample of predictors is taken at each split utilizing a user-specified number of predictors (Mtry). By expanding the RF up to a user-specified number of trees (Ntree), RF generates high variance and low bias trees. Therefore, new sets of input (unlabeled) data are assessed against all decision trees that are generated in the ensemble, and each tree votes for a class's membership. The membership with the majority votes will be the one that is eventually selected (see Fig. 6) [113], [114]. This process, hence, should obtain a global optimum [115]. To reach a global optimum, two-thirds of the samples, on average, are used to train the bagged trees, and the remaining samples, namely the out-of-bag (OOB) are employed to cross validate the quality of the RF model independently. The OOB error is used to calculate the prediction error and then to evaluate variable importance measures (VIMs) [116], [117].

Of particular RFs' characteristic is VIM [118]–[121]. Specifically, VIM allows a model to evaluate and rank predictor variables in terms of relative significance [26], [45]. VIM calculates the correlation between high-dimensional datasets on the basis of internal proximity matrix measurements [122] or identifying outliers in the training samples by exploratory examination of sample proximities through the use of variable importance metrics [44]. The two major variable importance metrics are: mean decrease in Gini (MDG) and mean decrease in accuracy (MDA) [123]–[125]. MDG measures the mean decrease in node impurities as a result of splitting and computes the best split selection. MDG switches one of the random input variables while keeping the rest constant. It then measures the decrease in the accuracy, which has taken place by means of the OOB error estimation and Gini Index decrease. In a case where all predictors are continuous and mutually uncorrelated, Gini VIM is not supposed to be biased [26]. MDA, however, takes into account the difference between two different OOB errors, the one that resulted from a dataset obtained by randomly permuting the predictor variable of interest and the one resulted from the original dataset [102].

To run the RF model, two parameters have to be set: The number of trees (Ntree) and the number of randomly selected features (Mtry). RFs are reported to be less sensitive to the Ntree compared to Mtry [126]. Reducing Mtry parameter may result in faster computation, but reduces both the correlation between any two trees and the strength of every single tree in the forest and thus, has a complex influence on the classification accuracy [113]. Since the RF classifier is computationally efficient and does not overfit, Ntree can be as large as possible [127]. Several studies found 500 as an optimum number for the Ntree because the accuracy was not improved by using Ntrees higher than this number [115]. Another reason for this value being commonly used could be the fact that 500 is the default value in the software packages like R package; "randomForest" [128].

In contrast, the number of Mtry is an optimal value and depends on the data at hand. The Mtry parameter is recommended to be set to the square root of the number of input features in classification tasks and one-third of the number of input features for regression tasks [102]. Although methods based on bagging, such as RF, unlike other methods based on boosting, are not sensitive to noise or overtraining [129], [130], the above-stated value for Mtry might be too small in the presence of a large number of noisy predictors, i.e., in the case of noninformative predictor variables, the small Mtry results in building inaccurate trees [26].

The RF classifier has become popular for classification, prediction, studying variable importance, variable selection, and outlier detection since its emergence in 2001 by Breiman [102]. They have been widely adopted and applied as a standard classifier to a variety of prediction and classification tasks, such as those in bioinformatics [131], computer vision [132], and RS land cover classification [133]. RF has gained its popularity in land cover classification due to its clear and understandable decision-making process and excellent classification results [115], as well as easy implementation of RF in a parallel structure for geo-big data computing acceleration [134]. Other advantages of RF classifier can be summarized as follows:

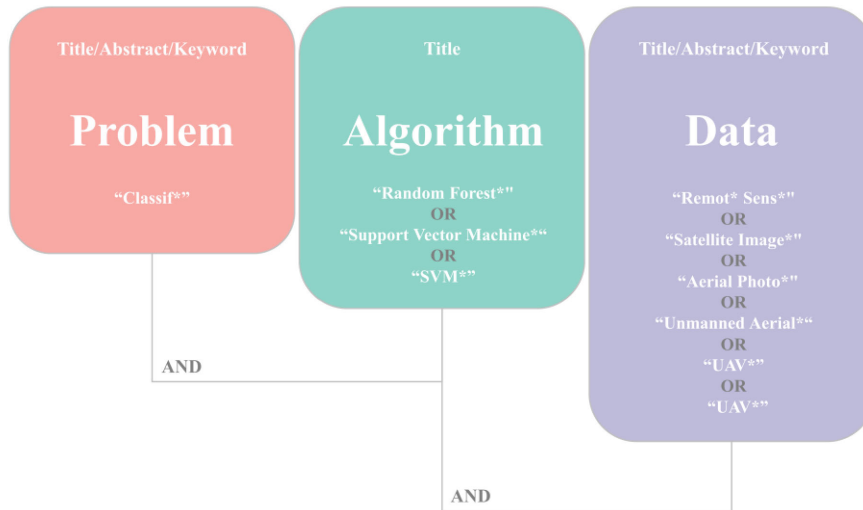1) handling thousands of input variables without variable deletion;

Fig. 7.   Article search query design.

2) reducing the variance without increasing the bias of the predictions;
3) computing proximities between pairs of cases that can be used in locating outliers;
4) being robust to outliers and noise; and
5) being computationally lighter compared to other tree ensemble methods, e.g., Boosting.

As such, many research works have illustrated the great capability of the RS classifier for classification of Landsat archive, hyper and multispectral image classification [135], and digital elevation model data [136]. Most RF research has demonstrated that RF has improved accuracy in comparison to other supervised learning methods and provide VIMs that is crucial for multisource studies, where data dimensionality is considerably high [137].

III. METHODS

To prepare for this comprehensive review, a systematic literature search query was performed using the Scopus and the Web of Science, which are two big bibliographic databases and cover scholarly literature from approximately any discipline. Notably, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) were followed for study selection [138]. After numerous trials, three groups of keywords were considered to retrieve relevant literature in a combination on and up to October 28, 2019 (see Fig. 7). The keywords in the first and last columns were searched in the topic (title/abstract/keyword) to include papers that used data from the most common RS platforms and addressed a classification problem. However, the keywords in the second column were exclusively searched in the title to narrow the search down. This resulted in obtaining studies that only employed SVM and RF algorithms in their analysis.

To identify significant research findings and to keep a manageable workload, only those studies that had been published in one of the well-known journals in the RS community (see Table II) have been considered in this literature review.
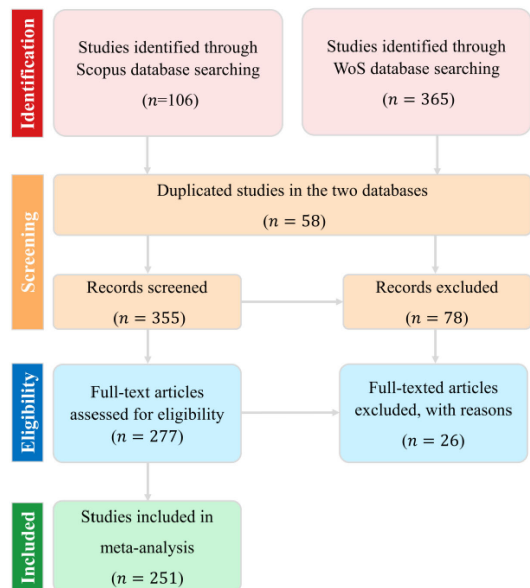

Fig. 8.   PRISMA flowchart demonstrating the selection of studies.

Of the 471 initial number of studies, 251 were eligible to be included in the meta-analysis with the following attributes: title, publication year, first author, journal, citation, application, sensor type, data type, classifier, image processing unit, spatial resolution, spectral resolution, number of classes, number of features, optimization method, test/train portion, software/library, single-date/multitemporal, overall accuracy, as well as some specific attributes for each classifier, such as kernel types for SVM and number of trees for RF. A summary of the literature search is demonstrated in Fig. 8.

IV. RESULTS AND DISCUSSION

A. General Characteristics of Studies

Following the in-depth review of 251 publications on supervised remotely-sensed image classification, relevant data

TABLE II
REMOTE SENSING JOURNALS USED TO COLLECT RESEARCH STUDIES FOR THIS LITERATURE REVIEW

Journal	Publication	Impact factor (2019)
Remote Sensing of Environment	Elsevier	8.218
ISPRS Journal of Photogrammetry and Remote Sensing	Elsevier	6.942
Transaction on Geoscience and Remote Sensing	IEEE	5.63
International Journal of Applied Earth Observation and Geoinformation	Elsevier	4.846
Remote Sensing	MDPI	4.118
GISience Remote Sensing	Taylor & Francis	3.588
Journal of Selected Topics in Applied Earth Observations and Remote Sensing	IEEE	3.392
Patterns Recognition Letters	Elsevier	2.81
Canadian Journal of Remote Sensing	Taylor & Francis	2.553
International Journal of Remote Sensing	Taylor & Francis	2.493
Remote Sensing Letters	Taylor & Francis	2.024
Journal of Applied Remote Sensing	SPIE	1.344

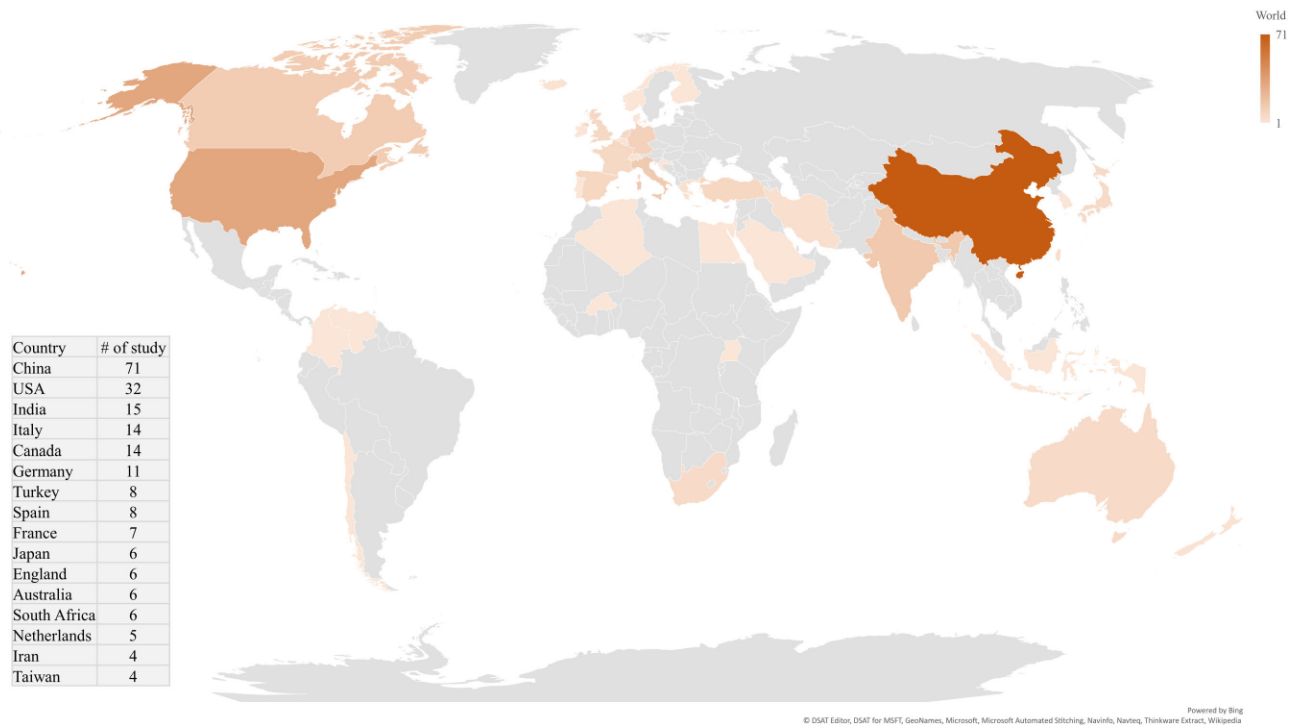| Country	# of study
China	71
USA	32
India	15
Italy	14
Canada	14
Germany	11
Turkey	8
Spain	8
France	7
Japan	6
England	6
Australia	6
South Africa	6
Netherlands	5
Iran	4
Taiwan	4

Fig. 9. Distribution of research institutions, according to the country reported in the article. The countries with more than three studies are presented in the table.

were obtained using the methods described in Section III. The primary sources of information were articles published in scientific journals. In this section, we conduct analysis about the geographical distribution of the research papers and discuss the frequency of those papers based on RF and SVM considering time, journals, and application domains. This was followed by statistical analysis of RS software packages used in the case studies, given RFs and SVMs. Furthermore, we reported the result and discussed the finding of classification performance

against essential features, i.e., data type, RS applications, spatial resolution, and finally, the number of extracted features.

Fig. 9 illustrates the geographical coverage of published papers based on the research institutions reported in the article on a global scale. More than three studies were published in 16 countries from 6 continents, including Asia 41%, Europe 32%, North America 18%, and the others 9%. As can be seen, most of the studies have been carried out in Asia and specifically in China by 71 studies; more than two times higher than the number
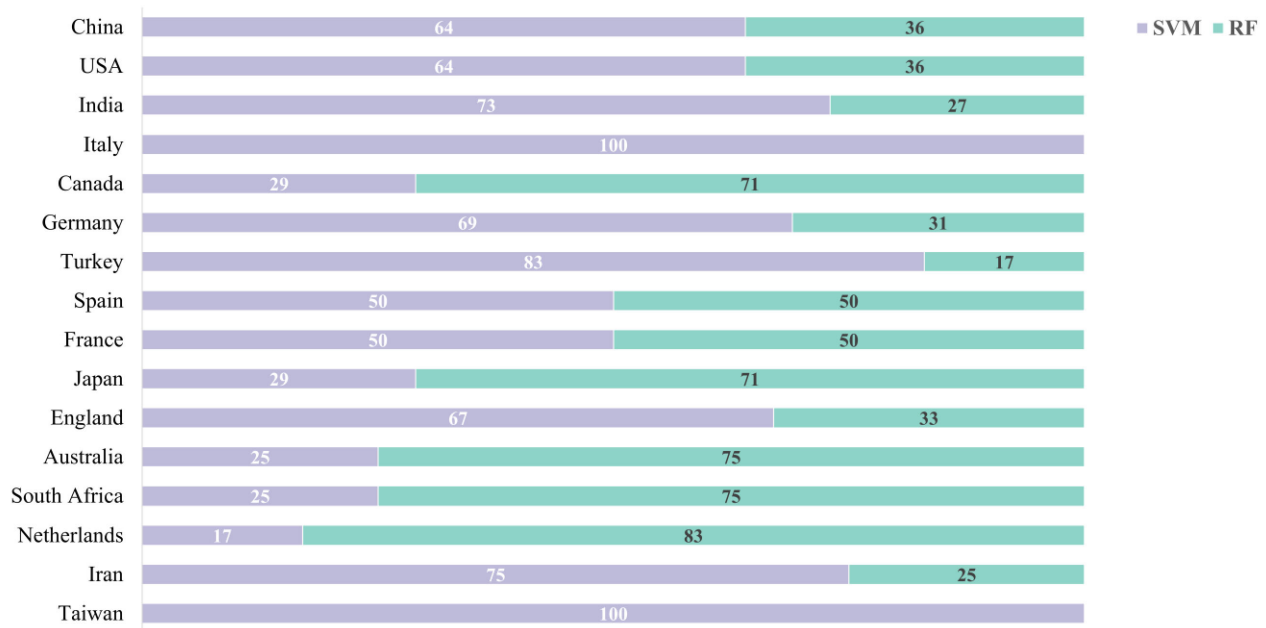
Fig. 10. Usage rate of SVM and RF for remote sensing image classification in countries with more than three studies.
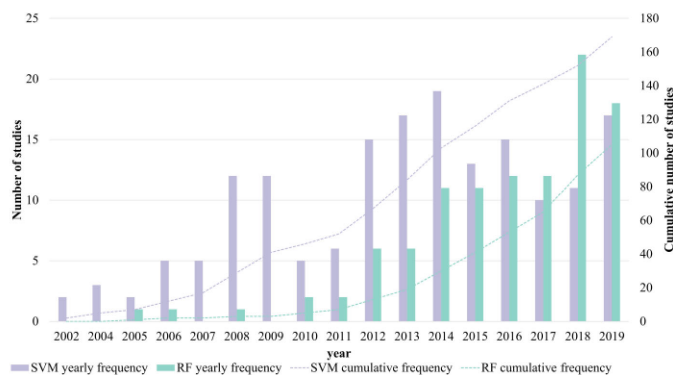

Fig. 11. Cumulative frequency of studies that used SVM or RF algorithms for remote sensing image classification.

of studies conducted in the following country (i.e., USA). The papers from China and the USA have been over 40% of the studies. This is maybe due to the extensive scientific studies conducted by the universities and institutions located in these countries, as well as the availability of the datasets. Fig. 10 also demonstrates the popularity of RF and SVM classifiers within these 16 countries. It is interesting that in the Netherlands and Canada, RF is frequently applied, while in China and the USA, SVM attracts more scientists. It was only in Italy and Taiwan that all the studies used merely one of the classifiers (i.e., SVMs).

Fig. 11 represents the annual frequency of the publications and the equivalent cumulative distribution for RF and SVM algorithms applied for RS image classification. The resulting list of papers includes 251 studies published in 12 different peer-reviewed journals. Case studies were assembled from 2002 to 2019, a period of 17 years, wherein the first study using SVM dates back to 2002 [139] and 2005 for RFs [137]. The graph shows the significant difference between studies using

SVM and RF over the time frame. Apart from a brief drop between 2010 and 2011, there was a moderate increase in case studies using SVM from 2002 to 2014. However, this slips back for four subsequent years, followed by significant growth in 2019. Studies using RFs, on the other hand, shows the steady increase in the given time-span, which resulted in an exponential distribution in the equivalent cumulative distribution function. The number of studies employed SVMs were always more than those that used RFs until 2016, wherein the number crossed over in favor of utilizing RFs. The sharp decrease in using both RF and SVM between 2014 and 2019 can be clearly explained by the advent of utilizing deep learning models in the RS community [140]. However, it seems that employing RF and SVM regained the researchers' attention from 2018 onward. Overall, 170 (68%) and 105 (42%) studies used SVM and RF for their classification task, respectively. It is noteworthy to mention that some papers include the implementation of both methods. Besides, the graph illustrates that utilizing SVMs fluctuated more than RFs, and both classifiers keep almost steady growth, given the time period. More information on datasets will be given in the next section.

Fig. 12 demonstrates the number of publications in each of these 12 peer-reviewed journals, as well as their contribution in using RF and SVM. Nearly one-fourth of the papers (24%) were published in Remote Sensing (RS MDPI) with the majority of the remaining published in the Institute of Electrical and Electronics Engineers (IEEE) hybrid journals (JSTARS, TGRS, and GRSL; 23%), *International Journal of Remote Sensing* (IJRS; 19%), *Remote Sensing of Environment* (RSE; 8%), and six other journals (26%). Although in most of the journals the number of SVM- and RF-related articles are high enough, three journals have published less than ten papers with RF or SVM classification implementations (i.e., GIS&RS, CJRS, and PRL).

The scheme for RS applications used for this study consisted of eight broad Classification groupings referred to as classes
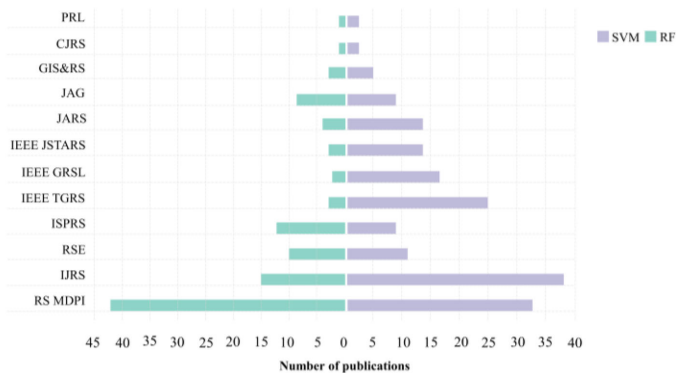
Fig. 12. Number of publications per journal; Remote Sensing of Environment (RSE), ISPRS Journal of Photogrammetry and Remote Sensing (ISPRS), Transaction on Geoscience and Remote Sensing (IEEE TGRS), International Journal of Applied Earth Observation and Geoinformation (JAG), Remote Sensing (RS MDPI), GIScience Remote Sensing (GIS & RS), Journal of Selected Topics in Applied Earth Observations and Remote Sensing (IEEE JSTARS), Pattern Recognition Letters (PRL), Canadian Journal of Remote Sensing (CJRS), International Journal of Remote Sensing (IJRS), Remote Sensing Letters (RSL), and Journal of Applied Remote Sensing (JARS).
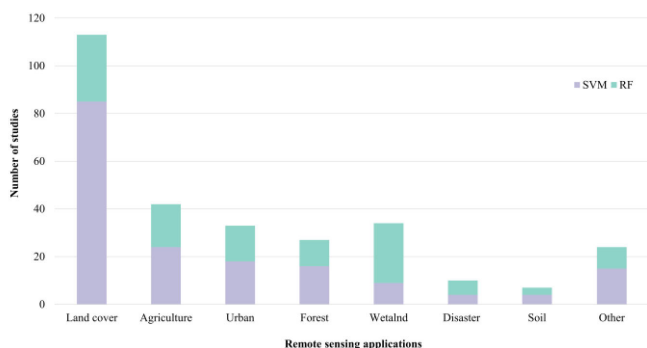

Fig. 13. Number of studies that used SVM or RF algorithms in different remote sensing applications.

distributed across the world (see Fig. 13). The most frequently investigated application, representing 39% of studies, was related to land cover mapping, with other categorial applications including agriculture (15%), urban (11%), forest (10%), wetland (12%), disaster (3%), and soil (2%). The remaining applications comprising about 8% of the case studies mainly consist of mining area classification, water mapping, benthic habitat, rock types, and geology mapping.

B. Statistical and Remote Sensing Software Packages

A comparison of the geospatial and image processing software and other statistical packages is depicted in Fig. 14. These software packages shown here were used for the implementation of both the SVM and RF methods in at least three journal papers. The software packages include eCognition (Trimble), ENVI (Harris Geospatial SolutionsInc.), ArcGIS (ESRI), Google Earth Engine, Geomatica (PCI Geomatics) as well as statistical and data analysis tools, which are MATLAB (MathWorks), and open-source software tools such as R, Python, OpenCV, and
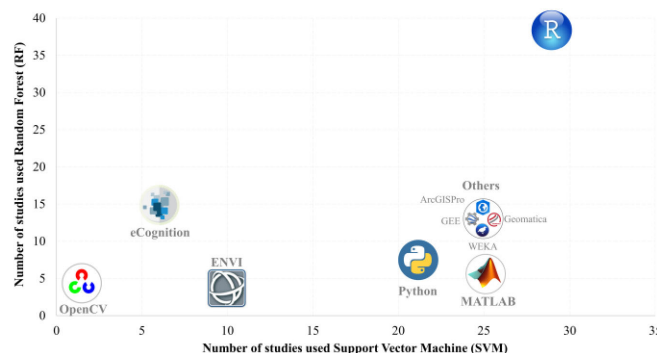

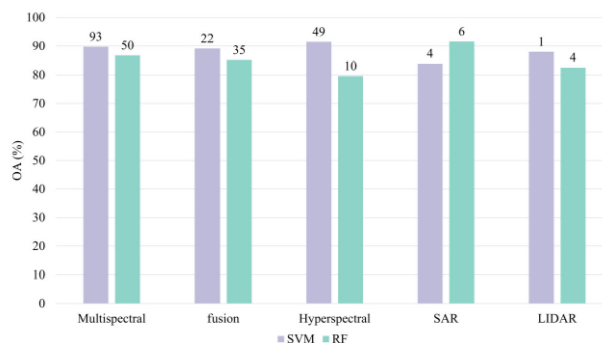Fig. 14. Software packages with tools/modules for the implementation of RF and SVM methods.

Fig. 15. Distribution of overall accuracies for different remotely sensed data (the numbers on top of the bars show the paper frequencies).

Weka data mining software tool (developed by Machine Learning Group at the University of Waikato, New Zealand). A detailed search through the literature showed that the free statistical software tool R appears to be the most important and frequent source of SVM (25%) and RF (41%) implementation. R is a programming language and free software environment for statistical computing and graphics and widely used for data analysis and development of statistical software. Most of the implementations in R were carried out using the caret package [141], which provides a standard syntax to execute a variety of machine learning methods, and e1071 package [142], which is the first implementation of SVM method in R. The dominance of statistical and data analysis software especially R, Python (scikit-learn package), and MATLAB is mainly because of the flexibility of these interfaces in dealing with extensive machine learning frameworks such as image preprocessing, feature selection and extraction, resampling methods, parameter tuning, training data balancing, and classification accuracy comparisons. In terms of the commercial software, for RF, eCognition is the most popular one, with about 17% of the case studies, while for the SVM classification method, ENVI is the most frequent software accounted for 9% of the studies.

C. Classification Performance and Data Type

Considering the optimal configuration for the available datasets, Fig. 15 shows the average accuracies based on the type

of remotely sensed data for both the SVM and RF methods. Multispectral RS images remain undoubtedly the most frequently employed data source amongst those utilized in both SVM and RF scenarios, with about 50% of the total papers mainly involving Landsat archives followed by MODIS imagery. On the other hand, the least data usage is related to LIDAR data, with less than 2% of the whole database. The percentages of the remaining types of RS data for SVM are as follows: fusion (21%), hyperspectral (21%), SAR (4%), and LIDAR (1.6%). These percentages for RF classifiers are about 32%, 10%, 6%, and 4%, respectively. In working with the multispectral and hyperspectral datasets, SVM gets the most attention. However, when it comes to SAR, LIDAR, and fusion of different sensors, RF is the focal point. The mean classification accuracy of SVM remains higher than the RF method in all sensor types except in SAR image data. Although it does not mean that in those cases, the SVM method works better than RF, it gives a hint about the range of accuracies that might be reached when using the SVM or RF. Moreover, it can be observed that, except for hyperspectral data in case of using the RF method, the mean classification accuracies are generally more than 82%. For SVM, the mean classification accuracy of hyperspectral datasets remains the highest at 91.5%, followed by multispectral (89.7%), fusion (89.14%), LIDAR (88.0%), and SAR (83.9%). This order for the RF method goes as SAR, multispectral, fusion, LIDAR, and hyperspectral with the mean overall accuracies of 91.60%, 86.74%, 85.12%, 82.55%, and 79.59%, respectively.

D. Classification Performance and Remote Sensing Applications

The number of articles focusing on different study targets (the number of studies is shown in the parenthesis) alongside the statistical analyses for each method is shown in Fig. 16. Other types of study targets, including soil, forest, water, mine, and cloud (comprising less than 10% of the total studies), were very few and are not shown here individually. The statistical analysis was conducted to assess OA (%) values that SVM and RF classifiers achieved for seven types of classification tasks. As shown in Fig. 16, most studies focused on LULC classification, crop classification, and urban studies with 50%, 14%, and 11% for SVM and 27%, 17%, and 14% for RF classifier. For LULC studies, the papers mostly adopted the publicly available benchmark datasets, as the main focus was on hyperspectral image classification. The most used datasets were from AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) and ROSIS (Reflective Optics System Imaging Spectrometer) hyperspectral sensors. For crop classification, the mainly used data was AVIRIS, followed by MODIS imageries. While in urban studies, Worldview-2 and IKONOS satellite imageries were the most frequently employed data. On the other hand, other studies mainly focused on the nonpublic image dataset for the region under study based on the application. Therefore, there is a satisfying number of studies that have focused on the real-world RS applications of both SVM and RF classifiers.

The assessment of classification accuracy regarding the types of study targets shows the maximum average accuracy in case
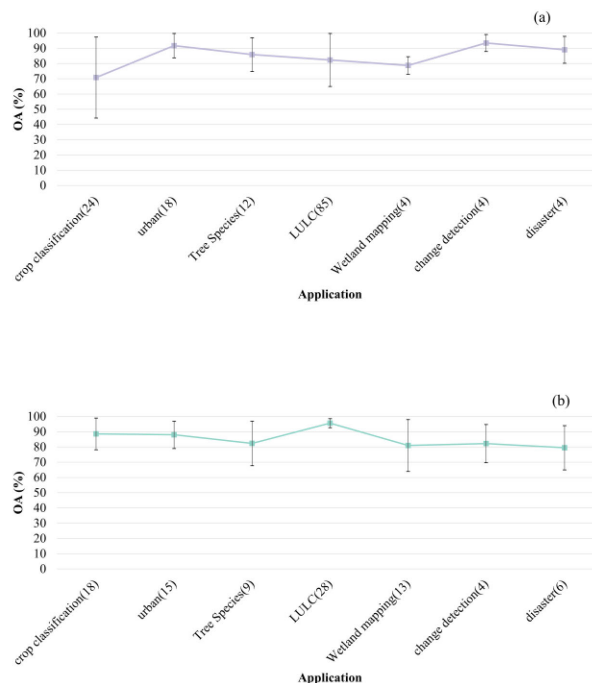
Fig. 16. Overall accuracies distribution of (a) SVM and (b) RF classifiers in different applications.
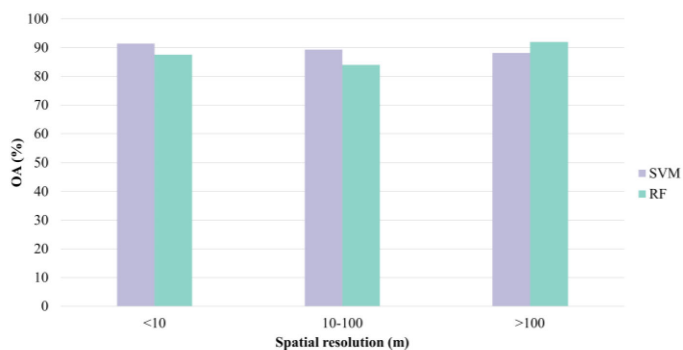
Fig. 17. Frequency and average accuracy of SVM and RF by the spatial resolution.

of using RF for LULC with approximately 95.5% and change detection with about 93.5% for SVM classification. LULC, as a mostly used application in both SVM and RF scenarios, shows little variability for the RF classifier. It can be interpreted as the higher stability of the RF method than SVM in the case of classification. The same manner is going on in crop classification tasks (i.e., the higher average accuracy and little variability for RF compared to the SVM method). The minimum amounts of average accuracies are also related to disaster-related applications and crop classification tasks for RF and SVM, respectively.

E. Classification Performance and Spatial Resolution

Fig. 17 shows the average obtained accuracy using RFs and SVMs based on the spatial resolution of the employed image data and their equivalent number of published papers. The papers were categorized based on the spatial resolution into high (<10

Fig. 18. Frequency and average accuracy of SVM and RF by the number of features.

m), medium (between 10 and 100 m), and low (>100 m). As seen, a relatively high number of papers (∼49%) were dedicated to the image data with spatial resolution in the range of 10–100 m. Datasets with a spatial resolution of smaller than 10 m also contribute to the high number of papers wit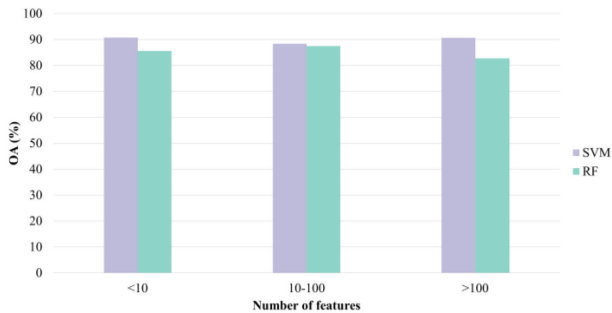h about 44% of the database. The remaining papers (∼7% of the database) deployed the image data with a spatial resolution of more than 100 m. The image data with high, medium, and low resolutions comprise 41%, 52%, and 7% of the studies for the SVM method and 52%, 44%, and 4% of the papers while using the RF method. Therefore, in the case of using high spatial resolution RS imagery, RF remains the most employed method, while in the case of medium- and low-resolution images, SVMs are the most commonly used method. A relatively robust trend is observed for the SVM method as the image data pixel size increases the average accuracy decreases. Such a trend can be observed for RF regarding that OA versus resolution is inconsistent among the three categories. Hence, in the case of the RF classifier, it is not possible to get a direct relationship between the classification accuracy and the resolution. However, in the high and medium resolution scenarios, the SVM method shows the higher average OAs.

F. Classification Performance and the Number of Extracted Features

Frequency and the average accuracies of SVMs and RFs versus the number of features are presented in Fig. 18. The papers were split into three groups (<10, between 10 and 100, and >100) based on the employed number of features. As can be seen, the number of papers for the SVM method is relatively the same for the three groups. However, in the case of RF, the vast majority of published papers (over than 60% of the total papers) focused on using 10 to 100 features, whereas a smaller number of papers used the number of features less than 10 or higher than 100, i.e., 22% and 18%, respectively. The comparison of the average accuracies of RF and SVM methods shows that the SVM method is reported to have higher accuracy. However, because of the inconsistency of the average accuracies for both SVM and RF, it is not possible to predict a linear relationship between the number of features and acquired accuracies. For the SVM method, the highest reported average accuracy is when using the lowest (<10) and highest (>100) number of features, but

RF shows the highest average accuracy while the number of features is between 10 and 100.

Fig. 19 displays the scatterplot of the sample data for pairwise comparison of RF and SVM classifiers. This figure illustrates the distribution of the overall accuracies and indicates the number of articles where one classifier works better than another. It can also help to interpret the magnitude of improvement for each sample article while considering the other classifier's accuracy. To further inform the readers, we marked the cases with different ranges of the number of features by three shapes (square, circle, and diamond), and with different sizes. The bigger size indicates the lower spatial resolution, i.e., bigger pixel size.

Moreover, to analyze the sensitivity of the algorithms to the number of classes, a colormap was used in which the brighter (more brownish) color shows the lower number of classes, and as the number of classes increases, the color goes toward the dark colors. The primary conclusion that is observed from the scatter plot is that most of the points are near the line 1:1, which shows somewhat similar behavior of the classifiers. However, in general, there are 32 papers with the implementation of both classifiers, and in 19 cases (about 60%), RF outperforms SVM, and in 40% of the remaining papers, SVM reports the higher accuracy.

Fig. 19 organizes the results based on the feature input dimensionality in three general categories by using different shapes. When it comes to the number of features, one can clearly observe that there is only one circle below the line 1:1, while the others are at the above. This shows the better performance of the SVM when the input data contains many more features. This result is in accordance with those exploited from Fig. 18. Conversely, in the case of features between 10 and 100, there are over 63% of the squares under the identity line (7 squares out of 11), which shows the high capability of RF while working with this group of image data. Finally, considering features fewer than 10, 58% of the diamonds are under the identity line (11 out of 17) and 42% above the line. Considering the fraction of RF supremacy over SVM 60–40, it is hard to infer which method performs better.

To examine the spatial resolution, three shape sizes were used. It is hardly possible to notice a specific pattern with respect to the pixel size of input data, but it is observed that most of the bigger shapes are under the one-to-one line. This represents the RF method offer consistently better results than SVM while dealing with images with bigger pixel size, which is in total accordance with Fig. 17. Looking at Fig. 19 and the distribution of points regarding the colors, the darker colors tend to offer more abundance under the identity line, while the points above the reference line are brighter. This can be proof of the efficiency of the SVM classifier to work with data with a lower number of classes. Statistically, 77% of the papers in which SVM shows higher accuracy include input data with the number of classes less than or equal to 6. The mean number of classes, in this case, is ∼5.5, whereas the mean number of classes in which RF works better is 8.4.

V. RECOMMENDATIONS AND FUTURE PROSPECT

Model selection can be used to determine a single best model, thus lending assistance to the one particular learner, or it can
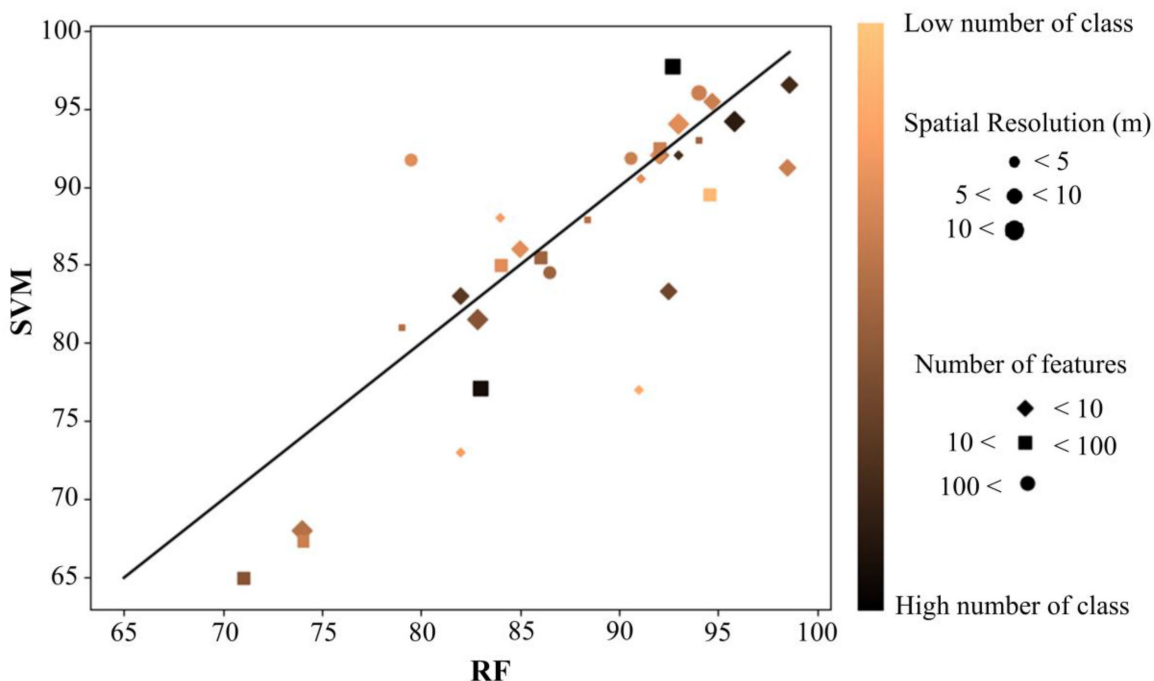
Fig. 19.   Comparison of the overall accuracy of SVM and RF classifiers by spatial resolution, number of features, and the number of classes in the under-study dataset.

be used to make inferences based on weighted support from a complete set of competing models. After a better understanding of the trends, strengths, and limitations of RFs and SVMs in the current study, the possibility of integrating two or more algorithms to solve a problem should be more investigated where the goal should be to utilize the strengths of one method to complement the weaknesses of another. If we are only interested in the best possible classification accuracy, it might be difficult or impossible to find a single classifier that performs as well as an excellent ensemble of classifiers. Mechanisms that are used to build the ensemble of classifiers, including using different subsets of training data with a single learning method, using different training parameters with a single training method, and using different learning methods, should be further investigated. In this sense, researchers may consider multiple variation of nearest neighbor techniques (e.g., K-NN) along with RF and SVM for both prediction and mapping.

Both SVM and RF are pixel-wise spectral classifiers. In other words, these classifiers do not consider the spatial dependences of adjacent pixels (i.e., spatial and contextual information). The availability of remotely sensed images with the fine spatial resolution has revolutionized image classification techniques by taking advantage of both spectral and spatial information in a unified classification framework [143], [144]. Object-based and spectral-spatial image classification using SVM and RF are regarded as a vibrant field of research within the RS community with lots of potential for further investigation.

The successful use of RF or SVM, coupled with a feature extraction approach to model a machine learning framework has been demonstrated intensively in the literature [145], [146]. The development of such machine learning techniques, which include a feature extraction approach followed by RF or SVM

as the classifier, is indeed a vibrant research line, which deserves further investigation.

As discussed in this study, SVM and RF can appropriately handle the challenging cases of the high-dimensionality of the input data, the limited number of training samples, and data heterogeneity. These advantages make SVM and RF well-suited for multisensor data fusion to further improve the classification performance of a single sensor data source [147], [148]. Due to the recent and sharp increase in the availability of data captured by various sensors, future studies will investigate SVM and RF more for these essential applications.

Several other ensemble machine learning toolboxes exist for different programming languages. The most widely used ones are scikit-learn [149] for Python, Weka [150] for Java, and mlj [151] for Julia. The most important toolboxes for R are mlr, caret [141], and tidy models [152]. Most recently, the mlr3 [153] package has become available for complex multistage experiments with advanced functionality that use a broad range of machine learning functionality. The authors of this study also suggest that researchers explore the diverse functionality of this package in their studies. We also would like to invite researchers to report their variable importance, class imbalance, class homogeneity, and sampling strategy and design in their studies. More importantly, to avoid spatial autocorrelation, if possible, samples for training should be selected randomly and spatially uncorrelated; purposeful samples should be avoided. For benchmark datasets, this study recommends researchers to use the standard sets of training and test samples already separated by the corresponding data provider. In this way, the classification performance of different approaches becomes comparable. To increase the fairness of the evaluation, RS societies have been developing evaluation leaderboards where researchers

can upload their classification maps and obtain classification accuracies on (usually) nondisclosed test data. One example of those evaluation websites is the IEEE Geoscience and Remote Sensing Society Data and Algorithm Standard Evaluation website (http://dase.grss-ieee.org/).

VI. CONCLUSION

The development of RS technology alongside with introducing new advanced classification algorithms have attracted the attention of researchers from different disciplines to utilize these potential data and tools for thematic applications. The choice of the most appropriate image classification algorithm is one of the hot topics in many research fields that deploy images of a vast range of spatial and spectral resolutions taken by different platforms while including many limitations and priorities. RF and SVM, as well-known and top-ranked machine learning algorithms, have gained the researchers' and analysts' attention in the field of data science and machine learning. Since there is ongoing employment of these methods in different disciplines, the common question is which method is highlighted based on the task properties. This article focused on a meta-analysis of comparison of peer-reviewed studies on RF and SVM classifiers. Our research aimed to statistically quantify the characteristics of these methods in terms of frequency and accuracy. The meta-analysis was conducted to serve as a descriptive and quantitative method of comparison using a database containing 251 eligible papers in which 42% and 68% of the database include RF and SVM implementation, respectively. The surveying carried out in the database showed the following.

1) The higher number of studies focusing on the SVM classifier is mainly due to the fact that it was introduced years before RF to the RS community. As can be concluded from the articles database, in the past three years, implementing the RF exceeded the SVM method. Nevertheless, in general, there is still an ongoing interest in using RF and SVM as standard classifiers in various applications.

2) The survey in the database revealed a moderate increase in using RF and SVM worldwide in an extensive range of applications such as urban studies, crop mapping, and particularly LULC applications, which got the highest average accuracy among all the applications. Although the assessment of the classification accuracies based on the application showed rather high variations for both RF and SVM, the results can be used for method selection concerning the application. For instance, the relatively high average accuracy and the little variance of the RF method for LULC applications can be interpreted as the superiority of RFs over SVMs in this field.

3) Medium and high spatial resolution images are the most used imageries for SVM and RF, respectively. It is hardly possible to notice a specific pattern concerning the spatial resolution of the input data. In the case of low spatial resolution images, the RF method offers consistently better results than SVM, although the number of papers using SVM for low spatial resolution image classification exceeded the RF method.

4) There is not a strong correlation between the acquired accuracies and the number of features for both the SVM and RF methods. However, a comparison of the average accuracies of RF and SVM methods suggests the superiority of the SVM method while classifying data containing many more features.

Contrary to the dominant utilization of SVM for the classification of hyperspectral and multispectral images, RF gets the attention while working with the fused datasets. For SAR and LIDAR datasets, the RF was also used more than the SVM method. However, its popularity cannot be concluded because of the low available number of published papers.

REFERENCES

[1] C. Toth and G. Jóźków, "Remote sensing platforms and sensors: A survey," *ISPRS J. Photogramm. Remote Sens.*, vol. 115, pp. 22–36, 2016.

[2] M. Wieland, W. Liu, and F. Yamazaki, "Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes," *Remote Sens.*, vol. 8, no. 10, Sep. 2016, Art. no. 792.

[3] K. Wessels *et al.*, "Rapid land cover map updates using change detection and robust random forest classifiers," *Remote Sens.*, vol. 8, no. 11, Oct. 2016, Art. no. 888.

[4] F. Traoré *et al.*, "Using multi-temporal landsat images and support vector machine to assess the changes in agricultural irrigated areas in the Mogtedo region, Burkina Faso," *Remote Sens.*, vol. 11, no. 12, Jun. 2019, Art. no. 1442.

[5] D. Seo, Y. Kim, Y. Eo, W. Park, and H. Park, "Generation of radiometric, phenological normalized image based on random forest regression for change detection," *Remote Sens.*, vol. 9, no. 11, Nov. 2017, Art. no. 1163.

[6] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, 2020.

[7] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[8] M. Mahdianpari, B. Salehi, F. Mohammadimanesh, and M. Motagh, "Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 13–31, 2017.

[9] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, nos. 1/2, pp. 107–117, 2004.

[10] E. Adam, O. Mutanga, J. Odindi, and E. M. Abdel-Rahman, "Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers," *Int. J. Remote Sens.*, vol. 35, no. 10, pp. 3440–3458, May 2014.

[11] O. S. Ahmed, S. E. Franklin, M. A. Wulder, and J. C. White, "Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm," *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 89–101, Mar. 2015.

[12] M. Mahdianpari, B. Salehi, F. Mohammadimanesh, S. Homayouni, and E. Gill, "The first wetland inventory map of newfoundland at a spatial resolution of 10 m using Sentinel-1 and Sentinel-2 data on the Google Earth Engine cloud computing platform," *Remote Sens.*, vol. 11, no. 1, Dec. 2018, Art. no. 43.

[13] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[14] R. Ramo and E. Chuvieco, "Developing a random forest algorithm for MODIS global burned area classification," *Remote Sens.*, vol. 9, no. 11, Nov. 2017, Art. no. 1193.

[15] B. Heung, C. E. Bulmer, and M. G. Schmidt, "Predictive soil parent material mapping at a regional-scale: A Random Forest approach," *Geoderma*, vol. 214/215, pp. 141–154, 2014.

[16] T. Bai, K. Sun, S. Deng, D. Li, W. Li, and Y. Chen, "Multi-scale hierarchical sampling change detection using Random Forest for high-resolution satellite imagery," *Int. J. Remote Sens.*, vol. 39, no. 21, pp. 7523–7546, Nov. 2018.

[17] F. Bovolo, L. Bruzzone, and M. Marconcini, "a novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.

[18] G. Cao, Y. Li, Y. Liu, and Y. Shang, "Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification," *Int. J. Remote Sens.*, vol. 35, no. 16, pp. 6255–6270, Aug. 2014.

[19] J. J. Gapper, H. El-Askary, E. Linstead, and T. Piechota, "Coral reef change detection in remote pacific islands using support vector machine classifiers," *Remote Sens.*, vol. 11, no. 13, Jun. 2019, Art. no. 1525.

[20] S. K. Karan and S. R. Samadder, "Improving accuracy of long-term land-use change in coal mining areas using wavelets and Support Vector Machines," *Int. J. Remote Sens.*, vol. 39, no. 1, pp. 84–100, Jan. 2018.

[21] T. Bangira, S. M. Alfieri, M. Menenti, and A. van Niekerk, "Comparing thresholding with machine learning classifiers for mapping complex water," *Remote Sens.*, vol. 11, no. 11, Jun. 2019, Art. no. 1351.

[22] R. Pouteau and B. Stoll, "SVM selective fusion (SELF) for multi-source classification of structurally complex tropical rainforest," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 5, no. 4, pp. 1203–1212, Aug. 2012.

[23] L. I. Wilschut *et al.*, "Mapping the distribution of the main host for plague in a complex landscape in Kazakhstan: An object-based approach using SPOT-5 XS, Landsat 7 ETM+, SRTM and multiple Random Forests," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 23, pp. 81–94, Aug. 2013.

[24] A. Mathur and G. M. Foody, "Crop classification by support vector machine with intelligently selected training data for an operational application," *Int. J. Remote Sens.*, vol. 29, no. 8, pp. 2227–2240, Apr. 2008.

[25] K. Millard and M. Richardson, "On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping," *Remote Sens.*, vol. 7, no. 7, pp. 8489–8515, Jul. 2015.

[26] A.-L. Boulesteix, S. Janitza, and J. Kruppa, "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics," *Biologie Unserer Zeit*, vol. 22, no. 6, pp. 323–329, 2012.

[27] Y. Shao and R. S. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points," *ISPRS J. Photogramm. Remote Sens.*, vol. 70, pp. 78–87, 2012.

[28] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259–272, Mar. 2012.

[29] S. E. Jozdani, B. A. Johnson, and D. Chen, "Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification," *Remote Sens.*, vol. 11, no. 14, Jul. 2019, Art. no. 1713.

[30] P. Kumar, D. K. Gupta, V. N. Mishra, and R. Prasad, "Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data," *Int. J. Remote Sens.*, vol. 36, no. 6, pp. 1604–1617, Mar. 2015.

[31] D. Turner, A. Lucieer, Z. Malenovský, D. King, and S. A. Robinson, "Assessment of Antarctic moss health from multi-sensor UAS imagery with random forest modelling," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 68, pp. 168–179, Jun. 2018.

[32] S. Tian, X. Zhang, J. Tian, and Q. Sun, "Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China," *Remote Sens.*, vol. 8, no. 11, Nov. 2016, Art. no. 954.

[33] A. Mellor, S. Boukir, A. Haywood, and S. Jones, "Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 155–168, Jul. 2015.

[34] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.

[35] B. Demir and S. Erturk, "Clustering-based extraction of border training patterns for accurate SVM classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 840–844, Oct. 2009.

[36] G. M. Foody and A. Mathur, "The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM," *Remote Sens. Environ.*, vol. 103, no. 2, pp. 179–189, Jul. 2006.

[37] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.

[38] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 38–53, Jul. 2015.

[39] X. Huang and L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines," *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 1977–1987, Apr. 2009.

[40] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[41] X. Liu *et al.*, "Spatial-temporal patterns of features selected using random forests: A case study of corn and soybeans mapping in the US," *Int. J. Remote Sens.*, vol. 40, no. 1, pp. 269–283, Jan. 2019.

[42] D. Phiri, J. Morgenroth, C. Xu, and T. Hermosilla, "Effects of pre-processing methods on Landsat OLI-8 land cover classification using OBIA and random forests classifier," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 73, pp. 170–178, Dec. 2018.

[43] A. S. Sahadevan, A. Routray, B. S. Das, and S. Ahmad, "Hyperspectral image preprocessing with bilateral filter for improving the classification accuracy of support vector machines," *J. Appl. Remote Sens.*, vol. 10, no. 2, Apr. 2016, Art. no. 025004.

[44] J. M. Corcoran, J. F. Knight, and A. L. Gallant, "Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota," *Remote Sens.*, vol. 5, no. 7, pp. 3212–3238, 2013.

[45] E. M. Abdel-Rahman, O. Mutanga, E. Adam, and R. Ismail, "Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 48–59, Feb. 2014.

[46] M. Dihkan, N. Guneroglu, F. Karsli, and A. Guneroglu, "Remote sensing of tea plantations using an SVM classifier and pattern-based accuracy assessment technique," *Int. J. Remote Sens.*, vol. 34, no. 23, pp. 8549–8565, Dec. 2013.

[47] N. Kranjčić, D. Medak, R. Župan, and M. Rezo, "Support vector machine accuracy assessment for extracting green urban areas in towns," *Remote Sens.*, vol. 11, no. 6, Mar. 2019, Art. no. 655.

[48] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, Jul. 2018, Art. no. 1119.

[49] R. Taghizadeh-Mehrjardi *et al.*, "Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran," *Geoderma*, vol. 376, Oct. 2020, Art. no. 114552.

[50] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, 2002.

[51] F. F. Camargo, E. E. Sano, C. M. Almeida, J. C. Mura, and T. Almeida, "A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian Tropical Savanna using ALOS-2/PALSAR-2 polarimetric images," *Remote Sens.*, vol. 11, no. 13, Jul. 2019, Art. no. 1600.

[52] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, 2017.

[53] F. E. Fassnacht *et al.*, "Review of studies on tree species classification from remotely sensed data," *Remote Sens. Environ.*, vol. 186, pp. 64–87, 2016.

[54] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," *Remote Sens. Environ.*, vol. 158, pp. 295–310, 2015.

[55] G. Shukla, R. D. Garg, H. S. Srivastava, and P. K. Garg, "An effective implementation and assessment of a random forest classifier as a soil spatial predictive model," *Int. J. Remote Sens.*, vol. 39, no. 8, pp. 2637–2669, Apr. 2018.

[56] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.

[57] P. Liu, K.-K. R. Choo, L. Wang, and F. Huang, "SVM or deep learning? A comparative study on remote sensing image classification," *Soft Comput.*, vol. 21, no. 23, pp. 7053–7065, 2017.

[58] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[59] V. Heikkinen, I. Korpela, T. Tokola, E. Honkavaara, and J. Parkkinen, "An SVM classification of tree species radiometric signatures based on the Leica ADS40 sensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4539–4551, Nov. 2011.

[60] T. Habib, J. Inglada, G. Mercier, and J. Chanussot, "Support vector reduction in SVM algorithm for abrupt change detection in remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 606–610, Jul. 2009.

[61] F. Giacco, C. Thiel, L. Pugliese, S. Scarpetta, and M. Marinaro, "Uncertainty analysis for the classification of multispectral satellite images using SVMs and SOMs," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3769–3779, Oct. 2010.

[62] S. S. Heydari, "Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites," *Remote Sens. Environ.*, vol. 204, pp. 648–658, 2018.

[63] S. S. Heydari and G. Mountrakis, "Meta-analysis of deep neural networks in remote sensing: A comparative study of mono-temporal classification to support vector machines," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 192–210, Jun. 2019.

[64] N. Mboga, C. Persello, J. R. Bergado, and A. Stein, "Detection of informal settlements from VHR images using convolutional neural networks," *Remote Sens.*, vol. 9, 2017, Art. no. 1106.

[65] M. Sheykhmousa, N. Kerle, M. Kuffer, and S. Ghaffarian, "Post-disaster recovery assessment with machine learning-derived land cover and land use information," *Remote Sens.*, vol. 11, no. 10, 2019, Art. no. 1174.

[66] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.

[67] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[68] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data.* Boca Raton, FL, USA: CRC Press, 2016.

[69] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.

[70] A. Statnikov, L. Wang, and C. F. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," *BMC Bioinformat.*, vol. 9, no. 1, 2008, Art. no. 319.

[71] L. Yu *et al.*, "Meta-discoveries from a synthesis of satellite-based land-cover mapping research," *Int. J. Remote Sens.*, vol. 35, no. 13, pp. 4573–4588, 2014.

[72] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Cambridge, MA, USA: MIT Press, 2001.

[73] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.

[74] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, Nov. 2006.

[75] B.-C. Kuo, H.-H. Ho, C.-H. Li, C.-C. Hung, and J.-S. Taur, "A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 1, pp. 317–326, Jan. 2014.

[76] X. Huang, T. Hu, J. Li, Q. Wang, and J. A. Benediktsson, "Mapping urban areas in China using multisource data with a novel ensemble SVM method," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4258–4273, Aug. 2018.

[77] L. Wang, *Support Vector Machines: Theory and Applications*, vol. 177. New York, NY, USA: Springer, 2005.

[78] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[79] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 11, no. 5, pp. 352–359, Oct. 2009.

[80] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, 2004.

[81] S. Ali and K. A. Smith-Miles, "A meta-learning approach to automatic kernel selection for support vector machines," *Neurocomputing*, vol. 70, nos. 1–3, pp. 173–186, 2006.

[82] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1224–1228, Sep. 2013.

[83] R. Khemchandani and S. Chandra, "Optimal kernel selection in twin support vector machines," *Optim. Lett.*, vol. 3, no. 1, pp. 77–88, 2009.

[84] X. Zhu, N. Li, and Y. Pan, "Optimization performance comparison of three different group intelligence algorithms on a SVM for hyperspectral imagery classification," *Remote Sens.*, vol. 11, no. 6, Mar. 2019, Art. no. 734.

[85] L. Su, "Optimizing support vector machine learning for semi-arid vegetation mapping by using clustering analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 4, pp. 407–413, Jul. 2009.

[86] F. Samadzadegan, H. Hasani, and T. Schenk, "Simultaneous feature selection and SVM parameter determination in classification of hyperspectral imagery using ant colony optimization," *Can. J. Remote Sens.*, vol. 38, no. 2, pp. 139–156, 2012.

[87] Y. Gu and K. Feng, "Optimized Laplacian SVM with distance metric learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 6, no. 3, pp. 1109–1117, Jun. 2013.

[88] N. Ghoggali and F. Melgani, "Genetic SVM approach to semisupervised multitemporal classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 212–216, Apr. 2008.

[89] N. Ghoggali, F. Melgani, and Y. Bazi, "A Multiobjective genetic SVM approach for classification problems with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1707–1718, Jun. 2009.

[90] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," *J. Appl. Remote Sens.*, vol. 10, no. 3, Sep. 2016, Art. no. 035021.

[91] S. Ch, N. Anand, B. K. Panigrahi, and S. Mathur, "Streamflow forecasting by SVM with quantum behaved particle swarm optimization," *Neurocomputing*, vol. 101, pp. 18–23, 2013.

[92] J. Milgram, M. Cheriet, and R. Sabourin, "'One against one' or 'one against all': Which one is better for handwriting recognition with SVMs?," in *Proc. 10th Int. Workshop Front. Handwriting Recognit.*, 2006.

[93] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.

[94] H. Guo and W. Wang, "Granular support vector machine: A review," *Artif. Intell. Rev.*, vol. 51, no. 1, pp. 19–32, 2019.

[95] L. Bruzzone, M. Chi, and M. Marconcini, "A Novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[96] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, 3804–3814, 2008.

[97] D. Sun, Y. Li, and Q. Wang, "A unified model for remotely estimating chlorophyll a in Lake Taihu, China, based on SVM and *in situ* hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2957–2965, Aug. 2009.

[98] Y. Chen, X. Zhao, and Z. Lin, "Optimizing subspace SVM ensemble for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 7, no. 4, pp. 1295–1305, Apr. 2014.

[99] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[100] Y. Wu *et al.*, "Approximate computing of remotely sensed data: SVM hyperspectral image classification as a case study," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 9, no. 12, pp. 5806–5818, Dec. 2016.

[101] C. Bo, H. Lu, and D. Wang, "Hyperspectral image classification via JCR and SVM models with decision fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 177–181, Feb. 2016.

[102] L. Breiman, "ST4_Method_Random_Forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[103] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: From early developments to recent advancements," *Syst. Sci. Control Eng. Open Access J.*, vol. 2, no. 1, pp. 602–609, 2014.

[104] S. Banks *et al.*, "Wetland classification with multi-angle/temporal SAR using random forests," *Remote Sens.*, vol. 11, no. 6, Mar. 2019, Art. no. 670.

[105] Z. Bassa, U. Bob, Z. Szantoi, and R. Ismail, "Land cover and land use mapping of the iSimangaliso Wetland Park, South Africa: Comparison of oblique and orthogonal random forest algorithms," *J. Appl. Remote Sens.*, vol. 10, no. 1, Mar. 2016, Art. no. 015017.

[106] L. Lam and S. Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Trans. Syst. Man Cybern. A, Syst. Humans*, vol. 27, no. 5, pp. 553–568, Sep. 1997.

[107] R. K. Shahzad and N. Lavesson, "Comparative analysis of voting schemes for ensemble-based malware detection," *J. Wireless Mobile Netw. Ubiquitous Comput. Dependable Appl.*, vol. 4, no. 1, pp. 98–117, 2013.

[108] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learning Theory*, 1995, pp. 23–37.

[109] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[110] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, 1995, vol. 1, pp. 278–282.

[111] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[112] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, 1997.

[113] J. M. Klusowski, "Complete analysis of a random forest model," vol. 13, pp. 1063–1095, 2018, *arXiv:1805.02587.*

[114] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[115] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[116] R. Genuer, J. M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, 2010.

[117] Ö. Akar and O. Güngör, "Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey," *Int. J. Remote Sens.*, vol. 36, no. 2, pp. 442–464, Jan. 2015.

[118] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010.

[119] H. Ishida, Y. Oishi, K. Morita, K. Moriwaki, and T. Y. Nakajima, "Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions," *Remote Sens. Environ.*, vol. 205, pp. 390–407, Feb. 2018.

[120] E. Izquierdo-Verdiguier, V. Laparra, L. Gomez-Chova, and G. Camps-Valls, "Encoding invariances in remote sensing image classification with SVM," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 981–985, Sep. 2013.

[121] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformat.*, vol. 9, no. 1, Dec. 2008, Art. no. 307.

[122] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, 2011.

[123] A. Behnamian, K. Millard, S. N. Banks, L. White, M. Richardson, and J. Pasher, "A systematic approach for variable selection with random forests: Achieving stable variable importance values," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1988–1992, Nov. 2017.

[124] X. Shen and L. Cao, "Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data," *Remote Sens.*, vol. 9, no. 11, Nov. 2017, Art. no. 1180.

[125] P. S. Gromski, Y. Xu, E. Correa, D. I. Ellis, M. L. Turner, and R. Goodacre, "A comparative investigation of modern feature selection and classification approaches for the analysis of mass spectrometry data," *Anal. Chim. Acta*, vol. 829, pp. 1–8, Jun. 2014.

[126] T. Berhane *et al.*, "Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 580.

[127] H. Guan, J. Li, M. Chapman, F. Deng, Z. Ji, and X. Yang, "Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests," *Int. J. Remote Sens.*, vol. 34, no. 14, pp. 5166–5186, 2013.

[128] P. Probst, M. N. Wright, and A. L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *Wiley Interdisciplinary Rev. Data Mining Knowl. Discov.*, vol. 9, no. 3, pp. 1–15, 2019.

[129] 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, 2008.

[130] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.

[131] C. Strobl, A. L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformat.*, vol. 8, 2007, Art. no. 25.

[132] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, "Robust and accurate shape model matching using random forest regression-Voting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1862–1874, Sep. 2015.

[133] M. Mahdianpari *et al.*, "Big Data for a big country: The first generation of canadian wetland inventory map at a spatial resolution of 10-m using Sentinel-1 and Sentinel-2 data on the google earth engine cloud computing platform," *Can. J. Remote Sens.*, vol. 46, pp. 15–33, 2020.

[134] M. N. Wright and A. Ziegler, "Ranger: A fast implementation of random forests for high dimensional data in C++ and R," *J. Statist. Softw.*, vol. 77, no. 1, pp. 1–17, 2017.

[135] J. S. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[136] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[137] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[138] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Ann. Internal Med.*, vol. 151, no. 4, pp. 264–269, 2009.

[139] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, 2002.

[140] S. S. Heydari and G. Mountrakis, "Meta-analysis of deep neural networks in remote sensing: A comparative study of mono-temporal classification to support vector machines," *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 192–210, 2019.

[141] M. Kuhn, "Caret: Classification and regression training," *ascl*, ascl-1505, 2015.

[142] D. Meyer *et al.*, "Package 'e1071'," *R J.*, 2019.

[143] P. Ghamisi *et al.*, "New frontiers in spectral-spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.

[144] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[145] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[146] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[147] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[148] P. Ghamisi and B. Hoefle, "LiDAR data classification using extinction profiles and a composite kernel support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 659–663, May 2017.

[149] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.

[150] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Exploration Newslett.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

[151] A. Blaom, F. Kiraly, T. Lienart, and S. Vollmer, Alan-turing-institute/MLJ.jl: v0.5.3 (Version v0.5.3). Zenodo. Nov. 14, 2019. [Online]. Available: http://doi.org/10.5281/zenodo.3541506

[152] H. Wickham *et al.*, "Welcome to the Tidyverse," *J. Open Source Softw.*, vol. 4, no. 43, Nov. 2019, Art. no. 1686.

[153] M. Lang *et al.*, "mlr3: A modern object-oriented machine learning framework in R," *J. Open Source Softw.*, vol. 4, no. 44, 2019, Art. no. 1903.

Mohammadreza Sheykhmousa received the M.Sc. degree in geoinformation science and earth observation from the University of Twente (UT), Enschede, The Netherlands, in 2018.

He designed and implement a novel approach to assess post-disaster recovery using satellite imagery and machine learning, analyzed results of the experiments, and wrote an original article for his M.Sc. thesis. He is currently working as a Researcher with OpenGeoHub (a not-for-profit research foundation), Wageningen, The Netherlands. His research interests include ensemble machine learning, (geo)data analysis, image processing, disaster recovery, and predictive mapping.

Masoud Mahdianpari (Member, IEEE) received the B.S. degree in surveying and geomatics engineering and the M.Sc. degree in remote sensing engineering from the School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran, in 2010 and 2013, respectively, and the Ph.D. degree in electrical engineering from the Department of Engineering and Applied Science, Memorial University, St. John's, NL, Canada, in 2019.

He was an Ocean Frontier Institute (OFI) Postdoctoral Fellow with Memorial University and C-CORE, in 2019. He is currently a Remote Sensing Technical Lead with C-CORE and a Cross-Appointed Professor with the Faculty of Engineering and Applied Science, Memorial University. His research interests include remote sensing and image analysis, with a special focus on PolSAR image processing, multisensor data classification, machine learning, geo big data, and deep learning.

Dr. Mahdianpari was the recipient of the Top Student Award from the University of Tehran for four consecutive academic years from 2010 to 2013, and the Research and Development Corporation Ocean Industries Student Research Award, organized by the Newfoundland industry and innovation center, amongst more than 400 submissions, in 2016. In 2016, he won T. David Collett Best Industry Paper Award organized by IEEE. He also received The Com-Adv Devices, Inc., Scholarship for Innovation, Creativity, and Entrepreneurship from the Memorial University of Newfoundland, in 2017, the Artificial Intelligence for Earth Grant organized by Microsoft, in 2018, and the Graduate Academic Excellence Award organized by Memorial University, in 2019. He currently serves as an editorial team member for the Remote Sensing Journal and the Canadian Journal of Remote Sensing.

Hamid Ghanbari received the B.Sc. degree in geomatics engineering and the master's degree in remote sensing from the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, in 2014 and 2016, respectively. He is currently working toward the Ph.D. degree with the Department of Geography, Université Laval, Québec, QC, Canada.

His research interests include hyperspectral image classification, multitemporal and multisensor image analysis, and machine learning.

Fariba Mohammadimanesh received the B.Sc. degree in surveying and geomatics engineering and the M.Sc. degree in remote sensing engineering from the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, in 2010 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Department of Engineering and Applied Science, Memorial University, St. John's, NL, Canada, in 2019.

She is currently a Remote Sensing Scientist with C-CORE's Earth Observation Team, St. John's, NL, Canada. Her research interests include image processing, machine learning, segmentation, and classification of satellite remote sensing data, including SAR, PolSAR, and optical, in environmental studies, with a special interest in wetland mapping and monitoring. Another area of her research includes geo-hazard mapping using the interferometric SAR technique (InSAR).

Dr. Mohammadimanesh was the recipient of several awards, including the Emera Graduate Scholarship (2016–2019), CIG-NL award (2018), and the IEEE NECEC Best Industry Paper Award (2016).
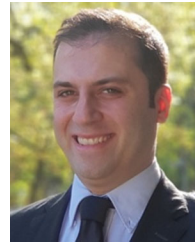
Pedram Ghamisi (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2015.

He works as the Head of the machine learning group with Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Freiburg, Germany, and as the CTO and co-founder of VasoGnosis Inc., USA. He is also the Co-Chair of IEEE Image Analysis and Data Fusion Committee. His research interests include interdisciplinary research on machine (deep) learning, image and signal processing, and multisensor data fusion.

Dr. Ghamisi was the recipient of the Best Researcher Award for M.Sc. students at the K. N. Toosi University of Technology in the academic year of 2010–2011, the IEEE Mikio Takagi Prize for winning the Student Paper Competition at IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2013, the Talented International Researcher by Iran's National Elites Foundation in 2016, the first prize of the data fusion contest organized by the Image Analysis and Data Fusion Technical Committee (IADF) of IEEE-GRSS, in 2017, the Best Reviewer Prize of IEEE Geoscience and Remote Sensing Letters (GRSL) in 2017, the IEEE Geoscience and Remote Sensing Society 2019 Highest Impact Paper Award, the Alexander von Humboldt Fellowship from the Technical University of Munich, and the High Potential Program Award from HZDR. He serves as an Associate Editor for MDPI-Remote Sensing, MDPI-Sensor, and IEEE GRSL.

Saeid Homayouni (Senior Member, IEEE) received the B.Sc. degree in surveying and geomatics engineering from the University of Isfahan, Isfahan, Iran, in 1996, the M.Sc. degree in remote sensing and geographic information systems from Tarbiat Modares University, Tehran, Iran, in 1999, and the Ph.D. degree in signal and image from Télécom Paris Tech, Paris, France, in 2005.

From 2006 to 2007, he was a Postdoctoral Fellow with the Signal and Image Laboratory, University of Bordeaux Agro-Science, Bordeaux, France. From 2008 to 2011, he was an Assistant Professor with the Department of Surveying and Geomatics, College of Engineering, University of Tehran, Tehran. From 2011 to 2013, through the Natural Sciences and Engineering Research Council of Canada (NSERC) visitor fellowship program, he worked with the Earth Observation Group of the Agriculture and Agri-Food Canada (AAFC) at the Ottawa Center of Research and Development. In 2013, he joined the Department of Geography, Environment, and Geomatics, University of Ottawa, Ottawa, as a Replacing Assistant Professor of remote sensing and geographic information systems. Since April 2019, he works as an Associate Professor of environmental remote sensing and geomatics with the Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Quebec, QC, Canada. His research interests include optical and radar earth observations analytics for urban and agro-environmental applications.