# Machine learning based hyperspectral image analysis: A survey

Utsav B. Gewali[1], Sildomar T. Monteiro[2], and Eli Saber[1,3]

[1]Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, Email: ubg9540@rit.edu
[2]Boeing, Huntsville, AL
[3]Department of Electrical & Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY

## Abstract

Hyperspectral sensors enable the study of the chemical properties of scene materials remotely for the purpose of identification, detection, and chemical composition analysis of objects in the environment. Hence, hyperspectral images captured from earth observing satellites and aircraft have been increasingly important in agriculture, environmental monitoring, urban planning, mining, and defense. Machine learning algorithms due to their outstanding predictive power have become a key tool for modern hyperspectral image analysis. Therefore, a solid understanding of machine learning techniques have become essential for remote sensing researchers and practitioners. This paper reviews and compares recent machine learning-based hyperspectral image analysis methods published in literature. We organize the methods by the image analysis task and by the type of machine learning algorithm, and present a two-way mapping between the image analysis tasks and the types of machine learning algorithms that can be applied to them. The paper is comprehensive in coverage of both hyperspectral image analysis tasks and machine learning algorithms. The image analysis tasks considered are land cover classification, target detection, unmixing, and physical parameter estimation. The machine learning algorithms covered are Gaussian models, linear regression, logistic regression, support vector machines, Gaussian mixture model, latent linear models, sparse linear models, Gaussian mixture models, ensemble learning, directed graphical models, undirected graphical models, clustering, Gaussian processes, Dirichlet processes, and deep learning. We also discuss the open challenges in the field of hyperspectral image analysis and explore possible future directions.

***Index terms***— Hyperspectral image analysis; machine learning; imaging spectroscopy; A review

## 1   Introduction

With applications to fields such as agriculture [67], ecology [295], mining [284], forestry [106], urban planning [300], defense [312], and space exploration [235], hyperspectral imaging is a powerful remote sensing modality to study the chemical properties of scene materials. Hyperspectral imaging, also known as imaging spectroscopy, captures the reflected or emitted electromagnetic energy from a scene over hundreds of narrow, contiguous spectral bands, from visible to infrared wavelengths [88]. Each pixel in a hyperspectral image is composed of a vector of hundreds of elements measuring the reflected or emitted energy as a function of wavelength, known as the spectrum. Hence, a hyperspectral image can be interpreted as a three dimensional data structure with two spatial axes, carrying the information about the location of objects, and one spectral axis, carrying the information about the objects' chemical composition. The spectrum captures the chemical information because the interaction between light at different wavelengths and the material is governed by material's atomic or molecular structure. Hyperspectral sensors mounted on aircrafts and satellites can image and map the land cover and land use, detect and locate objects, or understand the physical properties of materials over a large geographical area.

Since hyperspectral images contain hundreds of bands, not just three bands as in color photography, it is rarely analyzed by visual inspection, instead algorithms are developed to extract meaningful information from the images. Machine learning and pattern recognition based methods have been very successful for this purpose, as they are able to automatically learn the relationship between the spectrum captured at each pixel of the image and the information that is desired to be extracted. They are also more robust in handling the noise and uncertainties in the measurements than traditional methods, such

as manually designed normalized indices and physics-based models. The remote sensing community has shown a great deal of interest in machine learning recently. Many journals have published special issues on machine learning for remote sensing [280, 60, 5, 36], numerous articles have been published on the topic of rise of machine learning in remotes sensing [35, 158], and all of the winning methods of the recent annual remote sensing GRSS data fusion competition [72, 180, 209, 281] and the top performing methods on ISPRS benchmark tests [1] have been based on machine learning. The hyperspectral remote sensing community has been particularly active in this field and produced a great number of new methods.

This survey paper aims to provide a broad coverage of both the hyperspectral image analysis tasks and the machine learning algorithms, unlike previous surveys and tutorials which have either focused on a task [29, 184, 191, 39, 57] or a particular machine learning algorithm [40, 318, 211, 24]. All the methods reviewed here were published in peer-reviewed journals. The reviewed methods are able to analyze both radiance and reflectance images, unless otherwise stated. The hyperspectral data analysis tasks are categorized as land cover classification [39], target detection [222], unmixing [29]and physical parameter estimation [278]. The machine learning algorithms covered are Gaussian models [277], linear regression [208], logistic regression [199], support vector machines [257], Gaussian mixture models [28], latent linear models [147], sparse linear models [189], ensemble learning [246], directed graphical models [292], undirected graphical models [31], clustering [144], Gaussian processes [301] , Dirichlet processes [274], and deep learning [26].

The main contributions of this paper are: *a)* extensive review of recently published hyperspectral analysis methods, *b)* categorization of each method by remote sensing task and machine learning algorithm (which is neatly summarized in Table 1), and *c)* exploration of current trends and problems along with future directions.

This paper is organized as follows. Section 2 discusses about various types of analysis that can be performed on hyperspectral data, Section 3 provides a brief background on common machine learning techniques and terminologies, Section 4 explains the problem of high dimensionality of hyperspectral data, and Section 5 reviews various machine learning-based hyperspectral image analysis methods found in recent literature. Section 6 discusses open challenges in the field of hyperspectral data analysis, and Section 7 summarizes the paper.

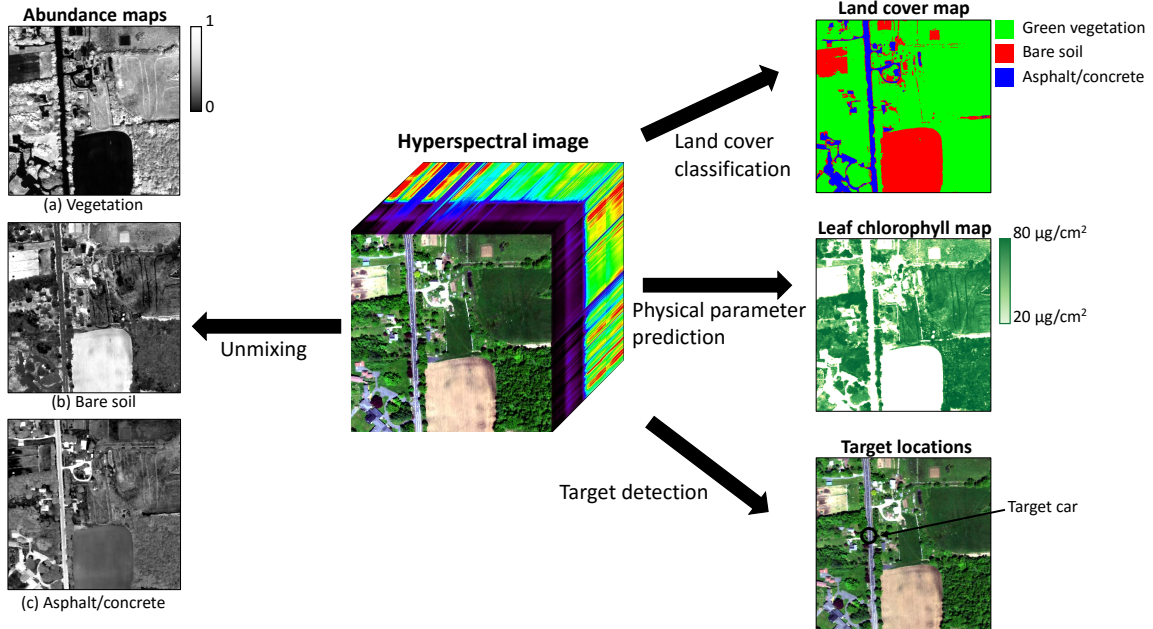# 2 Taxonomy of hyperspectral data analysis tasks



Figure 1: Hyperspectral image analysis tasks.[1]

---

[1] The image obtained from [2].

We discuss methods developed for reflective hyperspectral remote sensing in this survey. The reflective regime contains the visible, the near infrared, and the shortwave infrared wavelengths (350 nm to 2500 nm) and the radiance reaching the sensor is dominated by the solar energy reflected off the objects in the scene. The material property that determines the magnitude of the reflected radiance is the directional reflectance of the object. The radiance reaching the sensor however contains contributions due to the scatterings from the atmosphere, which can be removed using atmospheric compensation algorithms [101] to estimate surface reflectance. Hence, the pixels in reflective hyperspectral image are measured in radiance or reflectance. The reflectance image is more preferred for hyperspectral image analysis tasks because reflectance is a surface property and reflectance image generally tends to produce better results as the atmospheric interference is reduced. The majority of algorithms discussed in this paper are agnostic to the units (radiance or reflectance) the image is measured. However, many methods required the image and the ground truth to be converted to same units before application.

The data analysis tasks for reflective hyperspectral images can be divided into four distinct groups: land cover classification, target detection, spectral unmixing, and physical parameter estimation, as shown in Figure 1. These are further divided into sub-tasks.

## 2.1 Land cover classification

Land cover classification [39], also called land cover mapping and land cover segmentation, is the process of identifying the material under each pixel of a hyperspectral image. The goal is to create a map showing how different materials are distributed over a geographical area imaged by the hyperspectral sensor. Common applications of land cover mapping are plant species classification [69], urban scene classification [74], mineral identification [218], and change analysis [238]. The main advantage of using hyperspectral images to produce land cover maps is that hyperspectral images allow for the discrimination of land covers into finer class compared to other modalities, such as multispectral images, because hyperspectral images capture more information about the chemistry of the materials.

Many land cover mapping methods require a prior knowledge about the types of materials present in the scene along with examples of the spectra belonging to them. This information is generally provided from the image pixels by an expert, collected using a spectrometer during field campaign of the study area, or adapted from a third party spectral library. However, there are also many land cover mapping techniques that require no prior information about the materials in the scene.

Change detection [136] is the task of determining how the land cover and the land use of an area has changed over time by using hyperspectral images of that area taken at different times. Applications of change detection include urban growth mapping [223], forest monitoring [77], wetland degradation study [253], invasive vegetation growth monitoring [238], and post-fire vegetation regeneration mapping [242]. Many times, change detection study is performed by comparing a set of land cover maps or fractional land cover maps an area generated by applying land cover classification or unmixing algorithm on georegistered hyperspectral image of the area taken at different times [238, 237, 214].

## 2.2 Target detection

Target detection [191] is the task of finding and localizing target objects in a hyperspectral image, given a reference spectrum of the object. The target occurs very sparsely in the image, and can be composed of few pixels or even be smaller than a single pixel. Targets smaller than the size of a pixel are called sub-pixel targets. The reference spectrum is generally obtained from a spectral library. Generally, one or only few samples of reflectance spectrum of the target are only available.

Anomaly detection [195] is a related task with the objective of labeling anomalous objects in a hyperspectral image, without the prior knowledge of the object's spectrum. The size of the anomalous objects can also be in sub-pixel scale. Target and anomaly detection are widely used for reconnaissance and surveillance, and also in other areas like detection of special species in agriculture and rare mineral in geology [45, 73, 190].

## 2.3 Spectral unmixing

The energy captured by a pixel of a hyperspectral sensor is rarely reflected from a single surface of a single material. In airborne and space-borne imaging, the instantaneous field of view of a pixel (area covered by a pixel) on the ground is in meters, and it is highly likely that this area is covered by more than one material. For example, when imaging an agricultural land, the area under the pixel may contain vegetation and bare soil. Therefore, the measured spectrum is a combination of the spectra of different

materials in the scene. This can be modeled as a linear mixture of spectra, simply called linear mixing. Each material reflects energy in proportion to their coverage of the pixel area. Hence, the spectrum observed at the sensor is linear combination of spectra of the individual materials, weighted by their areal coverage. However, due to multiple scatterings of the light in the scene, the observed spectra is rarely a linear combination of spectra but, in fact, is a nonlinear combination [129]. Two common types of non-linear mixing models are bilinear mixing and intimate mixing. Bilinear mixing occurs when there are multiple reflections of the incident light on different materials on the scene. For instance, in forests the energy from the sun could get reflected from a leaf onto bare soil and then get reflected to the sensor. Intimate mixing occurs in fine mixtures, such as minerals, due to several multiple scattering from the particles in the mixture.

Hyperspectral unmixing is the process of recovering the proportions of pure material (called abundances) at each pixel of the image. The "pure" material spectra are called endmembers. The endmembers present in a scene may be known a priori, or obtained from the image using an endmemeber extraction algorithm [48, 331], or jointly estimated with the abundances. The methods that require endmembers to be supplied are referred to as supervised unmixing and the methods that estimate endmembers simultaneously with the abundances are referred to as unsupervised. Applications of hyperspectral unmixing include mapping of green vegetation, non-photosynthetic vegetation and soil cover [202], minerals exploration [245], urbanization study [205], fire disaster severity study [243], and water quality mapping [226].

## 2.4 Physical parameter estimation

Physical (or biophysical) parameter estimation is the process of predicting material properties, such as contents of chemicals, granularity and density of particles, from reflectance spectra. The chemical or mineral composition of the material is manifested as spectral absorption features in the reflectance spectra. The depth and the width of the absorption features are correlated to the amount or concentration to the corresponding physical parameters. Hence, regression models can be applied to relate spectra with the parameters of interest. Some examples of applications of physical parameter estimation are prediction of leaf biochemistry [321], sand and snow grain size [227, 107], vegetation biomass and structural parameter [64, 286], plant water stress [268], and soil nutrient [13].

# 3 Machine learning approaches

Machine learning algorithms attempt to predict variables of interest by learning a model from data. This section provides a brief background on machine learning techniques and terminology. As general references, the books by Kevin Murphy [216] and Christopher Bishop [30] provide a complete and detailed coverage of machine learning techniques.

## 3.1 Types of learning

Based on the type of learning, the machine learning methods can be broadly categorized into five groups as supervised learning, unsupervised learning, semi-supervised learning, active learning, and transfer learning.

In supervised learning [124], the relationship between the input and the output variables is established using a set of labeled examples, i.e., the examples for which the corresponding output variable values are known. The problem is called regression if the output variable is real and called classification if the output variable is discrete.

Unsupervised learning [125] discovers the structure or the characteristics of the input data using unlabeled examples (examples for which corresponding output values are unavailable). For instance, k-means is an unsupervised learning algorithm that clusters the input data into homogeneous groups. The principal component analysis (PCA) is another unsupervised learning algorithm that can be used to find an uncorrelated linear low dimensional representation of the input data.

Semi-supervised learning [330] utilizes the unlabeled data along with the labeled data to build relationship between the input and the output variables. The unlabeled examples are used to learn the structure of the input variables, so that this information can be exploited to better learn the input-output relationship using the labeled data.

Active learning [258] iteratively selects examples from the unlabeled data for manual labeling, and adds them to the labeled training set. The picked examples are the ones deemed most important for improving the input-output predictive performance. This cycle is repeated until the model exhibits a

desired performance. In this way, the goal of active learning is to produce results similar to supervised or semi-supervised learning methods with fewer training examples.

Transfer learning [229] utilizes the information learned from one problem to solve another problem. The tasks (output variable), the domains (input variables) or both can be different between the source and the destination problems. Hence, compared to traditional learning, transfer learning allows the task, the domain or their distributions to be different during training and testing. Domain adaptation is a subset of transfer learning in which the source and the destination have different domains. Similarly, multitask learning is a type of transfer learning where multiple related tasks are simultaneously learned with the objective of improving the performances on both tasks.

## 3.2 Non-probabilistic and probabilistic models

### 3.2.1 Non-probabilistic

Non-probabilistic models produce point estimation of the output and do not model the probability distribution of the output. These models have a decision or a regression function that estimates the value of the output based on the value of the input. These functions have some free parameters which are estimated by minimizing some cost function during training, with the goal of learning the input-output relationship. Certain penalties might be enforced on the possible values of the parameters via regularization to control the complexity of the model or to encourage certain properties in the solution.

### 3.2.2 Probabilistic

The probabilistic models infer the probability distribution of the output. The generative probabilistic models consider both the input and the output as random variables, and model a joint distribution of the input and the output variables. In contrast, the discriminative probabilistic models consider the input to be deterministic and the output to be random, and model the distribution of the output variables as a function of the input variables, i.e., they model conditional distribution of the output given the input. The generative model learns the process by which the input and the output are generated, while the discriminative model only learns how to predict the output when the input is given. In probabilistic terms, generative model learns $p(x, y) = p(y|x)p(x)$ while discriminative model only learns $p(y|x)$, where $x$ and $y$ are input and output respectively. Generative models have the advantage of being able to generate samples of the data, however discriminative models typically perform better than generative models for classification and regression as it requires larger number of samples to learn a generative model (because they need to model $p(x)$ along with $p(y|x)$ to learn $p(x, y)$). The parameters in probabilistic models are also considered to be random variables, whose point values or distribution are to be inferred from the data. The parameters can be learned by maximum likelihood estimation, maximum a posteriori estimation, or Bayesian inference.

- The maximum likelihood estimate (ML) estimation, makes point estimates for the parameters by maximizing the likelihood (probability) of observing the data given the parameters ($p(y|\theta)$ where $\theta$ represents parameters). The maximum a posteriori (MAP) estimation finds point estimates for the parameters by maximizing the probability of observing the parameters given the data (or posterior probability of parameters, $p(\theta|x)$). The ML estimation is equivalent to minimizing a cost function, and the MAP estimation equivalent to minimizing a cost function with regularization in non-probabilistic settings.

- The Bayesian inference finds the probability distribution of the parameters using the Bayes theorem, rather than just making point estimates. Exact Bayesian inference is intractable for many models, so different approximate inference techniques such as Laplace approximation, variational inference, and Markov chain Monte Carlo sampling have been developed. The main advantage of Bayesian inference over ML and MAP estimates is that Bayesian inference can properly model the prior believes about the model and handle uncertainties in the data and the model.

## 3.3 Bias and variance

The error in supervised learning models can be generally attributed to two sources: bias and variance. Bias is the error resulting from the model making wrong assumptions about the data. For instance, if a linear equation is used to model data whose input-output relationship is quadratic, the model will underfit and have high bias. It is characterized by high training error and high generalization (or testing)

error. Underfitting occurs when the complexity of the model (the space of all the functions that the model can explain) is insufficient to represent the data. On the other side, variance is the error resulting from the sensitivity of the model to small changes in training data. It occurs when the model tries perfectly fit all the training data points, rather than generalizing the trend. High variance arises when a model overfits the data, such that it has low training error but high generalization error. Overfitting typically occurs in complex models with large number of parameters when the amount of training data is small.

## 3.4    Parametric and non-parametric models

Models with fixed number of parameters have fixed complexity and are called parametric models. These models will underfit if the data complexity is greater than the model's complexity. For instance, a linear regression model which has fixed number of parameters (equal to the number of input features) will have high bias if fitted on quadratic data. In contrast, a non-parametric model can increase the number of parameters (and hence its complexity) with the available training data. An example of a non-parametric model is the nearest neighbor classifier. In nearest neighbor, the predicted class of a test sample is the class of the most similar training sample. In this model, the training data itself are the parameters. Hence, increasing the number of training samples increases the number of parameters and the complexity of decision boundary that can be modeled.

## 3.5    Model selection and performance evaluation

Model selection is the process of choosing the best model for a task from a set of candidates. A performance on the training data is not a good metric for choosing best model as it does not capture generalization performance. The goal of learning is to accurately make predictions on new data, not on the training data. Therefore, models are evaluated on a separate set of independent samples called the validation set. In model selection, the set of candidate models do not have to be entirely different algorithms, but could be same algorithm under different hyperparameter settings. Hyperparameters are the variables that control the properties of the model and are typically set before training. For instance, the number of layers in a neural network and the number of clusters in k-means algorithm are hyperparameters. Whenever enough data is not available for building separate training and validation sets, k-fold cross-validation technique can be applied. In this, the whole dataset is randomly divided into k disjoint subsets (folds), and one subset is used for validation while remaining k-1 are used for training. This process is repeated k-1 more times until each subset is chosen once for validation. Then, the results from all the folds is accumulated. In the extreme case, the value of k could be set as high as the number of training examples to get leave-on-out-cross-validation, which is useful when the number of training examples is very small.

If validation set is used to tune the hyperparameters of a model, the performance on validation set cannot be used as proxy for generalization performance of the method, because the hyperparameters were fitted to maximize performance on validation set. In this case, a third independent, separate set of samples, called testing set, should be used for performance evaluation to get unbiased estimate of the generalization performance.

# 4    High dimensionality of hyperspectral datasets

Hyperspectral datasets are high dimensional and suffer from scarcity of ground truth data. The dimensionality of spectra is equal to the number of bands, with each band representing a dimension in the feature space. Typically, there are hundreds of bands in a hyperspectral image, and consequently the dimensionality of spectral data is in hundreds. Additionally, most hyperspectral dataset come with very few ground truth information. This is due to the difficulties in collecting ground truth measurements, as it may be necessary to travel large distance from one location to another to obtain it. Moreover for unmixing and physical parameter prediction tasks, the process of determining ground truth chemical contents of samples in the laboratory is expensive and time consuming, limiting the amount of ground truth data. This unfortunate combination of high dimensional input spectra and low number of labeled examples causes problems of model overfitting and low generalization performance in the algorithms. This is referred as the curse of dimensionality or Hughes phenomenon. Since, collecting more training example is virtually impossible, one possible solution is to somehow reduce the dimensionality of the input spectra, while preserving the information contents. This is possible as the information contents of a hyperspectral image lies in a lower dimensional subspace, due to the high correlations between the

reflectance of the neighboring bands. Consequently, dimensionality reduction techniques such as the feature extraction [123] (transformations to generate lower dimensional spectral representation) and the band or feature selection [16] (methods that select a subset of most significant bands) are commonly used as pre-processing for hyperspectral data analysis. The principal component analysis [147] and the stepwise regression are the examples of commonly used feature extraction and band selection techniques. An open challenge in hyperspectral algorithm development is in designing methods that can efficiently exploit the underlying low dimensional nature of hyperspectral datasets.

Similar to the correlations across spectral dimension, there is also spatial dependencies between the neighboring pixels in a hyperspectral image. This is due to the fact that the material properties in a natural scene varying smoothly in space and the presence of a material can increase or decrease the likelihood of the occurrence of another material in its vicinity. The spatial information can be exploited be build more robust models, as it is seen with the success of spatial-spectral classification [97] and unmixing algorithms [262].

# 5    Machine learning for hyperspectral analysis

In this Section, we review recently published machine learning based hyperspectral image analysis methods. Each Subsection discusses methods that utilize a particular type of machine learning algorithm. The categorization of the machine learning algorithms is loosely based on the one used in Kevin Murphy's book [216]. Within each Subsection, we divide methods by the type of hyperspectral data analysis task they solve.

## 5.1    Gaussian models

Multivariate Gaussian models are the basis for most classical algorithms for land cover classification and target detection. A popular hyperspectral land cover classification algorithm is the quadratic discriminant analysis, also known as Gaussian maximum likelihood classifier or just maximum likelihood classifier [69]. It is a discriminative model where the class conditional distribution of the data is assumed to be described by a multivariate Gaussian distribution, with the mean vectors and the covariance matrices estimated using maximum likelihood. A special case where all of the class covariance matrices are assumed to be the same is called linear discriminant analysis [19, 174].

Gaussian models have also been extensively used in hyperspectral anomaly and target detection. The Mahalanobis distance detector [47] models the pixel values of a hyperspectral image using a multivariate Gaussian, and labels the pixels having low likelihood under this distribution as anomalies. The Reed-Xiaoli (RX) detector [47, 194] extends this by modeling only the neighborhood around the test pixel by a Gaussian distribution, not the entire image. Common target detection algorithms, such as spectral matched filter and adaptive cosine detector [192, 279], also assume Gaussian distributions for the target and the background pixels.

Gaussian models can also be found as components in more advanced algorithms. For instance, the classification method by Persello et al. [234], which performs both active learning and transfer learning, utilizes Gaussian models. In this method, the class probabilities of the data were modeled by Gaussian distributions and query functions defined over the class probabilities were used to iteratively remove examples belonging to the source dataset from the training set and add examples belonging to the target dataset to the training set.

## 5.2    Linear regression

Linear regression is a widely used method for hyperspectral data analysis. It has been applied to physical parameter estimation and unmixing problems. Linear regression is a supervised method that learns a linear relationship between a set of real input variables and a output variable by modeling the output variable as the weighted sum of input variables plus a constant. In physical parameter estimation, it is used to relate the parameter of interest with the spectral reflectance values or features derived from the spectra [293]. Some of the common spectral features used are spectral derivatives [278], tied spectra [155], and continuum removed spectra [156]. Most of the studies use step-wise regression technique to select bands that have higher correlation to the parameter of interest. In step-wise regression, bands are one-by-one added or removed from the predictive model depending on whether their presence increases or decreases the predictive performance. When used for linear unmixing, the reflectance of observed spectra at each band is modeled as a weighted sum of reflectance of the endmembers at that band,

with the weights being constant for all bands and corresponding to the abundances [128]. Using data transformations, non-linear unmixing problems (such as bilinear mixtures and Hapke mixtures) can be solved using linear unmixing framework [130].

## 5.3 Logistic regression

Logistic regression is a discriminative model primarily used for land cover classification in remote sensing. It models the class probability distribution as the logistic function of weighted sum of input features. It has been primarily used for pixel-wise classification, but as we will discuss later, logistic regression serves as a building component for more sophisticated algorithms that use ensemble learning, random fields, and deep learning. Logistic regression can perform classification with band selection using step-wise learning procedure [58] or using sparsity regularizer on the weights [327, 230, 303]. The sparsity regularizer forces many weights to be equal to zero during training, thus removing the corresponding bands from the model and keeping only the relevant bands. For improved performance, logistic regression have been trained on features derived from hyperspectral data. In [153], the squared projections on subspaces derived from class-specific spectral correlation matrices were used with logistic regression. Qian et al. [239] have proposed using 3D discrete wavelet transforms to obtain texture features from hyperspectral data cube for classification, and using mixture of subspace sparse logistic classifier to build a non-linear classifier. The 3D discrete wavelet transform based features have advantage of capturing both spatial and spectral contextual information of the scene. Spatial context can be also be incorporated to logistic regression by using morphological features [134].

Semi-supervised methods using logistic regression have also been proposed. These methods label the unlabeled data using heuristics and augment the training set with them. In [80], unlabeled pixels within the 4-neighborhood of labeled pixels were assigned the class of the labeled pixel and added to the training set, and in [168], the class labels of the unlabeled pixels were predicted by a Markov random fields based segmentation technique [169] and added to the training set.

## 5.4 Support vector machines

Support vector machines (SVMs) are the most used algorithms for hyperspectral data analysis [212]. They have been successfully applied to all data analysis tasks (land cover classification, target detection, unmixing, and physical parameter estimation). SVM generates a decision boundary with the maximum margin of separation between the data samples belonging to different classes. The decision boundary can be linear, or be non-linear through the use of kernels [257]. Using kernels, the data can be projected into higher dimensional space where a linear decision hyper-plane is fitted, which in turn is equivalent to fitting a non-linear decision surface in the original feature space. The Gaussian radial basis function kernel is used by majority of the hyperspectral SVM algorithms, however several kernels specifically designed for modeling hyperspectral data [200, 95, 255, 102] have been proposed. Since their introduction to the hyperspectral remote sensing in [132, 197], SVM have been considered state-of-the-art classifier for land cover mapping.

The most accurate SVM-based land cover mapping methods utilize spatial-spectral features, such as extended morphological (EMP) features [25, 93]. The EMP features are generated by applying a series of morphological opening and closing operations with structural element of different sizes on principle component bands of the hyperspectral image. It has been shown in [93] that appending features generated by discriminant analysis to the morphological features can further increase the accuracy. Feature selection has also been incorporated to hyperspectral classification with SVM. For instance, genetic algorithms can be utilized to select the bands and optimize the kernel parameters [21]. Similarly, a step-wise feature selection can be performed on the SVM [228]. Semi-supervised SVM that can utilize unlabeled data for training have also been developed [61]. Relevance vector machine (RVM) [276], a Bayesian probabilistic classification algorithm related to SVM, has also been applied for hyperspectral classification [33, 75, 203].

Multiple kernel learning tries to find a convex linear combination of an optimized set of kernel functions with optimized parameter that best describes the data. It has been shown that SVMs with multiple kernel can outperform SVM with single kernel for hyperspectral classification [283, 116, 298]. Using EMP features, multiple kernel learning framework can be used for spatial-spectral classification [115, 182, 171]. Kernels defined over the spectra of neighboring regions (square blocks of pixels [41] or superpixels obtained via segmentation [91] around the test pixel) have also been combined with the kernel defined over the spectrum of the test pixel to perform spatial-spectral classification with SVM.

A multiple kernel learning based transfer learning/domain transfer approach that simultaneously

minimizes the maximum mean discrepancy between the source and the destination datasets along with the structural risk functional of the SVM was proposed for classification in [266]. This method was found to be better than regular SVMs and other SVM-based transfer learning schemes. Similarly, an active learning based domain adaptation method with reweighting and possible removal of samples from the source dataset was introduced in [233]. The pixels in the source dataset misclassified by the SVM in each iteration were removed, while the target dataset pixels with the most uncertain class assignments (based on the votes of binary SVMs trained in one-vs-all approach) were manually labeled and added to the training set.

In [296, 117], binary class SVM classification was used for unmixing by assuming that the pixels lying on or separated by the max margin hyperplanes to be the pure pixels, and the pixels occurring within the margin to be the mixed pixels. The abundances of impure pixels was then given by the ratio of the distance from the margin to the margin width. Using one-vs-all scheme this method was extended for scenes with more than two endmembers. Another approach for SVM-based unmixing is to generate an artificial mixed spectra dataset with known abundances, and learn a SVM model to classify the proportion of each endmember present in a test spectra at single percentage increments [204]. The artificial dataset can be generated by calculating the randomly weighted sum of spectra belonging to a set of classes chosen at random from a list of endmembers. In another study [287], probabilistic SVM was used to generate per pixel probability of the pixel belonging to an endmember. The pixels with high probability of belonging to any one endmember were considered to be pure pixels, while the remaining pixels were considered mixed. The abundances in the mixed pixels were calculated using linear unmixing. The mixed pixels were further divided into subpixels, with the subpixels arbitrarily assigned to one of the endmember classes in numbers proportional to the class abundances. Simulated annealing was then used to arrange the subpixels in each mixed pixels to have spatial smoothness. This produced a sub-pixel mapping of the scene.

Anomaly and target detection have been performed using a SVM related algorithm called support vector data description [273]. This method generates a minimum enclosing hypersphere containing all the training data. Kernel trick can be used to find minimum enclosing hypersphere in a transformed domain. For anomaly detection, any pixel falling outside of the hypersphere enclosing all the pixels in the image are considered to be anomalies [20, 152, 118]. While for target detection, an artificial training set of target pixels can be created by adding multinomial Gaussian noise to the target reference spectra, and any test pixel falling inside of the hypersphere enclosing the artificial dataset can be labeled as a target [249]. Nemmour et al. [223] performed change detection in an area by training a SVM classifier on the concatenation of spectra from images collected at multiple dates to predict the change in the land cover.

Previously, SVM regression was applied to predict biophysical parameters from multi-spectral imagery [34, 37, 22]. While these method are also applicable to hyperspectral data, some newer methods have been developed specifically for hyperspectral data. In [38], a semi-supervised method that uses kernel matrix deformed by labeled and unlabeled data was proposed. Active learning approaches for biophysical parameter estimation that select new samples based on distance from support vectors and the disagreement between the pool of SVM regressors trained on different subsets of training data have been proposed in [231]. The idea of learning related biophysical parameter simultaneously, exploiting the relationship between them using multitask SVMs was introduced in [282]. The multitask SVMs were found to be more accurate than the individual SVMs in predicting biophysical parameters.

## 5.5   Gaussian mixture models

Gaussian mixture models [28] represent the probability density of the data with a weighted summation of a finite number of Gaussian densities with different means and standard deviations. They are commonly used to model data that are non-Gaussian in nature or to group data into finite number of Gaussian clusters. Gaussian mixture model is a good choice to model class conditional probability in maximum likelihood classifier when the image spectra do not show Gaussian characteristics [83, 176, 175]. Same is the case with anomaly and target detection algorithms which traditionally utilized Gaussian distribution to model pixel and background probability density [272].

Gaussian mixture model have also been used to cluster hyperspectral data. [271] used Gaussian mixture model followed by connected component analysis to segment the hyperspectral image into homogeneous areas. A related method called independent component analysis mixture model, which models cluster density by non-Gaussian density have also been applied for unsupervised classification of hyperspectral data [259].

The popular clustering algorithm k-means [15] is a special case of Gaussian mixture model clustering [30]. K-means starts with initial guesses for cluster centers, assigns all the data points to a cluster based on the distance to the cluster centers, calculates the mean of data in each cluster, and updates each cluster center with the mean of that cluster. The process of grouping data, calculating the mean and updating the cluster centers is repeated until convergence. The biggest issue with k-means is that it requires the number of clusters in the data to be known a priori. ISODATA [18] is a method based on k-means that does not require the number of clusters to be known a priori and works by merging and splitting the clusters in every k-means iteration on the basis of the distance between the clusters and the standard deviation of the data in each cluster. The k-means and the ISODATA are widely used for unsupervised classification of hyperspectral data [17, 220]. Unsupervised classification maps produced by them have been fused with the results of pixel-wise supervised classification to perform spatial-spectral classification [309, 271].

K-means have also been used for anomaly detection and dimensionality reduction. [84] performed anomaly detection was by labeling pixels which were distant from the cluster centers found by k-means or ISODATA as anomalous pixel and [263] used the cluster centers obtained from k-means as features for classification.

## 5.6 Latent linear models

Latent linear models find a latent representation of the data by performing a linear transform. The common latent linear models used in hyperspectral image analysis are the principal component analysis (PCA) [147] and the independent component analysis (ICA) [137]. The PCA linearly projects the data onto an orthogonal set of axes such that the projections onto each axis are uncorrelated. The projection onto the first axis captures the largest portion of variance in the data, the projection on to the second axis captures the second largest portion of variance in the data and so on, such that the axes towards the end do not capture any variance in the data but only represent the noise. On the other hand, the ICA linearly projects the data onto a non-orthogonal set of axes such that the projections onto each axis are statistically independent as possible. For PCA, the number of data samples has to be greater than or equal to the dimensionality of the data. Similarly, for ICA, the number axes onto which the data is projected has to be smaller or equal to the number of data samples. The PCA is primarily used for dimensionality reduction in hyperspectral images. The PCA is applied to the spectra in the image and only the projections which explain a significant proportion of the variance are kept [260]. Reducing dimensionality makes models less likely to overfit and also removes noise. Hence, it is widely used as a preprocessing tool for hyperspectral analysis [306, 55, 207, 92]. The minimum noise fraction (MNF) transform [164], which whitens the noise in the image before applying the PCA, is generally preferred over the PCA when reducing the dimensionality of highly noisy images. The PCA and the MNF can be used to perform non-linear dimensionality reduction using their kernalized variants [96, 224]. The spatial-spectral features can be obtained by applying morphological operations after the PCA or the MNF [236].

The partial least squares (PLS) [290] regression is a widely used method for physical parameter estimation from hypespectral data and is closely related to the PCA. The PLS fits a linear regression by projecting the input and the output onto separate linear subspaces where the covariance between their projections are maximized. [43] showed that PLS performs better than the combination of PCA and linear regression for hyperspectral data. It has been successfully applied to predict physical quantities, such as, soil organic carbon [109], biomass [65], nitrogen [122], and water stress [71].

The ICA can also be used to reduce the dimensionality of hyperspectral data. [294] observed that better classification results can be obtained if the dimensionality of hyperspectral image is reduced using the ICA compared to using the PCA or the MNF. Apart from dimensionality reduction, the ICA have also been used for unmixing [221] and unsupervised classification [82]. These methods assume that ICA axes are the endmembers and the projections are the abundances. Mixture model using ICA have been also proposed for unsupervised classification [259]. Similar to the PCA and the MNF, spatial-spectral features have been generated using morphological profiles on the image after ICA [68].

## 5.7 Sparse linear models

Linear sparse models [189] model the observed output to be the weighted linear combination of the elements (atoms) of a large dictionary with the restriction that most of the weights are equal to zero while the remaining few weights have significant magnitudes. The sparsity on the values of weights is

imposed by using a a sparse prior in probabilistic setting and a sparse regularizer in non-probabilistic setting. The dictionary can be supplied manually or be learned from the data itself. When a dictionary is learned from the data, it is automatically able to capture the data statistics. The linear sparse model is widely used for unmixing, because its formulation resembles to the linear mixing model, with the abundances being the weights and the endmembers being the dictionary elements. Iordache et al. [138] have proposed the use of sparse linear model with a spectral library as dictionary, to linearly unmix images using $L_1$ regularizer on the weights. This method is able to automatically select a subset of spectra in the spectral library as endmembers for each pixel of the image. Note that the regular least squares cannot be used in this kind of settings, since the number of spectra in a spectral library is much larger than the number of bands in a hyperspectral image. In this problem, sparsity is imposing additional constraints on a ill-conditioned problem to make it solvable. A modified spatial-spectral version of this method [139], additionally imposes the spatial contextual information by applying total variational regularization, i.e., minimizing the $L_1$ norm of the endmember-wise abundance differences between the neighboring pixels.

A multitask spatial-spectral extension to the same method, where sparsity is imposed to all the pixels of an image simultaneously to force the neighboring pixels to be composed of same endmembers, using $L_{2,1}$ norm on the abundance matrix was proposed in [140, 141]. In all these methods, non-negativity constraint on the abundances was imposed during optimization, but the sum-to-one constraint was not applied. A hierarchical Bayesian approach to sparse unmixing was introduced in [275]. Zero mean Laplace prior, estimated by a truncated Gaussian distribution, was used as prior on the abundance for sparsity and non-negativity constraints, and a deterministic heuristic to impose sum-to-one constraints was suggested. In [44], multiple spectra belonging to each endmember classes were added one by one to the dictionary until there was no gain in reconstruction accuracy. The authors also proposed using non-local coherence regularizer that promotes coherence between coefficients corresponding to the atoms belonging to the same endmember class along with local neighborhood coherence regularizer.

There is also a non-regularization based sparse endmember extraction and unmixing technique in literature. It first performs fully constrained least squares unmixing, and then iteratively removes endmembers with smallest abundances greedily until the required accuracy and sparsity were obtained [113]. This method was shown to perform better that $L_1$ regularizer based sparse unmixing methods in the experiments. Manifold based regularizers have also been used for sparse unmixing with the assumption that hyperspectral data is sparse in a lower dimensional non-linear manifold in a high dimension space [185].

Classification can be performed with linear sparse models by either using the sparse representation as features for a classifier [49, 81] or by selecting the class that has minimum class-wise reconstruction error [251, 50, 52, 44]. In [49], dictionary learning was used to infer sparse representation of the spectra, which was used as feature for SVM classification. In the experiments, it was found that using sparse feature performed better than using the raw spectra or the principal component analysis (PCA) transformed spectra. A compressed sensing based deblurring method to reconstruct hyperspectral signal from multispectral measurement was also introduced in this work. Similarly, in [81] dictionary learning with total variation regularizer was used to learn discriminative spatial-spectral sparse representation jointly with the sparse multinomial logistic regression trained on them.

The sparse reconstruction based classification algorithms learn the sparse representation of a test example using a dictionary containing examples of all classes, and then reconstruct that example using only dictionary atoms belonging to a specific class. The class with minimum reconstruction error is the predicted class for the test example. These methods use basis pursuit [44], greedy pursuit [50] or homotopy based algorithms [251] to obtain $L_1$ regularized sparse coefficients. Kernelized versions of sparse reconstruction algorithm can be used to construct dictionary in the higher dimensional kernel space [52, 181]. In [50, 52] spatial contextual information was utilized by enforcing that the Laplacian operation on the reconstructed image be equal to zero, and by optimizing joint sparsity that promotes neighboring pixels to be composed of the same dictionary elements. Similarly, a spatial-spectral classification method in which the surrounding pixels were weighted based on their similarity to the test pixel and reconstructed jointly with the test pixel using sparse model with spatial coherence regularizer was introduced in [317].

A new set of sparse code can be learned from the learned sparse code. This process can be repeated multiple times, with a new set of sparse code being learned from the sparse code learned in previous step, to obtain deep sparse code. A superpixel guided deep spatial-spectral sparse code learning method was published in [90]. In this method, first the image was segmented into superpixels and sparse code was computed for spectrum at all the pixels of image. The learned sparse codes were averaged over the

superpixels and assigned to all the pixels inside the superpixels. This process was repeated multiple times to generate a stack of features which were concatenated and classified using a support vector machine.

Chen et al. [51] performed target detection by formulating the problem as a two class classification problem and using a previously proposed spatial-spectral sparse classification technique [50]. Their method generated target training examples by running MODTRAN [27] with randomly varying parameters on the reference target reflectance spectra, and used randomly selected pixels from the test image as training examples for the background. This method was shown to out-perform spectral match filters, matched subspace detectors, adaptive subspace detector, and SVM based binary classification.

## 5.8   Ensemble learning

Ensemble learning [246] is a supervised learning technique of merging the results from multiple base predictors to produce a more accurate result. The outputs of the base predictors must be diverse and uncorrelated for ensemble learning methods to produce superior results. It has been applied successfully for hyperspectral classification. There are basically three types of ensemble learning approaches: bagging, boosting, and random forest. Bagging (also called bootstrapping) combines the results from multiple predictors trained on randomly sampled subsets of training set, where as boosting combines the results from multiple separate weak predictors trained on the whole training set. Random forests is bagging of decision/regression trees with random selection of features (also called feature bagging).

AdaBoost [98], an adaptive boosting technique, has been widely used to build robust hyperspectral classifiers. It was used with support vector machines (SVMs) trained on clusters of bands [241], multiple kernel support vector machines trained on screened training samples [114], support cluster machines with different number of clusters [62], and linear and quadratic decision stumps trained on randomly selected features [150] . Random forest and other random feature subspace based ensemble learning methods are considered more attractive for hyperspectral data as they use a reduced feature set to learn each ensemble member, which makes them less prone to overfitting. A bias-variance analysis in [201] found that random forest with embedded feature selection and Markov random field (MRF) based post-processing are best suited for hyperspectral data.

A random subspace SVM that trains multiple SVM classifiers on random subsets of bands and combines the result of each SVM on the test spectrum via majority voting was proposed in [299]. This method performed better than regular SVM and random forest, particularly when very few training examples were available. This method was further improved by optimizing every random subspace, by selecting the optimal bands using genetic algorithm with Jeffries-Matusita (JM) distance as fitness function in [56]. An adaptive boosting technique for random subspace SVM which jointly learns the kernel parameters of the SVMs and the coefficients of adaptive boosting was formulated in [119]. Ham et al. [121] introduced random forest algorithm to hyperspectral classification, and also proposed a novel idea of optimally selecting bands in each random subspaces projection using simulated annealing. Their method was extended to perform transfer learning by using the class hierarchy learned on the source image, when making predictions in the test (destination) image with few or no labeled data [240]. Transfer learning version of AdaBoost algorithm, TrAdaBoost [66], was used to reweight the samples from source image in the training set of a SVM classifier after unlabeled pixels in the target image were manually labeled and added to the training set in [193].

Extreme learning machines (ELM) [133] are three layered feed-forward neural network, where the weights from input layer to hidden layer are assigned at random and the weights from the hidden to the output layer are learned. They were introduced to hyperspectral classification as base classifier for bagging and AdaBoost in [250]. They were found to be very fast compared to the methods using SVM as base classifier, while provide similar accuracy. The authors also proposed using extended morphological map (EMP) features for spatial-spectral classification with ELM. Algorithms that dynamically select a separate subset of classifiers from an ensemble for each test pixel, considering the validation performance of pixels similar and near to the test pixel, have been proposed [70]. This method used the SVM and the ELM as base classifiers, and were more accurate than similar methods with fixed set of ensemble members.

A rotation forest [244] is a classifier which builds ensemble members by dividing the features of the training data into random groups, performing rotational transform on each group, and then collecting all the rotation vectors into a single transformation matrix which is used to rotate the entire training set before training a decision tree. The study by Xia et al. [306] applied rotation forest for hyperspectral data, with rotational transformations generated by principal component analysis (PCA), minimum noise fraction (MNF), independent component analysis (ICA), and Fisher's discriminative analysis and

showed their method to outperform bagging, random forest, and AdaBoost. The authors later proposed improving the classification performance by exploiting spatial context using Markov random fields [304] and extended morphological map features [305].

Peerbhay et al. [232] applied random forest classifier to identify anomaly by training a binary classifier on synthetic dataset which considered the image pixels to be the training samples of the non-anomaly class and the samples from empirical marginal distribution of image pixels to be the training samples of the anomaly class. Though primarily used for classification, random forest have also used for physical parameter estimation. In previous studies, random forest regression was used for band selection and retrieval of biophysical parameters, such as biomass [4, 219], nitrogen [3], and water stress [142].

## 5.9    Directed graphical models

Directed Graphical Models [292], also known as Bayesian networks, define a factorization of the joint probability of a set of variables over the structure of a directed graph. Each variable is represented by a node and the directed edges represent the conditional independence properties. Each variable in the joint distribution is considered to be conditionally independent given all its parents in the graph and the joint distribution is defined by the product of conditional distributions of all the variables given their parents. Bayesian networks have been primarily applied to hyperspectral unmixing. They have the advantages of providing uncertainty about the abundance estimates and probabilistically modeling the endmember spectral variability. Most methods follow a similar hierarchical Bayesian framework. They start with a linear mixing model or one of its nonlinear transforms, and assume prior distributions over the abundances and the endmembers, and then use non-informative priors on the hyperparameters. The likelihood or the noise model used is mostly Gaussian. Since, exact inference is not possible in these models, they use a Markov Chain Monte Carlo (MCMC) [12] method to estimate the posterior distribution of the abundances and sometimes also the endmembers.

Based on the distributions used for priors and hyperpriors and the inference algorithm selected, different characteristics are obtained. A method to linearly unmix pixels when the endmembers of the scene are known was proposed in [79]. The prior used on the abundances was a uniform sampling on a simplex to enforce non-negativity and sum-to-one constraints. This model was later extended to bilinear unmixing [120] and post-nonlinear unmixing [9]. In [275], the authors proposed the use of zero mean Laplace prior on abundance to promote non-negativity and sparsity on the retrieved abundance coefficients. For computational purposes, the Laplace prior was approximated by truncated Gaussian distribution. This model lacks sum-to-one constraint in its formulation, so the authors have suggested a heuristic to enforce this constraint.

The variability in endmember spectra was addressed in [85], where the endmember spectra were considered to be a Gaussian distributed vectors, with the means set to the endmember spectra extracted from the image and the covariance matrix learned from the data. In [213], joint estimation of the endmembers and the abundances distributions were made using a linear mixing model with independent Gamma priors on the mixing coefficient and the endmember reflectance values. The Gamma priors enforced positivity constraints on the mixing coefficient and the endmember reflectance values. Dobigeon et al. [78] later added sum-to-one constraint to this model by replacing the independent Gamma prior on mixing coefficient by a Dirichlet prior. [254] proposed hardware and software implementation strategies to scale up these methods to large scale. Methods have also been devised to jointly estimate the abundances and the endmembers in bilinear and post-nonlinear methods [8].

## 5.10    Undirected graphical models

The strong dependencies between neighboring pixels in hyperspectral images can been exploited for classification and unmixing using undirected graphical models (also called Markov random fields) [103]. Undirected graphical models (UGMs) [225] are the probabilistic models that define a joint probability distribution of a set of random variables using the structure of an undirected graph, such that the joint distribution can be factorized over maximally connected sub-graphs, called cliques. Similar to directed graphical models, the nodes of the graph represent the variables while the undirected edges encode the conditional independence properties. In UGMs, each node is conditionally independent to all of the the nodes given the neighbors of the node. The joint probability is equal to the normalized product of positive functions, called potential functions, defined over all of the cliques in the graph.

The most common type of undirected graphical model used in hyperspectral image analysis is grid-structured pairwise models. They have been widely used since their introduction to hyperspectral land

cover mapping in [143]. The graph structure used in these models is a grid, with pixel labels representing the nodes and an edge between every 4-connected or 8-connected neighbors. Two type of potential function is defined for these models, namely, the unary potential function and the pairwise potential function. The unary potential function is defined for each node and captures spectral information, while the pairwise potential function is defined for each edge and captures the spatial information by promoting neighboring similar pixels to be labeled the same class. It is common to use a pixel-wise classifier, such as Gaussian maximum likelihood classifier [143], logistic regression [32], probabilistic SVM [270, 307], Gaussian mixture model [175], and ensemble method [201], to derive unary potentials. Ising/Potts based models [325, 32, 270, 307, 165] are primarily used for the pairwise potential. In these methods, two types of inference is generally performed–(a) maximum a posteriori (MAP) inference using algorithms such as iterated conditional mode [143], simulated annealing [270], and graph cuts [32, 307, 165] and (b) probabilistic inference (also called marginal inference) by using loopy belief propagation [325, 170].

Tarabalka et al. [270] introduced a novel Potts-based spatial energy function that only created dependencies between the neighboring pixels if there was no intensity edge in between them. They found that SVM followed with edge-based Markov random fields performed better than SVM followed by non-edge based Markov random fields, which in turn was better than SVM followed by majority voting. The conditional random fields (CRFs) are the Markov random fields that model conditional distribution by having their potential functions be parameterized by input features [267]. Studies [325, 165] have successfully applied grid-structured pairwise CRFs to hyperspectral classification. CRFs are generally preferred over MRF because being discriminative model, they can better utilize training data for classification. However, to train a full CRF a large number of training examples are required because CRFs have much more parameters than MRFs. This is a problem for hyperspectral image analysis tasks, which are well-known for having limited ground truth. To tackle this problem, simpler formulations for the pairwise energy functions based on the similarity of the neighboring pixels have been proposed [315, 328, 320]. Band selection can also be jointly performed with the learning of a CRF to model land cover by applying Laplace prior on the CRF parameters as in [324]. This helps in reduction of the complexity of CRF model and hence is better suited when there are limited training examples.

All the methods discussed so far have only unary and pairwise potentials, and cannot express higher level relationship occurring between the different regions in the image. Higher order MRF/CRF have potential functions defined over a set with more than two nodes. Zhong and Wang proposed using robust $P^n$ model [154] for hyperspectral classification [326]. In this method, higher order potentials were defined over pixels inside each segment obtained by unsupervised segmentation of the hyperspectral image by mean-shift algorithm. This potential encouraged every pixel inside each segment to be assigned to the same class and was used along the unary and the pairwise potential functions. A heuristic approach is to use pairwise models to model segment labels directly and assign all the pixels in the segment to the same class. This approach is more rigid than $P^n$ model where all the pixels in the segment not necessarily have to be assigned to the same class. Similarly, [108] integrated information from hierarchical segmentation map into pairwise energy function to incorporate higher order information into the pairwise MRF.

A semi-supervised classification method that utilized MRF with semi-supervised graph priors on parameters was introduced in [166]. Li et al. [167] proposed an active learning based classification method consisting of a pipeline of logistic regression followed by MRF, in which the unlabeled pixel having the largest difference in class probabilities for the top two most probable classes were selected for manual labeling and added to training set. Similarly, another heuristic approach for active learning that selected unlabeled pixels which were assigned to different classes by logistic regression and the combination of logistic regression and MRF was proposed in [265].

Eches et al. [87] proposed using Markov random field and hierarchical Bayesian linear unmixing for simultaneous spatial-spectral unmixing and classification of a hyperspectral image. They modeled each pixel as the result of linear mixing of the endmembers with added zero mean Gaussian white noise. Potts-Markov model was used as prior for the class distribution over the image and separate sets of abundances prior were defined for each class. The abundances had softmax logistic regression as prior to enforce additivity and non-negativity constraints, and the coefficients of logistic regression had Gaussian prior with the mean and the variance distributed as normal and inverse gamma respectively. Bayesian inference on this model was performed using Metropolis-within-Gibbs algorithm. In the experiments, the proposed method performed better than pixel-wise Bayesian unmixing [79] and fully constrained least squares unmixing, but was more computationally intensive. The authors solved this problem by proposing an adaptive MRF which established relationships between homogeneous regions in image rather than all the pixels to reduce computational complexity in [86]. The new method used self-complementary area filter to segment the image into homogeneous regions as pre-processing. Potts-Markov model was used

as prior to define class probability over these regions. Instead of using softmax logistic regression prior over class conditional abundances, this method used Dirichlet distribution with uniform prior on the parameters. This method performed as good as the previous method [87] with lesser computational cost.

A hierarchical Bayesian method for non-linear unmixing was proposed in [11]. Each pixel in the image is modeled as sum of linear mixture of endmembers, a nonlinear term and Gaussian noise. All the pixels in the image were segmented into different classes by a MRF and the class conditional non-linear term was modeled by Gaussian processes. In the following paper [10], the authors proposed to model the nonlinear term by a gamma MRF instead, to enforce the non-linear term to be non-negative and be not limited to finite levels.

Sub-pixel mapping is the process of creating a classification map at scale smaller than the size of the pixel. The MRFs and CRFs have been used to create sub-pixel maps by first predicting the the content of coarse pixels, using methods like Gaussian likelihood estimated probability [297] and linear unmixing [319], then using this information to generate unary potentials for the sub-pixels inside all the coarse pixels, and finally using MRF/CRF to globally assign labels to all sub-pixels in the image while promoting spatial smoothness.

## 5.11   Clustering

Clustering algorithms have been primarily applied for unsupervised land cover classification of hyperspectral images [288], however they have been also used for band selection, semi-supervised classification, dimensionality reduction, and unmixing. The commonly used clustering algorithms are k-means [30], ISODATA [18], meanshift [59], affinity propagation [99], graph-based clustering [291], and Dirichlet process mixture models [274]. The k-means and the ISODATA methods were discusses earlier in Section 5.5. K-means and ISODATA are the most popular clustering algorithm used in remote sensing and they have found usages in unsupervised classification [17, 220], anomaly detection [84], and dimensionality reduction [263].

Meanshift is an algorithm that can be used to find the locations of the modes in multi-modal distributions of data. It iteratively updates the estimates of the modes by the mean of all the data points weighted by a kernel function placed at the locations of the mode estimates. Meanshift does not require prior knowledge of the number of clusters unlike k-means, which is an advantage. [135] proposed using meanshift for unsupervised classification of a hyperspectral image. Then, they classified the mean spectra of each cluster using a supervised classifier and assigned all the pixels in the cluster with the predicted class.

Affinity propagation is a clustering technique that uses message passing between data points and does not require prior knowledge of the number of clusters as well. It has been used primarily for band selection by clustering the bands in an image and selecting a representative band for each cluster [145, 308]. Graph based clustering techniques represent the structure of the data in graphs, with data points being nodes and the similarity between them being edges and express clustering as graph partitioning problem. Graph based clustering was used in [42] to perform semi-supervised land cover classification. Dirichlet process mixture models will be discussed in Section 5.13. Dirichlet process mixture models have been used for semi-supervised land cover classification [149], unmixing [206], and endmember extraction [313, 314].

## 5.12   Gaussian processes

Gaussian processes [301] (GPs) are non-parametric models that assume all the observed and the unobserved data are jointly distributed by a multivariate normal distribution. The mean vector of the multivariate normal distribution is typically assumed to be a zero vector and the covariance matrix is estimated using a covariance function (same as kernel functions). Predictions are made using Bayesian inference. A popular covariance function for hyperspectral data analysis is the squared exponential function. The square exponential covariance function is called the Gaussian radial basis function in context of the support vector machines. GPs are primarily used for supervised learning, but can also be used for unsupervised learning. Most of the use of GP in hyperspectral data analysis has been in non-linear physical parameter estimation. The advantages of the GP over the traditional approach are that being probabilistic GPs provide a prediction confidence map; and being Bayesian they can handle uncertainties better and are less likely to overfit; and finally being non-parametric their complexity can grow with the amount of data to model highly non-linear functions.

GPs have been applied to predict leaf chlorophyll, fractional vegetation cover, and leaf area index in [286]. This study first showed that using GPs to predict biophysical parameters from common vegetation

indices is more accurate than the standard technique of using linear regression. Then, they used GP to relate the raw spectrum to biophysical parameter and used step-wise backward elimination to select four bands whose reflectance could best predict the biophysical parameters. It was observed that GP with band selection performs the best. In the follow up study [285] auto-relevance detecting (ARD) squared exponential covariance function was used to select bands to predict chlorophyll instead of stepwise regression.

Both these techniques make assumption that the noise in each pixel is same, and it is independent of the signal. This assumption was relaxed in [160] by allowing the noise power to smoothly vary over observations using variational heteroscedastic Gaussian processes. This approach showed better performance than previous methods. The tutorial by Camps-Valls et al. [40] covers biophysical parameter estimation using GPs from hyperspectral imagery in detail. In a different application of Gaussian process regression, Murphy et al. [217] used GP to accurately locate a feature in ferric iron spectra which occurs at around 900 nm from only the shortwave infrared spectra (2000 nm-2500 nm), a commonly used spectral region in mineralogy.

Gaussian processes can be employed for classification by using discrete likelihood functions. However, when using these functions, Bayesian inference does not have an analytical solution and approximate inference methods have to be used. In [23], Gaussian process classification was used to classify pixels in the image using squared exponential and neural network covariance functions, and logistic likelihood with Laplace and expectation propagation inference methods. In the experiments, their method outperformed the support vector machines. It was seen that the runtime of Gaussian process classifier became very high as the number of training sample was increased, and the authors proposed using approximate sparse GP methods, such as informative vector machines, for faster operation at the cost of some performance in applications where runtime is critical. An active learning based classification method that introduces three heuristics for Gaussian process classification of hyperspectral images was proposed in [264].

A spatial-spectral classification scheme using GP was introduced in [148]. In this method, a set of GPs were used to model the per-class spatial variations in the reflectance in each band of the image pixels and for each class the mean reflectance estimate at each pixel was used as the mean of multivariate class-conditional normal probability distribution at that pixel. The observation angle dependent (OAD) function is a covariance function designed primarily for classifying minerals from hyperspectral data with GPs [198, 256]. Five parsimonious Gaussian processes for modeling class-conditional distribution in quadratic discriminant classifier, when limited training samples are available, were introduced in [94]. Parsimony was enforced using constraints on eigen decomposition of the covariance matrices, assuming that the discriminating information lies on a lower dimensional subspace.

Gaussian process latent variable model (GP-LVM) [159] can be used for non-linear unmixing and endmember extraction [7]. This approach uses GP-LVM to perform non-linear dimensionality reduction to obtain abundances at each pixel of the image while enforcing positivity and sum-to-one constraints. Then, a GP regression model is trained to predict the pixel spectra from the estimated abundances. Finally, the endmember spectra are predicted by that GP regression model using pure abundances as input. This method works in reverse order compared to most other combined endmember and abundance estimation methods, in that, the abundance is estimated before extracting the endmembers.

## 5.13 Dirichlet processes

Dirichlet process (DP) [274] is a Bayesian non-parametric model typically used to cluster data by modeling the data as a mixture of possibly infinite components. A set of infinitely many random variables are said to be distributed as Dirichlet process if the marginal of this joint is a Dirichlet distribution. When used for clustering, the number of clusters modeled by Dirichlet process can grow with the data, not requiring it to be set beforehand, unlike most clustering methods. Dirichlet process has been applied for classifying and unmixing hyperspectral data.

In [149], a spatially adaptive, semi-supervised DP based classification algorithm was proposed. This algorithm modeled the data by a DP based infinite mixture of Gaussian, with the Gaussian means given by spatially varying Gaussian processes. This method is capable of discovering new classes in the unlabeled samples in the training set, which is an uncommon feature in hyperspectral classification algorithms. A DP based joint endmember extraction and linear unmixing algorithm was proposed in [206]. This method used sticky hierarchical DP as spatial prior for the abundances in a Bayesian linear unmixing framework and Gibbs sampling to infer the posterior distributions of the endmembers and the abundances.

Piecewise convex endmember (PCE) detection algorithm, proposed in [313], models hyperspectral

data as collection of disjoint convex regions. DP was used to determine the number and the location of these convex regions, and a maximum a posteriori estimate based method was used to estimate the Gaussian distributed endmembers in those regions. A stochastic expectation maximization algorithm was used to iteratively refine the extracted abundances and endmembers. A fully Bayesian version of this algorithm that uses Metropolis-within-Gibbs sampling for inference was introduced by the authors in [314].

## 5.14   Deep learning

Deep learning [110, 26] methods apply a hierarchy of non-linear transforms to the data with the goal of generating an abstract, useful representation. The growth in the development of graphical processing units (GPUs), availability of large datasets, and innovations in training deep networks such as dropout, rectified linear unit, residual learning, batch normalization, and dense connections have led to state-of-the-arts performances in the fields of computer vision, speech recognition, natural language processing, and many more engineering disciplines. Remote sensing researchers have also developed numerous deep learning based remote sensing data analysis methods which has produced top performances. Currently, the main focus is on land cover classification task but in future we can expect deep learning to be used for other tasks as well. Due to its popularity, a tutorial [318] was also published recently to introduce deep learning to remote sensing researchers. In this section, we review the development of deep learning methods for hyperspectral analysis.

Chen et al. [54] proposed classifying the features learned from stacked autoencoders by logistic regression for hyperspectral classification. Autoencoders [289] are the neural networks that map the input to a hidden layer of size smaller than that of input and then try to reconstruct the input signal at the output using hidden layer activations, so that the activations in the hidden layer provides a compact, non-linear representation of the input. The autoencoders can be stacked and fine tuned, such that the hidden representation of one is the input of the other, to learn deep features of the input. Chen st al. showed that the features generated by stacked autoencoder from hyperspectral image was better than that generated by commonly used feature extraction methods, such as principal component analysis (PCA), kernel PCA (kPCA), independent component analysis (ICA), non-negative matrix factorization and factor analysis. They extended this work by using deep belief network for learning deep representation in [55]. Deep belief networks are formed by stacking a kind of generative probabilistic neural networks, called restricted Boltzmann machines [161]. In both papers, three kinds of features were tested–spectral, spatial-dominated and spatial-spectral. The spectral features were generated by learning features using the raw spectra taken from the pixels of hyperspectral image as input. The spatial-dominated features used dominant principal components (PCA components) of the pixel neighborhood as the input to the feature extractor and the spatial-spectral approach used the concatenation of the raw pixel spectra and the dominant PCA components of the pixel neighborhood as the input to the feature extractor. They found that the performance of the spatial-spectral features was better than that of spatial-dominated features, which in turn was better than the performance of spectral features. Their method with spatial-dominated and spatial-spectral features also outperformed widely used spatial-spectral support vector machines (SVM). It was seen that the accuracy generally grew with the number of PCA components used up to a limit, and then remained constant. They found that accuracy is highly dependent on the depth of the network and suggest using schemes like cross-validation to learn the best depth.

Liu et al. [183] combined stacked denoising autoencoder (an autoencoder trained on noisy input) and super-pixel segmentation [178] for spatial-spectral classification. Their method used a three-layered stacked denoising autoencoder trained on pixels of an image to generate features to prepare a classification map. This classification map was then segmented using a super-pixel approach and all pixels in a segment was classified to a common class based on majority voting. Stacked autoencoders that jointly learn spatial-spectral features have also been proposed. Ma et al.[187] proposed the use of the PCA bands of the patch around the pixel and the pixel's spectra as joint input to stacked autoencoders for generating spatial-spectral features. They also showed that promoting pixels with smaller spectral angle to have same hidden layer representation in autoencoders trained on pixels and spatially averaging hidden layer representation of each autoencoder before stacking subsequent autoencoder can produce good spatial-spectral features in [188].

A popular deep learning architecture for vision tasks is a convolutional neural network [157] (CNN). Inspired by the mammalian visual system, these neural networks contain a series of convolution layers, non-linearity layers, and pooling layers for learning low-level to high-level features. In the convolution layers, rather than being fully connected to the input, each hidden layer unit is connected via shared

weights to the local receptive field (pixel neighborhood for images) around the input. Non-linearity layer makes the activations non-linear function of the input. In the pooling layers, the responses at several input locations are summarized, via max operations, to build invariance to input translations. Networks with one dimensional [131], two dimensional [247], and three dimensional [53] convolutional layers have been developed for analyzing hyperspectral data. Methods with one dimensional convolution layer take spectra as input and learn features that capture only spectral information. [131] proposed a new five layered convolution neural network trained on the spectra for classification. This architecture contains an input layer, a convolution layer, a max pooling layer, a fully connected layer and an output layer. It was tested on two three hyperspectral datasets, and it was found that this method outperformed the SVM, and the existing shallow and deep neural networks, namely, two layer fully connected neural network, LeNet-5 [162] and stacked autoencoder based method [54]. Li et al. [173] devised a scheme to perform spatial-spectral classification using 1D CNN. They trained a 1D CNN on pairs of spectra to classify the common class if the spectra pair belong to same class and predict 0-th class otherwise. During testing, the test pixel and its neighboring pixels were paired one-by-one, fed to the network, and the class of the test pixel was found by majority voting.

Generally, methods that use two dimensional convolution layers, first reduce the number of bands in the image by using a dimensionality reduction technique, such as PCA, and then apply two dimensional convolutional network on image patches to extract features. These methods reduce the number of bands to control the size of network, as larger network require more training data which is limited. In [322], the feature vector extracted from the last layer of a convolutional neural network, which was trained on the principal components bands of image patches, was concatenated with spectral feature vector obtained by applying a manifold based dimensionality reduction technique on spectra to obtain spatial-spectral features. Yue et al. [311] generated spectral feature maps by performing outer products between different segments of the pixel's spectra and stacked it with spatial feature maps consisting of dominant PCA bands of image patch around the pixel to form spatial-spectral input, which was used to train a six layered convolutional neural network. In [247], features generated by sparse unsupervised convolutional neural network, trained using Enforcing Population and Lifetime Sparsity (EPLS) [248] algorithm, showed better performance for classification than the features generated by the PCA and the kernel PCA with radial basis function kernel. Ghamisi et al. [104] proposed a band selection method that chose the bands which maximized the accuracy of a 2D CNN over validation set using swarm particle optimization. Liang et al. [179] integrated 2D CNN and dictionary learning by learning sparse code from the last layers of a supervised 2D CNN, and using reconstruction loss to predict the class. They found that using dictionary learning was better than using support vector machines.

The scale of the spatial features in a hyperspectral image is highly dependent on the instantaneous field of view and the geometry of the imaging system, Zhao et al. [323] proposed a multiscale convolutional autoencoder with three PCA bands to build scale invariance for classification. Their model passes the Laplacian pyramid of three dominant PCA bands of an image through two layers of convolutional autoencoder to generate spatial features which are concatenated with spectra and fed to logistic regression. Because of scale invariance, their method performed better than state-of-the-art spatial-spectral classification algorithms, the extended morphological profile method [25] and the multilevel logistic based method [166]. Aptoula et al. [14] proposed using attribute profile [215] as input to 2-D CNN for hyperspectral image as these profiles can capture structural information in an image at various scales.

Chen et al. [53] proposed a three dimensional end-to-end convolutional neural networks to predict material class from the image patch around the test pixel. The term 'end-to-end' is used to denote networks that take raw input signal at the input layer and produce final output in the output layer and do not use any pre-processing or post-processing. Three dimensional convolutional networks are advantageous than two dimensional convolutional networks as they can directly learn spatial-spectral features as their filters span over both spatial and spectral axes. They also proposed using augmenting the training data with image patches generated by modeling the changes in illumination in the training examples and modeling linear mixing between the training examples to augment the training data. By augmenting the training data, they were able to train the proposed network without any dimensionality reduction of the image bands. Previous hyperspectral deep learning methods have relied on dimensionality reduction to decrease the number of bands in the image and hence limit the number of parameters in the model. The same idea of using 3D CNN for hyperspectral classification was again presented in [177]. This work showed that 3D CNN outperformed baseline stacked autoencoder, deep belief network, and 2D CNN. Similarly, a 3D residual [126] CNN was proposed in [329]. Residual networks have residual blocks which learn the difference between the target function and the input instead of the learning the whole target function which makes them more robust against overfitting when there are a large number of layers in

the network.

Recently, generative adversarial networks (GANs) [111] were proposed for semi-supervised classification by Zhi et al. [127]. GANs are generative models that learn to generate samples from the data distribution using two competing neural networks, one called the generator and the other called the discriminator. Zhi et al. trained a generator and a discriminator on spatial-spectral features obtained from hyperspectral image. Then, they added a softmax layer to the discriminator network at the end and fine tuned the network to perform classification. Their method was found to be effective when number of training examples were small.

An architecture with two processing streams, a 1-D CNN stream for spectral information (trained on spectrum) and a 2-D CNN stream for spatial information (trained on image patch around the pixel), was proposed in [316]. It is different from methods that concatenate spectral and spatial features before classifier in that this method jointly optimizes the spectral and the spatial feature extractors. Similarly, Yang et al. [310] performed transfer learning using a two stream spatial-spectral network. They trained the network using two fully annotated source images, fixed the weights of earlier layers, and fine-tuned final layers on test images. They found that transferring weights produces better results than training the entire network from scratch on the test image. Santara et al. [252] proposed an end-to-end network that selects groups of non-overlapping bands, processes them in identical parallel processing streams, and combines them to produce the final output.

A fully convolutional network is an end-to-end deep network with only convolutional layers and no fully connected layers. It is trained to map an arbitrary sized image directly to its classification map. The method by Jiao et al. [146] used FCN-8 [261] network pretrained on RGB ground based images to generate multi-scale features, which were then weighted and concatenated with spectrum to form spatial-spectral features for classification. Since, FCN-8 is trained on 3 bands (Red-Green-Blue), this method only use 3 dominant PCA bands of hyperspectral as input. In [163], a novel fully convolutional architecture was proposed for hyperspectral images. The network was trained on hyperspectral image patches as input and corresponding ground truth map of the image patches as output. This network used a multi-scale convolutional filter bank to extract features at multiple scales in the first layer. The network also contained skip connections for residual learning [126] and data augmentation (mirroring along horizontal, vertical, diagonal axes) to make optimization easier. Due to fully convolutional nature of the network, at test time the entire test image was passed into the network to generate classification map the whole image simultaneously.

Tao et al. [269] proposed using two layered stacked sparse autoencoders (an autoencoder which promotes sparsity of the activation in hidden layer during training) to learn multiscale spatial-spectral features in an unsupervised manner. Their method applied PCA on an hyperspectral image and extracted random square patches of different sizes to learn a set of autoencoders from the dominant principal component bands at different scales. Results from all the autoencoders were concatenated to obtained a final feature vector which was used for classification with a linear SVM. This was the first study that investigated the transferability of features in hyperspectral images. They found that features learned on separate source image can be as good as features learned on the test image itself. Similarly, Kemker and Kanan [151] proposed using multi-scale ICA and stacked convolutional autoencoder to learn unsupervised spatial-spectral representation from images patches. Compared to [269], they used a larger training set from multiple sensors for feature learning. On the other hand, studies [310, 196] have investigated the idea of pretraining a CNN on an annotated source hyperspectral image and then finetuning only the end layers on the test hyperspectral image to learn spatial-spectral representation in a supervised manner. A domain adaptation based method that learns features that are invariant to the difference in the distribution of the source image and the destination image was proposed in [89]. Their method uses an end-to-end domain-adversarial neural network [100] to learn features that maximizes the separability of material classes but minimizes the separability based on whether the features came from the source image or the destination image. It was found that this feature was better than feature learned from applying denoising autoencoder, PCA, and KPCA on source image and baseline deep networks trained on the source and the destination image.

Even though all the algorithms discusses in this section have been classification algorithms, deep learning can also be used for other tasks. A deep and transfer learning based anomaly detector has been recently published in [172]. In this method, a two-class one dimensional CNN was trained to detect the dissimilarity between two spectra. Training set was generated by selecting pairs of spectra from a fully annotated separate training image. The pair of spectra were assigned to 0-class if they belonged to same material and 1-class if they belonged to different materials. During testing, each pixel in the test image was compared to all its neighbors and the score obtained from the network was averaged. Anomalies

were then detected by thresholding this score.

Recurrent neural networks (RNN) are popular architectures for modeling sequential data. They contain feedback loops in their computation allowing the current output to be dependent on the current input and the previous inputs. This is different from all the networks we have discussed previously which used feedforward computations to produce output. Mou et al. [210] proposed using RNN to model pixel spectra in a hyperspectral image as a 1-D sequences for classification. They experimented with architectures based on two kinds of recurrent units, namely, long short-term memory [112] (LSTM) and gated recurrent unit [63] (GRN). They found that the GRN worked better than the LSTM for modeling hyperspectral data and both of the recurrent networks outperformed traditional approaches and baseline CNN. Similarly, Wu et al. [302] showed that a convolutional RNN [332] (a network that has few convolutional layers followed by RNN) is better choice for spectra classification than LSTM and baseline CNN.

# 6 Open Issues and Future Challenges

## 6.1 Curse of dimensionality

The high dimensionality of hyperspectral data is a well-documented problem in remote sensing. Some of the technique proposed to tackle it are dimensionality reduction [260], fusion of spatial information [97], transfer and multitask learning [310, 282], and supplementation of training data with synthetic examples [53]. However, the problem of high dimensionality is far from being solved. Because hyperspectral images are oversampled in spectral dimension, redundancies exist between the bands of the hyperspectral image. A big question is what is the intrinsic dimension of hyperspectral data [46], i.e., the minimum number of bands required to carry all the information in an image. The quantification of the intrinsic dimensionality would not only lead to development more efficient analysis algorithms, but also aid in the development of efficient hyperspectral sensors with optimized number of bands and sampling.

## 6.2 Robustness and reliability of models

The grand challenge of hyperspectral data analysis is to build models that are invariant to the differences in the time and season of image acquisition, the site, the platform, the spatial resolution, the spectral resolution and sampling intervals, and the sensor technology. It is an ambitious task as there are massive variability in the spectra when any one those factors is even slightly altered. Therefore, studies so far have mostly concentrated on developing models that work for a particular image. These models are typically trained using the labeled pixels in the test image itself or using the ground spectra and the chemical analysis of samples collected over the imaged site. However, the eventual goal is to build universal models that once trained, can be applied to other images taken under conditions different from the original directly or with slight modifications.

Interest towards such models is beginning to grow in the remote sensing community, as seen by the contest problem in the 2017 IEEE Geoscience and Remote Sensing (GRSS) data fusion competition[2]. The task in the competition is to map the land use of new test cities using ground truth land use information of separate cites in the training images using multispectral data collected by different satellites at different seasons. Once the satellite based hyperspectral imaging becomes more mature, such datasets will be available for hyperspectral images as well and make it possible to classify land covers into finer classes than it is currently possible with multispectral images.

## 6.3 Big data without ground truth

After the initial cost of the instruments, it is often cheaper to obtain large quantity of hyperspectral images than to collect ground truth information for even a small area. For some of the ground truth information, such as chemical composition of the materials that is needed for unmixing and physical parameter estimation, it is only practical to obtained ground truth for few samples collected from the scene. This has lead to availability of large databases of images without ground truth or with very limited ground truth. With the increasing quality and miniaturization of sensors with the decreasing costs, this amount is only expected to grow at an exponential rate as unmanned aerial vehicle-based and satellite-based imaging becomes more popular.

---

[2]http://www.grss-ieee.org/community/technical-committees/data-fusion/

The availability of this massive amount of ground truth-less images raises an interesting question. Is it possible to utilize the huge quantity of unlabeled images in improving the data analysis models? The current unsupervised, semi-supervised and active learning algorithms may not be able to handle the huge volume of data, so there is a necessity of development of large-scale algorithms to utilize unlabeled images.

## 6.4 Lack of standardized datasets and experiments

There is a lack of benchmarking dataset for hyperspectral analysis. Without a set standard procedure to evaluate methods under real-world scenarios, researchers and practitioners cannot make an educated choice in picking a right method for their problem. Since, researchers have been using different datasets with different experimental conditions, it is impossible to compare two methods proposed in two different papers. On top of that, many times it is difficult to reproduce a study because the implementation of the methods is not readily available. These factors stifle the possibilities of follow up work and adoption of the methods. The IEEE Geoscience and Remote Sensing society's (IEEE GRSS) Data and Algorithm Standard Evaluation website[3], IEEE GRSS annual data fusion contest, and Rochester Institute of Technology's target detection blind test website[4] are few of the current efforts to provide benchmarks to compare different hyperspectral analysis methods.

There are only few publicly available standard datasets for hyperspectral analysis. Even though they are handful in number, datasets for land cover classification are most commonly found. Almost all new classification methods are tested on them. However, these test images are small compared to real-life images and lack variability and diversity. Most of them were captured by same sensor as well. This raises a question that whether current methods which work great on these few and small images generalize to performing good on large real-world images. There are even fewer datasets for other tasks. Most of unmixing algorithms demonstrate their efficacy on simulated data as there are almost none publicly unmixing datasets with ground truth abundances (one available dataset is [5]). Researchers working on target detection and physical parameter estimation tend to use their own data rather than standard datasets, although few standard datasets are available for these tasks. Hyperspectral change detection seems to have no standard datasets. Out of many reasons, this has contributed to more methods being developed for land cover classification, than any of the other data analysis tasks. This is evidenced by the shear number of land cover classification methods published compared to others. More datasets related to target detection, unmixing, and physical parameter estimation need to be prepared. We also need more diverse datasets for land cover classification. Development of newer datasets should be one of the top priorities of the hyperspectral remote sensing community.

## 7   Summary

Over the years, machine learning has become the primary tool for analysis of hyperspectral images. With the literature booming with new methods, choosing the right method for a problem can become a daunting task. This paper addressed this issue by creating a catalog of published methods. We have summarized all of the discussed methods in Table 1.

The current hot topics in hyperspectral data analysis are deep learning and data fusion. Deep learning has been primarily tested for land cover classification. Hyperspectral datasets are infamous for having small amount of ground truth data while the deep learning methods are infamous for requiring large amount of ground truth data. Current methods get around this problem by training the network on labeled pixels or small patches of an image and testing on the remaining unlabeled pixels in the same image, instead of training and testing on separate images. They additionally employ regularization schemes, such as data augmentation and early stopping, to prevent overfitting. Since, the current models are learned on training sets consisting of spectra from a single image, which is typically small, these models cannot capture the spectral variabilities that occur due to differences in factors, such as illumination, atmospheric conditions, sun angle, viewing geometry, and resolution. If there were large-scale datasets with labeled images covering diverse scenes for training, it is highly likely that the networks would be able to learn a spectral variabilities resistant representation of the data, and would perform well when tested on unseen new images. This highlights the fact that there is an urgent need of large-scale datasets for land cover classification. The arrival of such datasets could have as powerful of an impact to the field

---

[3]`http://dase.ticinumaerospace.com/`
[4]`http://dirsapps.cis.rit.edu/blindtest/`
[5]`http://www.planetary.brown.edu/relabdocs/LSCCsoil.html`

Table 1: Summary of all reviewed methods.

| Methods | Classification | | Target Detection | | Unmixing | | Parameter Estimation |
|---|---|---|---|---|---|---|---|
| | Pixel-wise | Spatial-spectral | Target | Anomaly | Linear | Non-linear | |
| Gaussian Models | [69, 19, 167], transfer learning[234] | | [192] | [47] | [128] | [130] | band selection [293, 278, 155, 156] |
| Linear Regression | | | | | | | |
| Logistic Regression | [153], band selection[58, 327, 230, 303] | [239, 134], semi-supervised[168, 80] | | | | | |
| Support Vector Machines | [132, 197, 33, 75, 203, 283, 116, 298], band selection[21, 228], semi-supervised[61], transfer learning[266, 233], change detection[223] | [25, 93, 115, 182, 171, 41, 91] | [249] | [20, 152, 118] | [204], endmember and sub-pixel mapping[287] | [296, 117] | semi-supervised[38],active learning[231],multi-output[282] |
| Gaussain Mixture Models | [83, 176, 175], unsupervised [271, 259], band selection [263] | [309, 271] | | [272, 84] | | | |
| Latent Linear Models | dimensionality reduction [260, 164, 96, 224, 294], unsupervised [82, 259] | feature extraction [236, 68] | | | [221] | | [43, 109, 65, 122, 71] |
| Sparse Linear Models | [251, 44], feature extraction [49] | [181, 50, 52, 317], feature extraction [81, 90] | spatial-spectral [51] | | [138, 275, 113, 185], spatial-spectral[139, 140, 141, 44] | | |
| Ensemble Learning | [299, 56, 119, 241, 62, 150, 121, 70, 306], transfer learning [240], transfer and active learning [193] | [201, 250, 304, 305] | | [232] | | | band selection [4, 219, 3, 142] |
| Directed Graphical Models | | | | | [79, 275, 85, 213, 78, 254], spatial-spectral [87, 86], sub-pixel mapping [319] | [120, 9, 8], spatial-spectral [11, 10] | |
| Undirected Graphical Models | | [325, 32, 270, 307, 201, 165, 326, 108, 175, 170, 143, 315, 328, 320], band selection [324], semi-supervised [166], active learning [167, 265], sub-pixel mapping [297] | | | | | |
| Clustering | unsupervised [288, 135, 17, 220, 149], band selection [145, 308], dimensionality reduction [263], semi-supervised | | | [84] | | | |
| Gaussian Processes | [23, 198, 94], active learning [264] | [148] | | | endmember extraction [149, 206, 313, 314] | endmember extraction[7] | [286, 285, 160, 217] |
| Dirichlet Processes | | [149] | | | | | |
| Deep Learning | supervised feature learning[210, 302], unsupervised feature learning[289, 54, 55] | unsupervised feature learning [54, 55, 183, 323], supervised feature learning [131, 247, 53, 322, 311, 187, 188, 177, 14, 316, 146, 163, 329, 252, 173, 179], semi-supervised[127], supervised transfer learning [310, 196, 89], unsupervised transfer learning [269, 151, 89], supervised band selection[104] | | [172] | | | |

of hyperspectral image analysis as the development of ImageNet [76] had for the growth of computer vision and deep learning.

While the development of robust supervised deep learning methods is contingent on the development of large-scale datasets, unsupervised deep learning can learn robust representation of hyperspectral data using vast amount of already available unlabeled images. Unsupervised learning is also more suited for other hyperspectral image analysis tasks, for which large-scale datasets are unlikely to be available due to the cost associated with collection of ground truth. Studies have already investigated the transferability of learned unsupervised features between images of diverse scenes [269, 151]. Generative adversarial networks (GANs) and variational autoencoders (VAs) look very promising for modeling unlabeled hyperspectral data. GANs and VAs could characterize the spectral variability by modeling the generative distribution of the spectra. Such generative models, in turn, can be used as prior for classification algorithms to make them invariant to spectral variabilities. If such generative models are conditioned on the set of biophysical parameters of the material, they could be used as data-driven forward models in lieu of physics-based radiative transfer models. These models could find usage in non-linear unmixing and physical parameter estimation. Similarly, the generative models conditioned on material class can also serve as a spectral library. GANs and VAs could also model the spatial prior of the land covers (similar to how GANs were used in [186]). Deep learning based spatial prior could turn out to be better than Markov random fields based prior for land cover classification. GANs and VAs models for hyperspectral images should also prove to be good for image processing tasks, such as pansharpening, superresolution, denoising, and inpainting.

Multi-sensor, multi-resolution, multi-temporal, multi-modal, multi-platform, and multi-site data fusion still remains an open problem in remote sensing. In our opinion, deep learning methods are best suited for problems where multiple sources of imaging data have to be combined. Deep networks, in particular convolutional neural networks, can be trained to generate features which combine relevant aspect of all sources at each pixel locations, which can be further be processed for desired task. Such architecture has been investigated for fusion of hyperspectral imagery with lidar data [105], but similar architectures can be used to combine multi-sensor, multi-resolution, multi-temporal, multi-modal, and multi-platform images of the same scene. Another kind of data fusion would be to incorporate non-imaging data with imaging data. For instance, information from digital maps and geotagged user data could be used alongside information from remotely observed hyperspectral images for land cover mapping of urban areas. Due to their capabilities to model semantic relationships between different variables, conditional random models (CRFs) could be the best for this kind of fusion of information. In fact, recently a higher order CRF was used to model the dependencies between land use and land cover of a remotely imaged area [6].

# References

[1] ISPRS benchmarks. `http://www.isprs.org/education/benchmarks.aspx`.

[2] National ecological observatory network. `http://data.neonscience.org`. National Ecological Observatory Network, Battelle, Boulder, CO, USA.

[3] Elfatih M. Abdel-Rahman, Fethi B. Ahmed, and Riyad Ismail. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34(2):712–728, 2013.

[4] Elhadi Adam, Onisimo Mutanga, Elfatih M. Abdel-Rahman, and Riyad Ismail. Estimating standing biomass in papyrus (Cyperus papyrus l.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, 35(2):693–714, 2014.

[5] Amir H. Alavi, Amir H. Gandomi, and David J. Lary. Progress of machine learning in geosciences: Preface. *Geoscience Frontiers*, 7(1):1 – 2, 2016. Special Issue: Progress of Machine Learning in Geosciences.

[6] Lena Albert, Franz Rottensteiner, and Christian Heipke. A higher order conditional random field model for simultaneous classification of land cover and land use. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:63–80, 2017.

[7] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret. Nonlinear spectral unmixing of hyperspectral images using Gaussian processes. *IEEE Transactions on Signal Processing*, 61(10):2442–2453, May 2013.

[8] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret. Unsupervised post-nonlinear unmixing of hyperspectral images using a hamiltonian monte carlo algorithm. *IEEE Transactions on Image Processing*, 23(6):2663–2675, June 2014.

[9] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Transactions on Image Processing*, 21(6):3017–3025, June 2012.

[10] Y. Altmann, M. Pereyra, and S. McLaughlin. Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis. *IEEE Transactions on Computational Imaging*, 1(3):174–185, Sept 2015.

[11] Yoann Altmann, Nicolas Dobigeon, Steve McLaughlin, and Jean-Yves Tourneret. Residual component analysis of hyperspectral imagesapplication to joint nonlinear unmixing and nonlinearity detection. *IEEE Transactions on Image Processing*, 23(5):2148–2158, 2014.

[12] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[13] Naveen J.P. Anne, Amr H. Abd-Elrahman, David B. Lewis, and Nicole A. Hewitt. Modeling soil parameters using hyperspectral image reflectance in subtropical coastal wetlands. *International Journal of Applied Earth Observation and Geoinformation*, 33:47 – 56, 2014.

[14] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu. Deep learning with attribute profiles for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1970–1974, Dec 2016.

[15] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[16] Peter Bajcsy and Peter Groves. Methodology for hyperspectral band selection. *Photogrammetric Engineering & Remote Sensing*, 70(7):793–802, 2004.

[17] Claire A Baldeck and Gregory P Asner. Estimating vegetation beta diversity from airborne imaging spectroscopy and unsupervised clustering. *Remote Sensing*, 5(5):2057–2071, 2013.

[18] Geoffrey H Ball and David J Hall. ISODATA, a novel method of data analysis and pattern classification. Technical report, DTIC Document, 1965.

[19] T.V. Bandos, L. Bruzzone, and G. Camps-Valls. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3):862–873, March 2009.

[20] A. Banerjee, P. Burlina, and C. Diehl. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8):2282–2291, Aug 2006.

[21] Y. Bazi and F. Melgani. Toward an optimal SVM classification system for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3374–3385, Nov 2006.

[22] Y. Bazi and F. Melgani. Semisupervised PSO-SVM regression for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1887–1895, June 2007.

[23] Y. Bazi and F. Melgani. Gaussian process approach to remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):186–197, 2010.

[24] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.

[25] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, March 2005.

[26] Yoshua Bengio, Aaron Courville, and Pierre Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[27] Alexander Berk, Lawrence S Bernstein, and David C Robertson. MODTRAN: A moderate resolution model for LOWTRAN. Technical report, DTIC Document, 1987.

[28] Jeff A Bilmes et al. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510):126, 1998.

[29] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[30] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[31] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov random fields for vision and image processing*. Mit Press, 2011.

[32] J.S. Borges, J.M. Bioucas-Dias, and A.R.S. Marcal. Bayesian hyperspectral image segmentation with discriminative class learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2151–2164, June 2011.

[33] A.C. Braun, U. Weidner, and S. Hinz. Classification in high-dimensional feature spaces - assessment using SVM, IVM and RVM with focus on simulated EnMAP data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):436–443, April 2012.

[34] L. Bruzzone and F. Melgani. Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1):159–174, Jan 2005.

[35] G. Camps-Valls. Machine learning in remote sensing data processing. In *IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6, Sept 2009.

[36] G. Camps-Valls, J. Bioucas-Dias, and M. Crawford. A special issue on advances in machine learning for remote sensing and geosciences. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):5–7, 2016.

[37] G. Camps-Valls, L. Bruzzone, J.L. Rojo-Alvarez, and F. Melgani. Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 3(3):339–343, July 2006.

[38] G. Camps-Valls, J. Munoz-Mari, L. Gomez-Chova, K. Richter, and J. Calpe-Maravilla. Biophysical parameter estimation with a semisupervised support vector machine. *IEEE Geoscience and Remote Sensing Letters*, 6(2):248–252, April 2009.

[39] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1):45–54, Jan 2014.

[40] G. Camps-Valls, J. Verrelst, J. Munoz-Mari, V. Laparra, F. Mateo-Jimenez, and J. Gomez-Dans. A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):58–78, June 2016.

[41] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muñoz-Marí, Joan Vila-Francés, and Javier Calpe-Maravilla. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 3(1):93–97, 2006.

[42] Gustavo Camps-Valls, Tatyana V Bandos Marsheva, and Dengyong Zhou. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3044–3054, 2007.

[43] Luis M. Carrascal, Ismael Galvn, and Oscar Gordo. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118(5):681–690, 2009.

[44] A. Castrodad, Zhengming Xing, J.B. Greer, E. Bosch, L. Carin, and G. Sapiro. Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4263–4281, Nov 2011.

[45] Chein-I Chang. *Hyperspectral data processing: algorithm design and analysis*. John Wiley & Sons, 2013.

[46] Chein-I Chang. A review of virtual dimensionality for hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018.

[47] Chein-I Chang and Shao-Shan Chiang. Anomaly detection and classification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(6):1314–1325, 2002.

[48] Chein-I Chang and Antonio Plaza. A fast iterative algorithm for implementation of pixel purity index. *IEEE Geoscience and Remote Sensing Letters*, 3(1):63–67, 2006.

[49] A.S. Charles, B.A. Olshausen, and C.J. Rozell. Learning sparse codes for hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):963–978, Sept 2011.

[50] Yi Chen, N.M. Nasrabadi, and T.D. Tran. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985, Oct 2011.

[51] Yi Chen, N.M. Nasrabadi, and T.D. Tran. Sparse representation for target detection in hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):629–640, June 2011.

[52] Yi Chen, N.M. Nasrabadi, and T.D. Tran. Hyperspectral image classification via kernel sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):217–231, Jan 2013.

[53] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.

[54] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, June 2014.

[55] Yushi Chen, Xing Zhao, and Xiuping Jia. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2381–2392, June 2015.

[56] Yushi Chen, Xing Zhao, and Zhouhan Lin. Optimizing subspace SVM ensemble for hyperspectral imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1295–1305, April 2014.

[57] Gong Cheng and Junwei Han. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:11–28, 2016.

[58] Qi Cheng, P.K. Varshney, and M.K. Arora. Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 3(4):491–494, Oct 2006.

[59] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[60] M. Chi, A. J. Plaza, J. A. Benediktsson, B. Zhang, and B. Huang. Foreword to the special issue on big data in remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4607–4609, Oct 2015.

[61] Mingmin Chi and L. Bruzzone. Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1870–1880, June 2007.

[62] Mingmin Chi, Qian Kun, J.A. Benediktsson, and Rui Feng. Ensemble classification algorithm for hyperspectral remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 6(4):762–766, Oct 2009.

[63] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[64] Moses Azong Cho, Andrew Skidmore, Fabio Corsi, Sipke E. van Wieren, and Istiak Sobhan. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9(4):414 – 424, 2007.

[65] Moses Azong Cho, Andrew Skidmore, Fabio Corsi, Sipke E. van Wieren, and Istiak Sobhan. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, 9(4):414 – 424, 2007.

[66] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Boosting for transfer learning. In *International conference on Machine learning (ICML)*, pages 193–200. ACM, 2007.

[67] Laura M Dale, André Thewis, Christelle Boudry, Ioan Rotar, Pierre Dardenne, Vincent Baeten, and Juan A Fernández Pierna. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review. *Applied Spectroscopy Reviews*, 48(2):142–159, 2013.

[68] Mauro Dalla Mura, Alberto Villa, Jon Atli Benediktsson, Jocelyn Chanussot, and Lorenzo Bruzzone. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3):542–546, 2011.

[69] M. Dalponte, H.O. Orka, T. Gobakken, D. Gianelle, and E. Naesset. Tree species classification in boreal forests with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2632–2645, May 2013.

[70] B.B. Damodaran, R.R. Nidamanuri, and Y. Tarabalka. Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2405–2417, June 2015.

[71] R. DE BEI, D. COZZOLINO, W. SULLIVAN, W. CYNKAR, S. FUENTES, R. DAMBERGS, J. PECH, and S. TYERMAN. Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian Journal of Grape and Wine Research*, 17(1):62–71, 2011.

[72] Christian Debes, Andreas Merentitis, Roel Heremans, Jürgen Hahn, Nikolaos Frangiadakis, Tim van Kasteren, Wenzhi Liao, Rik Bellens, Aleksandra Pižurica, Sidharta Gautama, et al. Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2405–2418, 2014.

[73] Sahar Dehnavi, Yasser Maghsoudi, and Mohammadjavad Valadanzoej. Using spectrum differentiation and combination for target detection of minerals. *International Journal of Applied Earth Observation and Geoinformation*, 55:9 – 20, 2017.

[74] F. Dell'Acqua, P. Gamba, A. Ferrari, J.A. Palmason, J.A. Benediktsson, and K. Arnason. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, 1(4):322–326, Oct 2004.

[75] B. Demir and S. Erturk. Hyperspectral image classification using relevance vector machines. *IEEE Geoscience and Remote Sensing Letters*, 4(4):586–590, Oct 2007.

[76] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.

[77] Baudouin Descle, Patrick Bogaert, and Pierre Defourny. Forest change detection by statistical object-based method. *Remote Sensing of Environment*, 102(12):1 – 11, 2006.

[78] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A.O. Hero. Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 57(11):4355–4368, Nov 2009.

[79] N. Dobigeon, J.-Y. Tourneret, and Chein-I Chang. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 56(7):2684–2695, July 2008.

[80] I. Dopido, Jun Li, P. Gamba, and A. Plaza. A new hybrid strategy combining semisupervised classification and unmixing of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8):3619–3629, Aug 2014.

[81] Peijun Du, Zhaohui Xue, Jun Li, and A. Plaza. Learning discriminative sparse representations for hyperspectral image classification. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1089–1104, Sept 2015.

[82] Qian Du, Ivica Kopriva, and Harold Szu. Independent-component analysis for hyperspectral remote sensing imagery classification. *Optical Engineering*, 45(1):017008–017008, 2006.

[83] M Murat Dundar and David Landgrebe. A model-based mixture-supervised classification approach in hyperspectral data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2692–2699, 2002.

[84] O. Duran and M. Petrou. A time-efficient method for anomaly detection in hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3894–3904, Dec 2007.

[85] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret. Bayesian estimation of linear mixtures using the normal compositional model. application to hyperspectral imagery. *IEEE Transactions on Image Processing*, 19(6):1403–1413, June 2010.

[86] Olivier Eches, Jon Atli Benediktsson, Nicolas Dobigeon, and Jean-Yves Tourneret. Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images. *IEEE Transactions on Image Processing*, 22(1):5–16, 2013.

[87] Olivier Eches, Nicolas Dobigeon, and Jean-Yves Tourneret. Enhancing hyperspectral image unmixing with spatial correlations. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4239–4247, 2011.

[88] Michael Theodore Eismann. *Hyperspectral remote sensing*. SPIE Press Bellingham, 2012.

[89] A. Elshamli, G. W. Taylor, A. Berg, and S. Areibi. Domain adaptation using representation learning for the classification of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9):4198–4209, Sept 2017.

[90] J. Fan, T. Chen, and S. Lu. Superpixel guided deep-sparse-representation learning for hyperspectral image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–1, 2017.

[91] Leyuan Fang, Shutao Li, Wuhui Duan, Jinchang Ren, and Jón Atli Benediktsson. Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12):6663–6674, 2015.

[92] Michael D Farrell and Russell M Mersereau. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geoscience and Remote Sensing Letters*, 2(2):192–195, 2005.

[93] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, Nov 2008.

[94] M. Fauvel, C. Bouveyron, and S. Girard. Parsimonious Gaussian process models for the classification of hyperspectral remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2423–2427, Dec 2015.

[95] Mathieu Fauvel, Jocelyn Chanussot, and Jon Atli Benediktsson. Evaluation of kernels for multiclass classification of hyperspectral remote sensing data. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 2, pages II–II. IEEE, 2006.

[96] Mathieu Fauvel, Jocelyn Chanussot, and Jón Atli Benediktsson. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP J. Adv. Signal Process*, 2009:11:1–11:14, January 2009.

[97] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[98] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.

[99] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315:972–976, 2007.

[100] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

[101] Bo-Cai Gao, Marcos J. Montes, Curtiss O. Davis, and Alexander F.H. Goetz. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment*, 113, Supplement 1:S17 – S24, 2009. Imaging Spectroscopy Special Issue.

[102] Utsav B Gewali and Sildomar T Monteiro. A novel covariance function for predicting vegetation biochemistry from hyperspectral imagery with gaussian processes. In *IEEE International Conference on Image Processing (ICIP)*, pages 2216–2220. IEEE, 2016.

[103] Utsav B Gewali and Sildomar T Monteiro. A tutorial on modeling and inference in undirected graphical models for hyperspectral image analysis. *arXiv preprint arXiv:1801.08268*, 2018.

[104] P. Ghamisi, Y. Chen, and X. X. Zhu. A self-improving convolution neural network for the classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 13(10):1537–1541, Oct 2016.

[105] Pedram Ghamisi, Bernhard Höfle, and Xiao Xiang Zhu. Hyperspectral and lidar data fusion using extinction profiles and deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6):3011–3024, 2017.

[106] Azadeh Ghiyamat and Helmi ZM Shafri. A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing*, 31(7):1837–1856, 2010.

[107] Habes A. Ghrefat, Philip C. Goodell, Bernard E. Hubbard, Richard P. Langford, and Raed E. Aldouri. Modeling grain size variations of aeolian gypsum deposits at white sands, new mexico, using {AVIRIS} imagery. *Geomorphology*, 88(12):57 – 68, 2007.

[108] M. Golipour, H. Ghassemian, and F. Mirzapour. Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–12, 2015.

[109] Ccile Gomez, Raphael A. Viscarra Rossel, and Alex B. McBratney. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma*, 146(34):403 – 411, 2008.

[110] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[111] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[112] Alex Graves and Jrgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18(5):602 – 610, 2005. IJCNN 2005.

[113] J.B. Greer. Sparse demixing of hyperspectral images. *IEEE Transactions on Image Processing*, 21(1):219–228, Jan 2012.

[114] Yanfeng Gu and Huan Liu. Sample-screening {MKL} method via boosting strategy for hyperspectral image classification. *Neurocomputing*, 173, Part 3:1630 – 1639, 2016.

[115] Yanfeng Gu, Tianzhu Liu, Xiuping Jia, Jón Atli Benediktsson, and Jocelyn Chanussot. Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3235–3247, 2016.

[116] Yanfeng Gu, Chen Wang, Di You, Yuhang Zhang, Shizhe Wang, and Ye Zhang. Representative multiple kernel learning for classification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7):2852–2865, 2012.

[117] Yanfeng Gu, Shizhe Wang, and Xiuping Jia. Spectral unmixing in multiple-kernel Hilbert space for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):3968–3981, July 2013.

[118] P. Gurram and Heesung Kwon. Support-vector-based hyperspectral anomaly detection using optimized kernel parameters. *IEEE Geoscience and Remote Sensing Letters*, 8(6):1060–1064, Nov 2011.

[119] P. Gurram and Heesung Kwon. Sparse kernel-based ensemble learning with fully optimized kernel parameters for hyperspectral classification problems. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):787–802, Feb 2013.

[120] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4153–4162, Nov 2011.

[121] Jisoo Ham, Yangchi Chen, M.M. Crawford, and J. Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, March 2005.

[122] P.M. Hansen and J.K. Schjoerring. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4):542 – 553, 2003.

[123] M. Hasanlou and F. Samadzadegan. Comparative study of intrinsic dimensionality estimation and dimension reduction techniques on hyperspectral images using k-nn classifier. *IEEE Geoscience and Remote Sensing Letters*, 9(6):1046–1050, Nov 2012.

[124] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Overview of supervised learning. In *The elements of statistical learning*, pages 9–39. Springer, 2009.

[125] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Unsupervised learning. In *The elements of statistical learning*, pages 485–585. Springer, 2009.

[126] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[127] Zhi He, Han Liu, Yiwen Wang, and Jie Hu. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sensing*, 9(10), 2017.

[128] D.C. Heinz and Chein-I Chang. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, Mar 2001.

[129] R. Heylen, M. Parente, and P. Gader. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):1844–1868, June 2014.

[130] R. Heylen, P. Scheunders, A. Rangarajan, and P. Gader. Nonlinear unmixing by using different metrics in a linear unmixing chain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2655–2664, June 2015.

[131] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 501:258619, 2015.

[132] C Huang, LS Davis, and JRG Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, 2002.

[133] Guang-Bin Huang, Dian Hui Wang, and Yuan Lan. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2(2):107–122, 2011.

[134] Xin Huang, Xuehua Guan, J.A. Benediktsson, Liangpei Zhang, Jun Li, A. Plaza, and M. Dalla Mura. Multiple morphological profiles from multicomponent-base images for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4653–4669, Dec 2014.

[135] Xin Huang and Liangpei Zhang. An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12):4173–4185, 2008.

[136] Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:91–106, 2013.

[137] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[138] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, June 2011.

[139] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4484–4502, Nov 2012.

[140] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza. Collaborative sparse regression for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):341–354, Jan 2014.

[141] M.-D. Iordache, J.M. Bioucas-Dias, A. Plaza, and B. Somers. MUSIC-CSR: Hyperspectral unmixing via multiple signal classification and collaborative sparse regression. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7):4364–4382, July 2014.

[142] Riyad Ismail and Onisimo Mutanga. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, 12:S45–S51, 2010.

[143] Qiong Jackson and David A Landgrebe. Adaptive bayesian contextual classification based on markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2454–2463, 2002.

[144] Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

[145] Sen Jia, Zhen Ji, Yuntao Qian, and Linlin Shen. Unsupervised band selection for hyperspectral imagery classification without manual band removal. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):531–543, 2012.

[146] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10):5585–5599, Oct 2017.

[147] Ian Jolliffe. *Principal Component Analysis*. John Wiley & Sons, Ltd, 2014.

[148] G. Jun and J. Ghosh. Spatially adaptive classification of land cover with remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(7):2662–2673, 2011.

[149] Goo Jun and J. Ghosh. Semisupervised learning of hyperspectral data with unknown land-cover classes. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):273–282, Jan 2013.

[150] S. Kawaguchi and R. Nishii. Hyperspectral image classification by bootstrap adaboost with random decision stumps. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3845–3851, Nov 2007.

[151] R. Kemker and C. Kanan. Self-taught feature learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2693–2705, May 2017.

[152] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi. Anomaly detection in hyperspectral images based on an adaptive support vector method. *IEEE Geoscience and Remote Sensing Letters*, 8(4):646–650, July 2011.

[153] M. Khodadadzadeh, Jun Li, A. Plaza, and J.M. Bioucas-Dias. A subspace-based multinomial logistic regression for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 11(12):2105–2109, Dec 2014.

[154] Pushmeet Kohli, Philip HS Torr, et al. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009.

[155] Raymond F Kokaly, Gregory P Asner, Scott V Ollinger, Mary E Martin, and Carol A Wessman. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, 113:S78–S91, 2009.

[156] Raymond F. Kokaly and Roger N. Clark. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, 67(3):267 – 287, 1999.

[157] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 1097–1105, 2012.

[158] David J. Lary, Amir H. Alavi, Amir H. Gandomi, and Annette L. Walker. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3 – 10, 2016. Special Issue: Progress of Machine Learning in Geosciences.

[159] Neil D Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in neural information processing systems (NIPS)*, 16(3):329–336, 2004.

[160] M. Lazaro-Gredilla, M.K. Titsias, J. Verrelst, and G. Camps-Valls. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. *IEEE Geoscience and Remote Sensing Letters*, 11(4):838–842, April 2014.

[161] Nicolas Le Roux and Yoshua Bengio. Representational power of restricted Boltzmann machines and deep belief networks. *Neural computation*, 20(6):1631–1649, 2008.

[162] Yann LeCun, LD Jackel, Léon Bottou, Corinna Cortes, John S Denker, Harris Drucker, Isabelle Guyon, UA Muller, E Sackinger, Patrice Simard, et al. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261:276, 1995.

[163] H. Lee and H. Kwon. Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing*, 26(10):4843–4855, Oct 2017.

[164] James B Lee, A Stephen Woodyatt, and Mark Berman. Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3):295–304, 1990.

[165] Fan Li, Linlin Xu, P. Siva, A. Wong, and D.A. Clausi. Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2427–2438, June 2015.

[166] J. Li, J.M. Bioucas-Dias, and A. Plaza. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4085–4098, Nov 2010.

[167] Jun Li, J.M. Bioucas-Dias, and A. Plaza. Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3947–3960, Oct 2011.

[168] Jun Li, J.M. Bioucas-Dias, and A. Plaza. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geoscience and Remote Sensing Letters*, 10(2):318–322, March 2013.

[169] Jun Li, José M Bioucas-Dias, and Antonio Plaza. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):809–823, 2012.

[170] Jun Li, José M Bioucas-Dias, and Antonio Plaza. Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):844–856, 2013.

[171] Jun Li, Xin Huang, Paolo Gamba, José M Bioucas-Dias, Liangpei Zhang, Jón Atli Benediktsson, and Antonio Plaza. Multiple feature learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1592–1606, 2015.

[172] W. Li, G. Wu, and Q. Du. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(5):597–601, May 2017.

[173] W. Li, G. Wu, F. Zhang, and Q. Du. Hyperspectral image classification using deep pixel-pair features. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):844–853, Feb 2017.

[174] Wei Li, S. Prasad, J.E. Fowler, and L.M. Bruce. Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 8(5):894–898, Sept 2011.

[175] Wei Li, Saurabh Prasad, and James E Fowler. Hyperspectral image classification using Gaussian mixture models and Markov random fields. *IEEE Geoscience and Remote Sensing Letters*, 11(1):153–157, 2014.

[176] Wei Li, Saurabh Prasad, James E Fowler, and Lori Mann Bruce. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1185–1198, 2012.

[177] Ying Li, Haokui Zhang, and Qiang Shen. Spectralspatial classification of hyperspectral imagery with 3d convolutional neural network. *Remote Sensing*, 9(1), 2017.

[178] Zhenguo Li, Xiao-Ming Wu, and Shih-Fu Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 789–796. IEEE, 2012.

[179] Heming Liang and Qi Li. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sensing*, 8(2), 2016.

[180] Wenzhi Liao, Xin Huang, Frieke Van Coillie, Sidharta Gautama, Aleksandra Pižurica, Wilfried Philips, Hui Liu, Tingting Zhu, Michal Shimoni, Gabriele Moser, et al. Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 ieee grss data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2984–2996, 2015.

[181] Jianjun Liu, Zebin Wu, Zhihui Wei, Liang Xiao, and Le Sun. Spatial-spectral kernel sparse representation for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(6):2462–2471, Dec 2013.

[182] Tianzhu Liu, Yanfeng Gu, Xiuping Jia, Jón Atli Benediktsson, and Jocelyn Chanussot. Class-specific sparse multiple kernel learning for spectral–spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7351–7365, 2016.

[183] Yazhou Liu, Guo Cao, Quansen Sun, and Mel Siegel. Hyperspectral classification via deep networks and superpixel segmentation. *International Journal of Remote Sensing*, 36(13):3459–3482, 2015.

[184] Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.

[185] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li. Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2815–2826, May 2013.

[186] Pauline Luc, Camille Couprie, Soumith Chintala, and Jakob Verbeek. Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*, 2016.

[187] Xiaorui Ma, Jie Geng, and Hongyu Wang. Hyperspectral image classification via contextual deep learning. *EURASIP Journal on Image and Video Processing*, 2015(1):20, 2015.

[188] Xiaorui Ma, Hongyu Wang, and Jie Geng. Spectral–spatial classification of hyperspectral image based on deep auto-encoder. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4073–4085, 2016.

[189] Julien Mairal, Francis Bach, and Jean Ponce. Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, 8(2-3):85–283, December 2014.

[190] Ihab Makki, Rafic Younes, Clovis Francis, Tiziano Bianchi, and Massimo Zucchetti. A survey of landmine detection using hyperspectral imaging. {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, 124:40 – 53, 2017.

[191] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1):29–43, Jan 2002.

[192] Dimitris Manolakis, David Marden, and Gary A Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.

[193] G. Matasci, D. Tuia, and M. Kanevski. Svm-based boosting of active learning strategies for efficient domain adaptation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(5):1335–1343, Oct 2012.

[194] S. Matteoli, T. Veracini, M. Diani, and G. Corsini. A locally adaptive background density estimator: An evolution for RX-based anomaly detectors. *IEEE Geoscience and Remote Sensing Letters*, 11(1):323–327, Jan 2014.

[195] Stefania Matteoli, Marco Diani, and Giovanni Corsini. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):5–28, 2010.

[196] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4520–4533, Aug 2017.

[197] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug 2004.

[198] Arman Melkumyan and Eric Nettleton. An observation angle dependent nonstationary covariance function for Gaussian process regression. In *International Conference on Neural Information Processing*, pages 331–339. Springer, 2009.

[199] Scott Menard. *Applied logistic regression analysis*. Number 106. Sage, 2002.

[200] Grégoire Mercier and Marc Lennon. Support vector machines for hyperspectral image classification with spectral-based kernels. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 1, pages 288–290. IEEE, 2003.

[201] A. Merentitis, C. Debes, and R. Heremans. Ensemble learning in hyperspectral image classification: Toward selecting a favorable bias-variance tradeoff. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1089–1102, April 2014.

[202] T. Meyer and G.S. Okin. Evaluation of spectral unmixing techniques using {MODIS} in a structurally complex savanna environment for retrieval of green vegetation, nonphotosynthetic vegetation, and soil fractional cover. *Remote Sensing of Environment*, 161:122 – 130, 2015.

[203] F.A. Mianji and Ye Zhang. Robust hyperspectral classification using relevance vector machine. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2100–2112, June 2011.

[204] F.A. Mianji and Ye Zhang. SVM-based unmixing-to-classification conversion for hyperspectral abundance quantification. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4318–4327, Nov 2011.

[205] Ryo Michishita, Zhiben Jiang, and Bing Xu. Monitoring two decades of urbanization in the poyang lake area, china through spectral unmixing. *Remote Sensing of Environment*, 117:3 – 18, 2012. Remote Sensing of Urban Environments.

[206] R. Mittelman, N. Dobigeon, and A.O. Hero. Hyperspectral image unmixing using a multiresolution sticky HDP. *IEEE Transactions on Signal Processing*, 60(4):1656–1671, April 2012.

[207] Sildomar Takahashi Monteiro, Yohei Minekawa, Yukio Kosugi, Tsuneya Akazawa, and Kunio Oda. Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery. {ISPRS} *Journal of Photogrammetry and Remote Sensing*, 62(1):2 – 12, 2007.

[208] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis.* John Wiley & Sons, 2015.

[209] G. Moser, D. Tuia, and M. Shimoni. Report on the 2015 ieee grss data fusion contest-extremely high resolution lidar and optical data: A data fusion challenge [technical committees]. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):118–121, Sept 2015.

[210] L. Mou, P. Ghamisi, and X. X. Zhu. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3639–3655, July 2017.

[211] Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. {ISPRS} *Journal of Photogrammetry and Remote Sensing*, 66(3):247 – 259, 2011.

[212] Giorgos Mountrakis, Jungho Im, and Caesar Ogole. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259, 2011.

[213] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret. Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. *IEEE Transactions on Signal Processing*, 54(11):4133–4145, Nov 2006.

[214] CA Mucher, KT Steinnocher, FP Kressler, and C Heunks. Land cover characterization and change detection for environmental monitoring of pan-europe. *International Journal of Remote Sensing*, 21(6-7):1159–1181, 2000.

[215] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747–3762, Oct 2010.

[216] Kevin P Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.

[217] R.J. Murphy, A. Chlingaryan, and A. Melkumyan. Gaussian processes for estimating wavelength position of the ferric iron crystal field feature at ∼900 nm from hyperspectral imagery acquired in the short-wave infrared (1002 -1355 nm). *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1907–1920, April 2015.

[218] R.J. Murphy, S.T. Monteiro, and S. Schneider. Evaluating classification techniques for mapping vertical geology using field-based hyperspectral sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3066–3080, Aug 2012.

[219] Onisimo Mutanga, Elhadi Adam, and Moses Azong Cho. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 18:399 – 406, 2012.

[220] Sunil Narumalani, Deepak R Mishra, Jared Burkholder, Paul BT Merani, and Gary Willson. A comparative evaluation of ISODATA and spectral angle mapping for the detection of saltcedar using airborne hyperspectral imagery. *Geocarto International*, 21(2):59–66, 2006.

[221] José MP Nascimento and Jose MB Dias. Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Transactions on Geoscience and Remote Sensing*, 43(1):175–187, 2005.

[222] N.M. Nasrabadi. Hyperspectral target detection : An overview of current and future challenges. *IEEE Signal Processing Magazine*, 31(1):34–44, Jan 2014.

[223] Hassiba Nemmour and Youcef Chibani. Multiple support vector machines for land cover change detection: An application for mapping urban extensions. {ISPRS} *Journal of Photogrammetry and Remote Sensing*, 61(2):125 – 133, 2006.

[224] Allan Aasbjerg Nielsen. Kernel maximum autocorrelation factor and minimum noise fraction transformations. *IEEE Transactions on Image Processing*, 20(3):612–624, 2011.

[225] Sebastian Nowozin, Christoph H Lampert, et al. Structured learning and prediction in computer vision. *Foundations and Trends® in Computer Graphics and Vision*, 6(3–4):185–365, 2011.

[226] Leif G. Olmanson, Patrick L. Brezonik, and Marvin E. Bauer. Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The mississippi river and its tributaries in minnesota. *Remote Sensing of Environment*, 130:254 – 265, 2013.

[227] Thomas H. Painter, Jeff Dozier, Dar A. Roberts, Robert E. Davis, and Robert O. Green. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, 85(1):64 – 77, 2003.

[228] M. Pal and G.M. Foody. Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2297–2307, May 2010.

[229] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[230] P. Pant, V. Heikkinen, I. Korpela, M. Hauta-Kasari, and T. Tokola. Logistic regression-based spectral band selection for tree species classification: Effects of spatial scale and balance in training samples. *IEEE Geoscience and Remote Sensing Letters*, 11(9):1604–1608, Sept 2014.

[231] E. Pasolli, F. Melgani, N. Alajlan, and Y. Bazi. Active learning methods for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10):4071–4084, Oct 2012.

[232] K.Y. Peerbhay, O. Mutanga, and R. Ismail. Random forests unsupervised classification: The detection and mapping of Solanum mauritianum infestations in plantation forestry using hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):3107–3122, June 2015.

[233] C. Persello. Interactive domain adaptation for the classification of remote sensing images using active learning. *IEEE Geoscience and Remote Sensing Letters*, 10(4):736–740, July 2013.

[234] C. Persello and L. Bruzzone. Active learning for domain adaptation in the supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4468–4483, Nov 2012.

[235] C. Pilorget and J.-P. Bibring. Automated algorithms to identify and locate grains of specific composition for {NIR} hyperspectral microscopes: Application to the micromega instrument onboard exomars. *Planetary and Space Science*, 99:7 – 18, 2014.

[236] Javier Plaza, Antonio J Plaza, and Cristina Barra. Multi-channel morphological profiles for classification of hyperspectral images using support vector machines. *Sensors*, 9(1):196–218, 2009.

[237] Ruiliang Pu, Peng Gong, Yong Tian, Xin Miao, Raymond I Carruthers, and Gerald L Anderson. Using classification and ndvi differencing methods for monitoring sparse vegetation coverage: a case study of saltcedar in nevada, usa. *International Journal of Remote Sensing*, 29(14):3987–4011, 2008.

[238] Ruiliang Pu, Peng Gong, Yong Tian, Xin Miao, RaymondI. Carruthers, and GeraldL. Anderson. Invasive species change detection using artificial neural networks and casi hyperspectral imagery. *Environmental Monitoring and Assessment*, 140(1-3):15–32, 2008.

[239] Y. Qian, M. Ye, and J. Zhou. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):2276–2291, April 2013.

[240] S. Rajan, J. Ghosh, and M.M. Crawford. Exploiting class hierarchies for knowledge transfer in hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(11):3408–3417, Nov 2006.

[241] P. Ramzi, F. Samadzadegan, and P. Reinartz. Classification of hyperspectral data using an adaboostsvm technique applied on band clusters. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2066–2079, June 2014.

[242] D Riao, E Chuvieco, S Ustin, R Zomer, P Dennison, D Roberts, and J Salas. Assessment of vegetation regeneration after fire through multitemporal analysis of {AVIRIS} images in the santa monica mountains. *Remote Sensing of Environment*, 79(1):60 – 71, 2002.

[243] Peter R. Robichaud, Sarah A. Lewis, Denise Y.M. Laes, Andrew T. Hudak, Raymond F. Kokaly, and Joseph A. Zamudio. Postfire soil burn severity mapping with hyperspectral image unmixing. *Remote Sensing of Environment*, 108(4):467 – 480, 2007.

[244] Juan José Rodriguez, Ludmila I Kuncheva, and Carlos J Alonso. Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1619–1630, 2006.

[245] Derek Rogge, Benoit Rivard, Karl Segl, Brian Grant, and Jilu Feng. Mapping of nicupge ore hosting ultramafic rocks using airborne and simulated enmap hyperspectral imagery, nunavik, canada. *Remote Sensing of Environment*, 152:302 – 317, 2014.

[246] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.

[247] A. Romero, C. Gatta, and G. Camps-Valls. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–14, 2015.

[248] Adriana Romero, Petia Radeva, and Carlo Gatta. Meta-parameter free unsupervised sparse feature learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1716–1722, 2015.

[249] W. Sakla, A. Chan, J. Ji, and A. Sakla. An SVDD-based algorithm for target detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(2):384–388, March 2011.

[250] A. Samat, Peijun Du, Sicong Liu, Jun Li, and Liang Cheng. E2LMs : Ensemble extreme learning machines for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1060–1069, April 2014.

[251] Q. Sami ul Haq, Linmi Tao, Fuchun Sun, and Shiqiang Yang. A fast and robust sparse approach for hyperspectral data classification using a few labeled samples. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6):2287–2302, June 2012.

[252] A. Santara, K. Mani, P. Hatwar, A. Singh, A. Garg, K. Padia, and P. Mitra. Bass net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9):5293–5301, Sept 2017.

[253] T. Schmid, M. Koch, J. Gumuzzio, and P. M. Mather. A spectral library for a semi-arid wetland and its application to studies of wetland degradation using hyperspectral and multispectral data. *International Journal of Remote Sensing*, 25(13):2485–2496, 2004.

[254] F. Schmidt, A. Schmidt, E. Treguier, M. Guiheneuf, S. Moussaoui, and N. Dobigeon. Implementation strategies for hyperspectral unmixing using Bayesian source separation. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4003–4013, Nov 2010.

[255] Sven Schneider, Arman Melkumyan, Richard J Murphy, and Eric Nettleton. Gaussian processes with oad covariance function for hyperspectral data classification. In *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, volume 1, pages 393–400. IEEE, 2010.

[256] Sven Schneider, Richard J. Murphy, and Arman Melkumyan. Evaluating the performance of a new classifier  the GP-OAD: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery. {*ISPRS*} *Journal of Photogrammetry and Remote Sensing*, 98:145 – 156, 2014.

[257] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[258] Burr Settles. Active learning literature survey. *Computer Sciences Technical Report 1648, University of Wisconsin, Madison*, 52(55-66):11, 2010.

[259] CA Shah, PK Varshney, and MK Arora. ICA mixture model algorithm for unsupervised classification of remote sensing imagery. *International Journal of Remote Sensing*, 28(8):1711–1731, 2007.

[260] Gary Shaw and Dimitris Manolakis. Signal processing for hyperspectral image exploitation. *IEEE Signal processing magazine*, 19(1):12–16, 2002.

[261] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, April 2017.

[262] Chen Shi and Le Wang. Incorporating spatial information in spectral unmixing: A review. *Remote Sensing of Environment*, 149:70 – 87, 2014.

[263] Hongjun Su, He Yang, Qian Du, and Yehua Sheng. Semisupervised band clustering for dimensionality reduction of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(6):1135–1139, 2011.

[264] Shujin Sun, Ping Zhong, Huaitie Xiao, and Runsheng Wang. Active learning with Gaussian process classifier for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1746–1760, April 2015.

[265] Shujin Sun, Ping Zhong, Huaitie Xiao, and Runsheng Wang. An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1074–1088, Sept 2015.

[266] Zhuo Sun, Cheng Wang, Hanyun Wang, and J. Li. Learn multiple-kernel SVMs for domain adaptation in hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 10(5):1224–1228, Sept 2013.

[267] Charles Sutton and Andrew McCallum. Foundations and trends® in machine learning. *Foundations and Trends in Machine Learning*, 4(4):267–373, 2011.

[268] L. Surez, P.J. Zarco-Tejada, G. Sepulcre-Cant, O. Prez-Priego, J.R. Miller, J.C. Jimnez-Muoz, and J. Sobrino. Assessing canopy {PRI} for water stress detection with diurnal airborne imagery. *Remote Sensing of Environment*, 112(2):560 – 575, 2008. Soil Moisture Experiments 2004 (SMEX04) Special Issue.

[269] C. Tao, H. Pan, Y. Li, and Z. Zou. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2438–2442, Dec 2015.

[270] Y. Tarabalka, M. Fauvel, J. Chanussot, and J.A. Benediktsson. SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 7(4):736–740, Oct 2010.

[271] Yuliya Tarabalka, Jón Atli Benediktsson, and Jocelyn Chanussot. Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2973–2987, 2009.

[272] Yuliya Tarabalka, Trym Vegard Haavardsholm, Ingebjørg Kåsen, and Torbjørn Skauli. Real-time anomaly detection in hyperspectral images using multivariate normal mixture models and GPU processing. *Journal of Real-Time Image Processing*, 4(3):287–300, 2009.

[273] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.

[274] Yee Whye Teh. Dirichlet process. In *Encyclopedia of machine learning*, pages 280–287. Springer, 2011.

[275] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas. A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing. *IEEE Transactions on Signal Processing*, 60(2):585–599, Feb 2012.

[276] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *The journal of machine learning research*, 1:211–244, 2001.

[277] Yung Liang Tong. *The multivariate normal distribution*. Springer Science & Business Media, 2012.

[278] Paul M Treitz and Philip J Howarth. Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems. *Progress in Physical Geography*, 23(3):359–390, 1999.

[279] E. Truslow, D. Manolakis, M. Pieper, T. Cooley, and M. Brueggeman. Performance prediction of matched filter and adaptive cosine estimator hyperspectral target detectors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2337–2350, June 2014.

[280] D. Tuia, E. Merenyi, X. Jia, and M. Grana-Romay. Foreword to the special issue on machine learning for remote sensing data processing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1007–1011, April 2014.

[281] D. Tuia, G. Moser, and B. Le Saux. 2016 ieee grss data fusion contest: Very high temporal resolution from space technical committees. *IEEE Geoscience and Remote Sensing Magazine*, 4(1):46–48, March 2016.

[282] D. Tuia, J. Verrelst, L. Alonso, F. Perez-Cruz, and G. Camps-Valls. Multioutput support vector regression for remote sensing biophysical parameter estimation. *IEEE Geoscience and Remote Sensing Letters*, 8(4):804–808, July 2011.

[283] Devis Tuia, Gustavo Camps-Valls, Giona Matasci, and Mikhail Kanevski. Learning relevant image features with multiple-kernel classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3780–3791, 2010.

[284] Freek D Van der Meer, Harald MA Van der Werff, Frank JA van Ruitenbeek, Chris A Hecker, Wim H Bakker, Marleen F Noomen, Mark van der Meijde, E John M Carranza, J Boudewijn de Smeth, and Tsehaie Woldai. Multi-and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):112–128, 2012.

[285] J. Verrelst, L. Alonso, J.P.R. Caicedo, J. Moreno, and G. Camps-Valls. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):867–874, April 2013.

[286] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5):1832–1843, May 2012.

[287] A. Villa, J. Chanussot, J.A. Benediktsson, and C. Jutten. Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):521–533, June 2011.

[288] A. Villa, J. Chanussot, J.A. Benediktsson, C. Jutten, and R. Dambreville. Unsupervised methods for the classification of hyperspectral images with low spatial resolution. *Pattern Recognition*, 46(6):1556 – 1568, 2013.

[289] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research*, 11:3371–3408, 2010.

[290] Vincenzo Esposito Vinzi, Wynne W Chin, Jörg Henseler, and Huiwen Wang. *Handbook of partial least squares: Concepts, methods and applications*. Springer Science & Business Media, 2010.

[291] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[292] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305, January 2008.

[293] Fu-Min Wang, Jing-Feng Huang, and Xiu-Zhen Wang. Identification of optimal hyperspectral bands for estimation of rice biophysical parameters. *Journal of integrative plant biology*, 50(3):291–299, 2008.

[294] Jing Wang and Chein-I Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote sensing*, 44(6):1586–1600, 2006.

[295] Kai Wang, Steven E Franklin, Xulin Guo, and Marc Cattet. Remote sensing of ecology, biodiversity and conservation: A review from the perspective of remote sensing specialists. *Sensors*, 10(11):9647–9667, 2010.

[296] Liguo Wang and Xiuping Jia. Integration of soft and hard classifications using extended support vector machines. *IEEE Geoscience and Remote Sensing Letters*, 6(3):543–547, July 2009.

[297] Liguo Wang and Qunming Wang. Subpixel mapping using Markov random field with multiple spectral constraints from subpixel shifted remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 10(3):598–602, May 2013.

[298] Qingwang Wang, Yanfeng Gu, and Devis Tuia. Discriminative multiple kernel learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):3912–3927, 2016.

[299] B. Waske, S. van der Linden, J.A. Benediktsson, A. Rabe, and P. Hostert. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(7):2880–2889, July 2010.

[300] Elizabeth A Wentz, Sharolyn Anderson, Michail Fragkias, Maik Netzband, Victor Mesev, Soe W Myint, Dale Quattrochi, Atiqur Rahman, and Karen C Seto. Supporting global environmental change research: A review of trends and knowledge gaps in urban remote sensing. *Remote Sensing*, 6(5):3879–3905, 2014.

[301] Christopher KI Williams and Carl Edward Rasmussen. Gaussian processes for machine learning. *the MIT Press*, 2(3):4, 2006.

[302] Hao Wu and Saurabh Prasad. Convolutional recurrent neural networks forhyperspectral data classification. *Remote Sensing*, 9(3), 2017.

[303] Zebin Wu, Qicong Wang, A. Plaza, Jun Li, Le Sun, and Zhihui Wei. Real-time implementation of the sparse multinomial logistic regression for hyperspectral image classification on GPUs. *IEEE Geoscience and Remote Sensing Letters*, 12(7):1456–1460, July 2015.

[304] Junshi Xia, J. Chanussot, Peijun Du, and Xiyan He. Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2532–2546, May 2015.

[305] Junshi Xia, M. Dalla Mura, J. Chanussot, Peijun Du, and Xiyan He. Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(9):4768–4786, Sept 2015.

[306] Junshi Xia, Peijun Du, Xiyan He, and J. Chanussot. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1):239–243, Jan 2014.

[307] Yang Xu, Zebin Wu, and Zhihui Wei. Spectral-spatial classification of hyperspectral image based on low-rank decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2370–2380, June 2015.

[308] Chen Yang, Sicong Liu, Lorenzo Bruzzone, Renchu Guan, and Peijun Du. A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 10(5):1152–1156, 2013.

[309] He Yang, Qian Du, and Ben Ma. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 7(4):875–879, 2010.

[310] J. Yang, Y. Q. Zhao, and J. C. W. Chan. Learning and transferring deep joint spectral-spatial features for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4729–4742, Aug 2017.

[311] Jun Yue, Wenzhi Zhao, Shanjun Mao, and Hui Liu. Spectral–spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, 6(6):468–477, 2015.

[312] P WT Yuen and M Richardson. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *The Imaging Science Journal*, 58(5):241–253, 2010.

[313] A. Zare and P. Gader. PCE: Piecewise convex endmember detection. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6):2620–2632, June 2010.

[314] A. Zare, P. Gader, and G. Casella. Sampling piecewise convex unmixing and endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3):1655–1665, March 2013.

[315] Guangyun Zhang and Xiuping Jia. Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters*, 9(5):856–860, 2012.

[316] Haokui Zhang, Ying Li, Yuzhu Zhang, and Qiang Shen. Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sensing Letters*, 8(5):438–447, 2017.

[317] Hongyan Zhang, Jiayi Li, Yuancheng Huang, and Liangpei Zhang. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2056–2065, June 2014.

[318] L. Zhang, L. Zhang, and B. Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, June 2016.

[319] Ji Zhao, Yanfei Zhong, Yunyun Wu, Liangpei Zhang, and Hong Shu. Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1049–1060, Sept 2015.

[320] Ji Zhao, Yanfei Zhong, and Liangpei Zhang. Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery. *IEEE transactions on Geoscience and Remote Sensing*, 53(5):2440–2452, 2015.

[321] Kaiguang Zhao, Denis Valle, Sorin Popescu, Xuesong Zhang, and Bani Mallick. Hyperspectral remote sensing of plant biochemistry using bayesian model averaging with variable and band selection. *Remote Sensing of Environment*, 132:102 – 119, 2013.

[322] Wenzhi Zhao and Shihong Du. Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4544–4554, 2016.

[323] Wenzhi Zhao, Zhou Guo, Jun Yue, Xiuyuan Zhang, and Liqun Luo. On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing*, 36(13):3368–3379, 2015.

[324] Ping Zhong and Runsheng Wang. Learning sparse CRFs for feature selection and classification of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(12):4186–4197, Dec 2008.

[325] Ping Zhong and Runsheng Wang. Learning conditional random fields for classification of hyperspectral images. *IEEE Transactions on Image Processing*, 19(7):1890–1907, July 2010.

[326] Ping Zhong and Runsheng Wang. Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):688–705, Feb 2011.

[327] Ping Zhong, Peng Zhang, and Runsheng Wang. Dynamic learning of SMLR for feature selection and classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Letters*, 5(2):280–284, April 2008.

[328] Yanfei Zhong, Xuemei Lin, and Liangpei Zhang. A support vector conditional random fields classifier with a mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery. *IEEE journal of selected topics in applied earth observations and remote sensing*, 7(4):1314–1330, 2014.

[329] Z. Zhong, J. Li, Z. Luo, and M. Chapman. Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–12, 2017.

[330] Xiaojin Zhu and Andrew B Goldberg. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130, 2009.

[331] Maciel Zortea and Antonio Plaza. A quantitative and comparative analysis of different implementations of n-findr: A fast endmember extraction algorithm. *IEEE Geoscience and Remote Sensing Letters*, 6(4):787–791, 2009.

[332] Zhen Zuo, Bing Shuai, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, and Yushi Chen. Convolutional recurrent neural networks: Learning spatial dependencies for image representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 18–26, 2015.