



## Review

# Computer Vision and Pattern Recognition for the Analysis of 2D/3D Remote Sensing Data in Geoscience: A Survey

Michalis A. Savelonas <sup>1,\*</sup>, Christos N. Veinidis <sup>2</sup> and Theodoros K. Bartsokas <sup>2</sup>

<sup>1</sup> Department of Computer Science and Biomedical Informatics, University of Thessaly, 35131 Lamia, Greece

<sup>2</sup> Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Athens, Greece

\* Correspondence: msavelonas@uth.gr

**Abstract:** Historically, geoscience has been a prominent domain for applications of computer vision and pattern recognition. The numerous challenges associated with geoscience-related imaging data, which include poor imaging quality, noise, missing values, lack of precise boundaries defining various geoscience objects and processes, as well as non-stationarity in space and/or time, provide an ideal test bed for advanced computer vision techniques. On the other hand, the developments in pattern recognition, especially with the rapid evolution of powerful graphical processing units (GPUs) and the subsequent deep learning breakthrough, enable valuable computational tools, which can aid geoscientists in important problems, such as land cover mapping, target detection, pattern mining in imaging data, boundary extraction and change detection. In this landscape, classical computer vision approaches, such as active contours, superpixels, or descriptor-guided classification, provide alternatives that remain relevant when domain expert labelling of large sample collections is often not feasible. This issue persists, despite efforts for the standardization of geoscience datasets, such as Microsoft's effort for AI on Earth, or Google Earth. This work covers developments in applications of computer vision and pattern recognition on geoscience-related imaging data, following both pre-deep learning and post-deep learning paradigms. Various imaging modalities are addressed, including: multispectral images, hyperspectral images (HSIs), synthetic aperture radar (SAR) images, point clouds obtained from light detection and ranging (LiDAR) sensors or digital elevation models (DEMs).

**Keywords:** geoscience; computer vision; pattern recognition; deep learning; LiDAR; multispectral imaging; hyperspectral imaging; SAR imaging; land cover mapping; target detection; change detection



**Citation:** Savelonas, M.A.; Veinidis, C.N.; Bartsokas, T.K. Computer Vision and Pattern Recognition for the Analysis of 2D/3D Remote Sensing Data in Geoscience: A Survey. *Remote Sens.* **2022**, *14*, 6017. <https://doi.org/10.3390/rs14236017>

Academic Editors: Anup Basu, Chengcai Leng and Hemanth Venkateswara

Received: 28 September 2022

Accepted: 24 November 2022

Published: 27 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The advent of various 2D/3D imaging technologies in the area of RS (remote sensing) raises multiple challenges for computational tools capable of assisting domain experts, such as earth scientists, on conducting land cover surveys, as well as in the study of a diverse range of phenomena. The growing availability of large amounts of imaging data in the form of 2D images (aerial, satellite, etc.), point clouds, hyperspectral images (HSIs), as well as in the form of image time-series, sets the stage for the design of various deep learning (DL)-based strategies, as well as of standard descriptor-based computer vision approaches.

Geoscience-related imaging data usually appear in the form of: (1) RS images acquired by satellites, (2) images derived by measurements from in situ sensors in the sea, land, or air, and (3) simulated images generated using physics-based or machine-learning models. Space research organizations, including the national aeronautics and space administration (NASA) and European space agency (ESA), provide a large, ever increasing, amount of RS data in the context of geoscience [1]. Earth-observing satellites monitor various geoscience-related parameters, such as surface temperature, humidity, optical reflectance and the chemical content of the atmosphere. These parameters are recorded at fine spatial scales

and regular time intervals for long periods, as is the case with the Landsat archives [2]. RS data can also be collected using imaging sensors on flying devices, such as unmanned aerial vehicles (UAVs) or aircrafts [3]. Such imaging sensors are RGB cameras, multispectral and hyperspectral sensors, thermal cameras, light detection and ranging (LiDAR) sensors and radars [4]. These data commonly comprise samples of geo-registered images over single time points or time-series of images associated with single spatial locations [1].

There are projects dedicated to the standardization of geoscience datasets [5], as is the case with Microsoft's AI for Earth. [6]. Google Earth engine [7] is a widely utilized cloud-based platform for planetary-scale geospatial analysis using Google's infrastructure. Google Earth features several modes of interaction: the Code Editor (Figure 1) is a web-based integrated development environment (IDE) for writing and running scripts, the Explorer is a lightweight web app for exploring Google's data catalog and running simple analyses, whereas the client libraries provide Python and JavaScript wrappers around a web application programming interface (API). Recently, Google introduced Dynamic World [8], which leverages advances in large-scale cloud computing, DL, high-performance open-source software frameworks, such as TensorFlow, as well as increased access to satellite image collections through platforms, including Google Earth engine, in order to create global land cover maps at a spatial resolution of less than 10m and near real-time. Figure 2 presents an example land cover map of a part of Europe obtained by Google's Dynamic World.

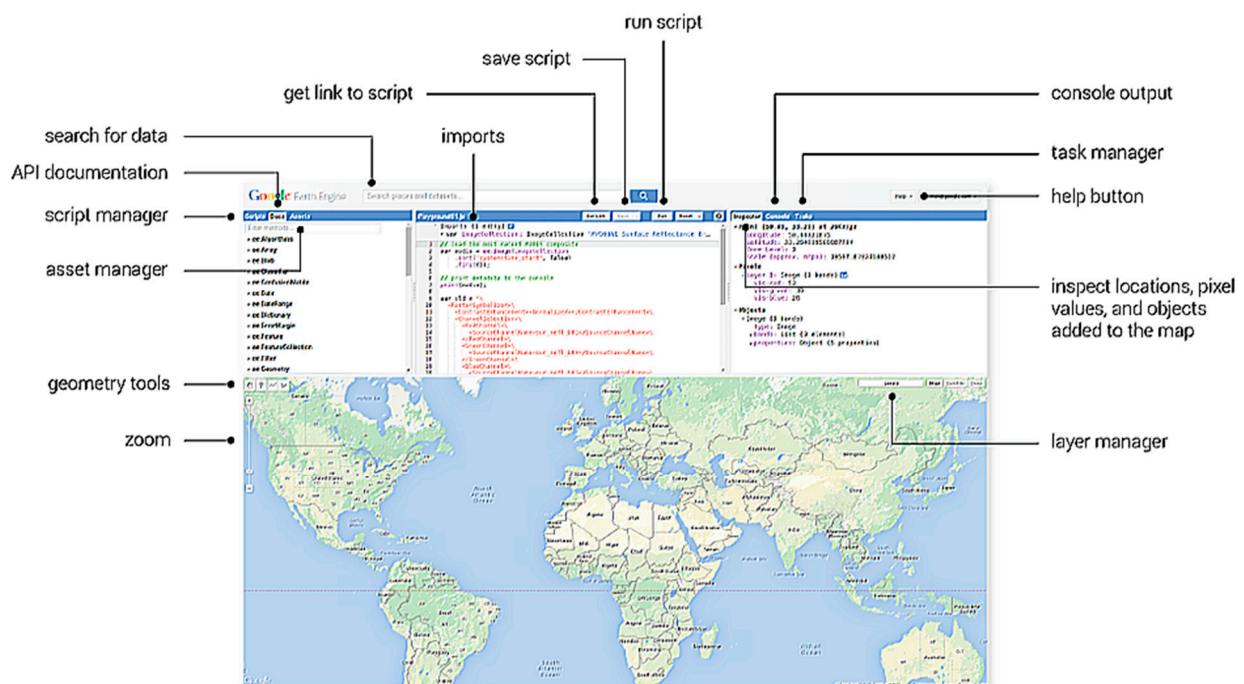


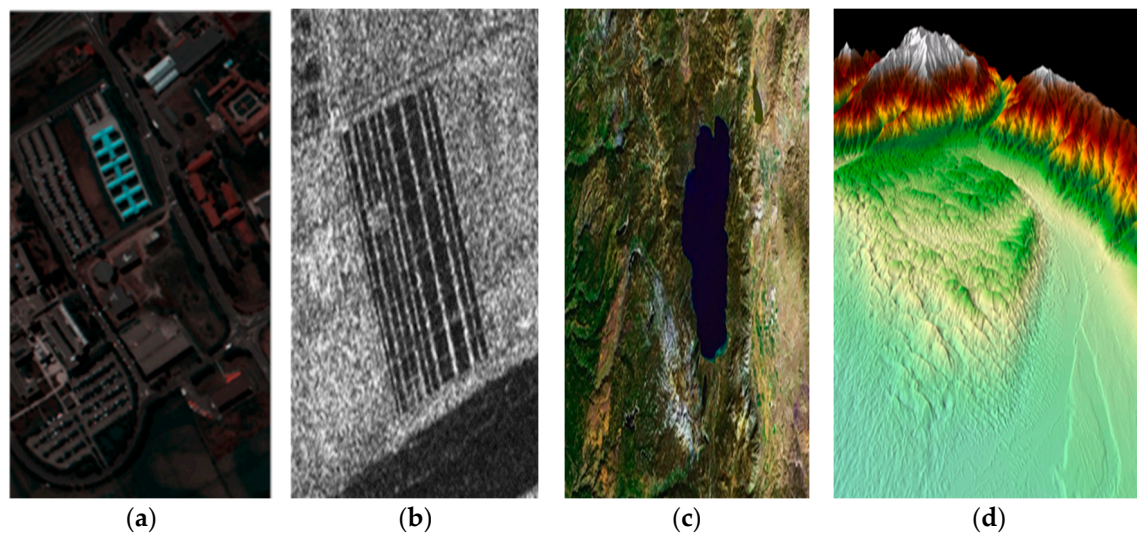
Figure 1. The elements of Google Earth Code Editor.



**Figure 2.** An example land cover map of a part of Europe received by the Google’s Dynamic World. The utilized classes are nine: water, trees, grass, flooded vegetation, crops, shrub and scrub, built, bare, snow and ice (<https://dynamicworld.app/explore/> (accessed on 26 November 2022)).

In terms of modalities, imaging data are acquired in the form of 2D images, multidimensional images or point-clouds. Consumer-level cameras have been adopted since the late 1990s for applications such as digital elevation models (DEMs) and RS of vegetation. Many of these aerial images are being used to produce topographic data by means of modern structure-from-motion (SfM) algorithms [9]. Satellite missions such as Sentinel, Venus, or Landsat data continuity mission (LDCM) provide time-series of high-resolution optical data. Hyperspectral imaging is a powerful RS technology, enabling the study of chemical and physical properties of scene materials. It provides multi-dimensional images capturing the reflected or emitted electromagnetic energy from a scene, over hundreds of narrow, contiguous spectral bands, from visible to infrared wavelengths [10,11]. These spectral bands represent different materials and form mixed pixels, which are the basic element of HSIs. The use of point clouds has promising perspectives in different fields of geosciences, for supporting high-resolution geological or geomorphological mapping, for studying the evolution of active processes, as well as for monitoring various kinds of natural hazards. Airborne LiDAR is a laser profiling and scanning technique for bathymetric and topographic applications, which emerged in mid-1990s. With the aid of direct geo-referencing, the laser scanning equipment installed in a UAV or aircraft collects a cloud of laser range measurements of the 3D coordinates of the area under inspection. In contrast to 2D planimetric data, the explicit LiDAR data point cloud describes the 3D topographic profile of a surface [12]. Figure 3 presents example images of various modalities, widely employed in geoscience-related applications.





**Figure 3.** Example data widely-used in geoscience-related applications: (a) the composite HSI image of Pavia University, northern Italy, acquired by the ROSIS sensor, (b) SAR image from the Yellow River dataset, acquired in June 2008, (c) image of Lake Tahoe acquired by Landsat 8 operational land imager (OLI), (d) National Elevation Data map of Australia DEM.

Historically, the progress of RS imaging has been closely associated with computer vision and pattern recognition. Several geoscience-related problems, including land cover mapping, target detection and boundary extraction, pose challenges for traditional image analysis tasks, such as image clustering, classification and segmentation. Descriptors (i.e., handcrafted feature vectors) such as local binary patterns (LBP) [13], wavelet vectors [14] and morphological profiles [15] have been applied for the classification of optical images and HSIs [16]. Spin images [17], point feature histograms (PFH) [18] and signatures of histogram of orientation (SHOT) [19] are descriptors that have been employed for point clouds. The standard approach for image clustering or classification is to provide 2D, 3D or multidimensional descriptors as inputs to classical clustering algorithms, such as *k*-means and fuzzy-*c*-means, or to classifiers such as decision trees (DTs), random forests (RFs), support vector machines (SVMs) and multi linear perceptrons (MLPs) (see Section 2). A problem with this approach is that descriptor parameterization depends on expert knowledge, limiting their applicability in difficult scenarios, featuring subtle inter-class or large intra-class variations [20]. Later, visual dictionaries, the so-called bag of visual words (BoVW) [21,22], incorporated the statistics associated with each problem at hand [23]. In the context of semantic segmentation, numerous approaches, including active contours [24], Markov random fields (MRFs) [25] and superpixels [26], have been combined with descriptors or BoVW and to be applied on 2D RS images or HSIs.

Despite some successful results of descriptor-based approaches, the geoscience community has been slow in adopting computer vision tools, partially due to limitations in the actual representation capability of descriptors or BoVW. The rapid evolution of powerful GPUs and the availability of large datasets aided extraordinary advances in DL-based computer vision, starting from the performance breakthrough of AlexNet in ILSVRC 2012 [27]. This new paradigm reinvigorated the interest of geoscientists. Numerous works are regularly published for the analysis of RS images with deep neural networks. Convolutional neural networks (CNNs) [28] and derivative architectures, such as VGG [29], ResNet [30] and Inception [31], play a prominent role in this direction. Another DL branch in RS consists of applications of recurrent neural networks (RNNs) [32], including long short-term memory (LSTM) networks [33], on time-series of images. Regularly, the training of these methods requires large amounts of data, preferably labelled. Labelling is costly in terms of time and domain expert resources. A remedy for the lack of a sufficient amount of labelled data, which is often adopted in machine learning, is provided by data augmentation: real

samples are transformed (e.g., by translation or rotation) to generate additional synthetic samples. Another related approach is to generate synthetic samples following a learning process on a labelled dataset. In this case, variational autoencoders (VAEs) [34] and generative adversarial networks (GANs) [35] are the most widely adopted methods.

A number of survey articles covering computational approaches in RS has appeared in the literature in the last couple of years. A part of these surveys [11,20] focus on machine learning applications on a single imaging modality, such as HSIs. Other surveys cover imaging and non-imaging data [1,36,37]. Overall, most recent surveys [20,38–40] focus on DL and miss a large amount of related work which is not based on neural networks. This survey maintains a focus on imaging modalities and geoscience-related applications, whereas it provides an overview that spans both non-DL-based and DL-based methods. Accordingly, it fills a gap in recent literature, since no other survey: (1) spans multiple imaging modalities in geoscience-related applications, (2) bridges non-DL and DL-based computer vision methods, keeping track of research activity in the former paradigm. This is particularly important considering that in several cases there is a lack of sufficient data for training deep neural networks, and classical approaches, such as superpixels or decision trees (DTs), provide a well-tested alternative. A remark should be added with respect to earthquakes. Due to the lack of distinctive visual information, this geologic event has received limited attention from the computer vision community [41]. Some works start from acoustic signals to derive 2D images to be analyzed. We consider such signal processing applications beyond the scope of this work.

The rest of this article is organized as follows: in Section 2, we provide an overview of the core computer vision and pattern recognition approaches, which have been prevalent in geoscience-related applications in the last three decades. In Section 3, we present applications of these approaches for land cover mapping, target detection, mining associations in geoscience imaging data, boundary extraction, change detection and RS image pre-processing. In Section 4, we discuss overall trends and aspects of the applications presented, as well as the main challenges raised, whereas in Section 5 we present the main conclusions of this work.

## 2. Computer Vision and Pattern Recognition Approaches

This Section presents prominent computer vision and pattern recognition approaches, which have been extensively applied in geoscience-related problems. Acknowledging the fact that DL induced a new era in computer vision, we divide the presentation into pre-DL and DL-based approaches. As is the case with computer vision in general, DL is prevalent in the state-of-the-art in most geoscience-related applications. However, for various reasons we maintain that pre-DL approaches are still relevant. These reasons include the difficulty in obtaining sufficiently large labelled datasets to train deep neural networks, the computational efficiency and intuitiveness of some pre-DL approaches, as well as the appearance of hybrid methods that combine pre-DL approaches with deep neural networks. Certain approaches, such as SVMs or CNNs, are fundamental and widely adopted, dictating a more detailed presentation in the text to follow.

### 2.1. Computer Vision and Pattern Recognition prior Deep Learning

In this Subsection, we present pre-DL-based approaches that are still relevant and continue to appear frequently in the literature.

#### 2.1.1. Descriptors

The standard approach, which was prevalent in computer vision before the DL breakthrough, was to employ descriptors: handcrafted, low-level feature vectors, representing color or textural information. Widely employed descriptors include gray-level co-occurrence matrices (GLCMs), introduced in the seminal work of Haralick et al. [42], local binary patterns (LBPs) [43], Gabor filtering [44], wavelet filters [14], morphological profiles [15] and scale-invariant feature transform (SIFT) [45]. Typically, descriptors guide

segmentation methods such as active contours (see Section 2.1.2) and Markov random fields (MRFs, see Section 2.1.3) or classifiers such as random forests (RFs, see Section 2.1.6) and support vector machines (SVMs, see Section 2.1.7). 3D vision employs shape descriptors, such as spin images [17], 3D shape context [46], point feature histograms (PFHs) [18], signature of histogram of orientations (SHOT) [19] and view-point feature histogram (VFH) [47]. Beyond the simple approach of directly employing such low-level features, global feature statistics can be encoded by means of mid-level features: bag of visual words (BoVW) [48] represent a sample with the frequencies of local visual words. Vectors of locally aggregated descriptors (VLAD) [49] concatenate the differences of local feature vectors and cluster center vectors, whereas Fisher vectors (FV) [50] use gaussian mixture models (GMMs) to encode the statistics of descriptors. Such features have been successfully used in various RS and geoscience-related applications, such as classification of sea-ice, extraction of built-up areas and land cover mapping (see Section 3) [51].

### 2.1.2. Active Contours

Active contours (ACs) [24], are physics-inspired image segmentation models, which employ iteratively evolving contours guided by energy minimization. The energy functional encompasses terms associated with internal and external ‘forces’. The former terms depend on contour features (e.g., ‘elasticity’), whereas the latter depend on image-derived features, either edge-based, such as intensity gradients, or region-based, such as texture. ACs were initially formulated as parametric curves. Later, Osher and Sethian [52] introduced the level-set formulation, which employs an implicit contour representation and is capable of handling topological contour changes, such as merging or splitting. Several AC-based RS applications (see Section 3) have appeared in the literature, especially in the pre-DL era, motivated by the attractive attributes of some AC variants, including noise robustness, insensitivity to initialization and intuitiveness [53].

### 2.1.3. Markov Random Fields

Markov random fields (MRFs) are statistical modeling tools, which are capable of integrating spatial context into image segmentation and classification algorithms [54,55]. MRFs formulate the maximum a posteriori decision rule as an energy minimization problem [56], which is graph representable [57] and can be efficiently solved by graph-cuts [58,59]. This efficiency as well as the accuracy in the results obtained are the reasons that MRFs are regarded as an effective tool for incorporating spatial information in image segmentation and classification. In particular, this has been proved in the case of HSIs [59,60]. Another advantage of MRFs in the context of HSI classification, is that MRFs can be combined with classifiers, such as support vector machines (SMVs, see Section 2.1.7) under Bayesian frameworks, and cope with the ‘curse of dimensionality’ (or Hughes phenomenon), which is attributed to this type of images, due to the large number of correlated spectral bands.

### 2.1.4. Superpixels

Superpixels are perceptually meaningful groups of pixels, formed on the basis of similarity with respect to some low-level feature, such as colour or texture [61]. The image oversegmentation resulting from superpixels have often been used in the context of image segmentation algorithms as a form of initialization. Unlike pixels, superpixels follow natural image boundaries, whereas they are robust against noise and artifacts and induce low computational cost [62]. In addition, superpixels provide an intuitive mechanism to bypass the need for seed-based initialization, which arises in several image segmentation pipelines [37]. Superpixels are often based into graphs or gradient ascent [63]. Several methods have been proposed to generate superpixels [37], such as superpixels lattice [64], turbopixels [65], quick shift [66] and simple linear iterative clustering (SLIC) [63]. The advantages of these methods enabled numerous geoscience-related superpixel applications, mostly for land cover mapping, where superpixels have been combined with SVMs, RFs,

NNs (including CNNs), in order to provide an initial oversegmentation, as described above (see Section 3).

#### 2.1.5. Clustering

Clustering is the grouping of a set of data samples in such a way that samples of the same group are close or similar, in some sense, to one another, whereas they are dissimilar from samples of other groups [67]. It is primarily a form of unsupervised learning, although semi-supervised or supervised variants have emerged. The concept is quite general and has been applied in various domains, including medical sciences, engineering and earth sciences.

The process of clustering can be divided into four stages [68]: (1) feature selection and/or extraction: feature selection aims to select a subset of the input feature set, whereas feature extraction refers to algorithms that process the original feature set and create new features [69], (2) clustering algorithm design or selection: this stage is usually combined with the adoption of a suitable proximity measure, which quantifies the similarity of the feature vectors, (3) cluster validation: the validity of the clustering results is tested, usually employing evaluation standards and criteria, which provide a degree of confidence, (4) results interpretation: a domain expert interprets the clustering results, aiming to discover meaningful insights.

A number of taxonomies has been proposed for clustering [70–72]. Hierarchical clustering [73,74] includes two types of clustering, namely agglomerative and divisive, which operate inversely and iteratively produce clustering results with decreasing or increasing number of clusters, respectively. Hard and fuzzy clustering algorithms [75], assign a feature vector to one or more than one clusters. Genetic clustering uses principles inspired by natural population genetics [76]. Possibilistic clustering [77] aims to measure the possibility of a feature vector to belong to a cluster. Density-based clustering [78] operates with the consideration that the valid clusters are regions with high sample density. Subspace clustering [79] is performed on subspaces of the original feature space. Some classical clustering algorithms are *k*-means [80], fuzzy *c*-means (FCM) [81], Gustafson-Kessel [82], density-based spatial clustering of applications with noise (DBSCAN) [83], mean-shift [84], ordering points to identify the clustering structure (OPTICS) [85], balanced iterative reducing and clustering using hierarchies (BIRCH) [86], and generalized principal component analysis (GPCA) [87].

Clustering has been employed in a wide range of geoscience applications. For land cover mapping, it has been employed for segmentation, which can be posed as a clustering problem at the pixel level. In addition, some clustering-based methods have been proposed for spectral band selection in HSIs, aiming to aid a subsequent HSI classification stage, as well as for target detection and change detection (see Section 3).

#### 2.1.6. Decision Trees and Random Forests

Decision trees (DTs) [88] and random forests (RFs) [89] are supervised learning approaches, suitable for classification and regression. DTs are aimed to investigate all the potential solutions of the problem at hand. Each DT is realized as a tree-structure consisting of two types of nodes: decision and leaf nodes. The former determines the path leading to a certain solution, whereas the latter are the outcomes. RFs consist of a number of DTs, which operate as an ensemble. Each of these DTs is trained by a different subset of the training set. In testing, the same input is provided to all trained DTs and each one independently results in an output. For classification, the RF classifies each input following the majority of DT outputs ('votes'). For regression, the RF output is the average of all DT outputs. Another supervised learning approach, which utilizes DT ensembles, is the gradient boosted decision tree (GBDT) [90]: each DT is trained sequentially, as opposed to RFs, where DTs are trained independently and only the DT outputs are combined to derive the RF output. The GBDT loss function is minimized via the sequential minimization of the training loss function of each subsequent DT.



When considering RS applications in geoscience, DTs and RFs have mostly been employed for land cover mapping, as well as for pattern mining in geoscience-related imaging data (see Section 3).

#### 2.1.7. Support Vector Machines

SVMs [91] belong to a type of classifiers which are trained by means of a predefined optimisation criterion. Supposing two, linearly separable classes, the training objective is to determine a hyperplane in the feature space, which has the maximum distance from the nearest samples of both classes. This hyperplane lies in the middle of a ‘stripe’ defined by these nearest samples, which accounts for no preference to either class. This approach aims to enhance the generalization capability of the resulting classifier.

The mathematical formulation of SVMs is based on the minimization of a cost function, under a set of constraints, named Karush–Kuhn–Tucker (KKT) conditions. A typical treatment for this problem is to minimize the corresponding Lagrange function, which is defined by embedding the KKT conditions in the main cost function. The solution of the optimisation problem dictates that the direction of the optimum hyperplane is determined as a linear combination of a subset of feature vectors of training samples, named support vectors. Support vectors are the nearest feature vectors to the hyperplane determined.

The generalization of this algorithm to the non-linearly separable case is based on the mapping of the input data space to a Hilbert space with higher dimensionality. The mathematical formulation in the new space is similar to the one in the linear case and is based on the inner products of the feature vectors. The inner products are realized as kernels and the matrix comprising the kernel values computed on each potential pair of feature vectors is called kernel matrix. Typical kernels utilized within the SVM context are the polynomials, the radial basis functions (RBFs) and the hyperbolic tangents. Unlike the linearly separable case, the optimisation result is a non-linear discriminating surface. Finally, generalizations of the multiclass problem include one-against-all, one-against-one and error correcting coding [92].

As evident in the text to follow, SVMs have been the pre-DL classifiers of choice for most types of RS applications: land cover mapping, target detection, pattern mining and change detection (see Section 3).

#### 2.1.8. Linear and Logistic Regression

Linear regression is a supervised approach to estimate the relationship between two variables by fitting a linear equation to the observed data. The training of a linear regression model is essentially parameter estimation of a linear system of equations. These parameters are estimated by adopting the least squares (LS) criterion, i.e., minimizing the sum of error squares, which are defined by the squared differences between the actual labels of the training data and their linear approximation. The solution is derived by the multiplication of the linear system’s pseudoinverse matrix with the vector containing the actual labels of the training data.

In the context of the Bayesian framework, the assignment of a feature vector to a class is based on the maximization of the posterior probabilities  $P(\omega_i|x)$ ,  $i = 1, 2, \dots, M$ , where  $\omega_i$  is the class  $i$ ,  $x$  is a feature vector and  $M$  is the total number of classes. The logarithm of the ratio of these probabilities is modeled as a linear function of  $x$ , by means of the logistic regression (LR) approach [93]. Model parameters are usually estimated by means of the maximum likelihood (ML), exploiting the feature vectors of the training set. The derived log-likelihood function is maximized using an optimisation algorithm, such as gradient descent or Newton’s algorithm [94]. In the case of  $M = 2$ , the LR model is called binary, whereas in the case of  $M > 2$ , it is called multinomial.

Beyond obvious applications for a classifier, such as land cover mapping, target detection and change detection, LR has often been employed for trend analysis, prediction, susceptibility mapping and pattern mining in geoscience-related imaging data (see Section 3).



### 2.1.9. Artificial Neural Networks

An artificial neural network (ANN) is an interconnected assembly of simple processing elements, called neurons or perceptrons [95], which have a functionality loosely based on the animal neuron [96]. Each neuron receives as input the values of the components of a feature vector, a set of weight values, called synaptic weights or synapses and a constant value named bias. The neuron alters these values by: (1) deriving the inner product of the feature vector and the weight vector. The former is augmented with the constant value 1, whereas the latter is augmented with the bias, (2) computing the value of a nonlinear function, called activation function, on the resulting inner product.

The basic model of a neuron is employed on several complicated ANN architectures. A basic taxonomy of these architectures [97] separates ANNs as: (1) feed-forward networks [98], including single-layer perceptrons, multilayer perceptrons (MLPs) and radial basis functions (RBFs), (2) RNNs [32] (see also Section 2.2.2), including competitive networks, self-organizing maps (SOMs) [99] and Hopfield networks [100]. Each of these architectures is defined by its neuron interconnections, the type of activation functions, the number of neuron layers, the existence or not of feedback loops between neuron of different layers, etc. Each ANN is trained to determine all synaptic weights and bias values that minimize a predefined cost function of network output.

ANNs have been employed as tools for various pattern recognition tasks, including clustering, classification, regression and prediction. The development of ANNs, spans decades of research, from the work of McCulloch and Pitts [101] to DL, presented in the following sections. In the context of RS, pre-DL NN-based applications were less than the SVM-based ones, which could be attributed to the comparative advantages of the latter, mostly related to generalization capability. Still, shallow MLPs can be found in the text to follow, mostly dated prior to 2012 (AlexNet breakthrough—see Section 3).

## 2.2. Deep Learning-Based Computer Vision and Pattern Recognition

In this Subsection, we present deep neural network architectures that have been successfully applied on geoscience-related problems.

### 2.2.1. Convolutional Neural Networks

CNNs [102] are used for image analysis in a wide range of applications and domains, such as bioinformatics, robotics and geosciences. One of the main reasons for their success is their effective image representations, which is attributed to their capability in preserving spatial information while reducing input dimensionality. This capability is based on two mathematical operations, convolution and pooling.

2D convolution operates as a feature extractor. A filter represented as a 2D array is sliding over potentially overlapping, equally sized, regions of the input image. In each sliding iteration, the dot product between the filter array and the image region array is registered as an element of the output array, which has smaller dimensions than the input array. Each element of the filter array corresponds to a different ‘neuron’, which is accompanied by an added bias term, followed by a nonlinear activation function, such as rectified linear unit (ReLU) and sigmoid. The resulting array is also called feature map. The elements of each filter constitute weights to be learned and different weight values result in different extracted features maps. Each such sequence of operations constitutes a convolutional layer.

Pooling is essentially an instrument for dimensionality reduction and, similarly to convolution, is performed on image regions. The two most common pooling operations are max pooling and average pooling. The former maintains the maximum value of each region, whereas the latter maintains the average value of each region. Apart from dimensionality reduction, pooling also enables translational invariance, which is a key attribute for image understanding. These operations constitute the pooling layer.

In CNNs, there are potentially multiple convolutional layers, followed by nonlinear activations and pooling layers, providing a sequence of image abstractions. Early layers

are associated with low-level image information (very often edges or ridges), whereas later layers tend to represent semantic information, associated with the task at hand. This first part of a CNN is devoted to feature learning. The second part of a CNN uses the extracted features to accomplish a specific task, such as classification and segmentation. For this, the last feature map is 'flattened', i.e., is converted to a 1D vector. Thus, spatial proximity is no longer maintained. This 1D vector is the input to a series of fully connected layers. In the case of image classification, the fully connected layers end with a 1D vector comprising the probabilities of each class.

Another CNN-based architecture is YOLO (you only look once) [103]. YOLO is a single-shot detector, which determines bounding boxes and class probabilities directly from full images. In Section 3.2, a number of YOLO-based target detection applications in the context of RS are reviewed.

### 2.2.2. Recurrent Neural Networks

RNNs [102] have been proposed much earlier than the DL era to handle time-series data by means of sequences of state and hidden state variables. Each state or hidden state variable depends on previous states and hidden states. Multiple layers of hidden states can be used in deeper architectures, aiming to capture underlying dynamics. Despite some successful applications on various domains, early variants of RNNs often fail to capture long-term dependencies due to the vanishing gradient effect [104]. A first solution to cope with this issue is provided by alternatives to stochastic gradient descent [104,105]. A second solution involves the design of a sophisticated activation function, which consists of an affine transformation and a simple element-wise nonlinearity obtained by gating units. The earliest such method employed a recurrent unit, called long short-term memory (LSTM) [33]. Later, the gated re-current unit (GRU) was proposed [104]. In the last decade, deep LSTMs or GRUs have been successfully applied in tasks that require capturing long-term dependencies, such as speech recognition [106], natural language processing [107] and driving behaviour classification [108]. In a similar fashion, RNNs have been applied to time-series of RS data (see Section 3). Starting from the observation that the temporal variability of a sequential signal is similar to the spectral variability of a hyperspectral pixel, the same idea has been applied to HSIs for image classification.

A DL-based alternative for processing sequential data is provided by Transformers. Transformers originate from natural language processing but have also attracted the interest of computer vision community [109]. Transformers are deep neural networks mainly based on the self-attention mechanism [110]. Unlike RNNs, Transformers process the entire input at once. Recently, Transformer-based architectures have been used in RS, mainly for target detection [111]. Related applications are reviewed in Section 3.2.

### 2.2.3. Deep Generative Models and GANs

Deep generative models (DGMs) aim to generate synthetic data by means of sampling from probability distribution functions (PDFs), which are learned from available datasets. A standard approach for DGM learning is to maximize the log-likelihood or a lower bound of the log-likelihood of the PDF [112]. Generative adversarial networks (GANs) [35] are prominent members of the DGM family and follow a different learning approach, based on a minmax game between two competing neural networks: the generator and the discriminator. The generator starts from a latent variable space, associated with a prior PDF. This PDF is sampled and projected to the data space. The discriminator aims to correctly classify synthetic data, as well as real data, whereas the generator aims to generate synthetic data as realistically as possible and deceive the discriminator. Accordingly, GAN optimisation ends up as a minmax game between the generator and the discriminator. Popular GAN variants include [113–115].

One limitation of the original GAN formulation is that it cannot provide any control mechanism over the output of the generator. Conditional generative adversarial networks

(cGANs) [116] are capable of conditioning both the generator and the discriminator and can be used for image-to-image translation.

### 3. Geoscience-Related Applications of Computer Vision and Pattern Recognition

In this Section, we present the most important types of geoscience-related applications of computer vision and pattern recognition. By far, the most frequently addressed application is land cover mapping, mostly on HSIs. Still, there are other types of applications, which include target/object detection, pattern mining in geoscience imaging data, shoreline boundary extraction (or extraction of other types of boundaries) and change detection. In addition, we dedicate a Subsection to image preprocessing. Although this is not a geoscience application itself, it involves core computer vision approaches and constitutes a building block in the context of such applications. Several works are hybrid, combining different approaches: superpixels with SVMs or CNNs, MRFs with SVMs, descriptors with deep neural networks, etc. The text in this Section is focused on the core approaches and the algorithmic elements introduced by each method. Information on datasets is provided in the discussion Section. For each application type, the works are presented following the order of the associated core approaches in Section 2. Within each paragraph associated with a distinct pair ‘type of application/core approach’, the works are presented mostly following chronological order, with some exceptions dictated by the need for logical cohesiveness.

#### 3.1. Land Cover Mapping

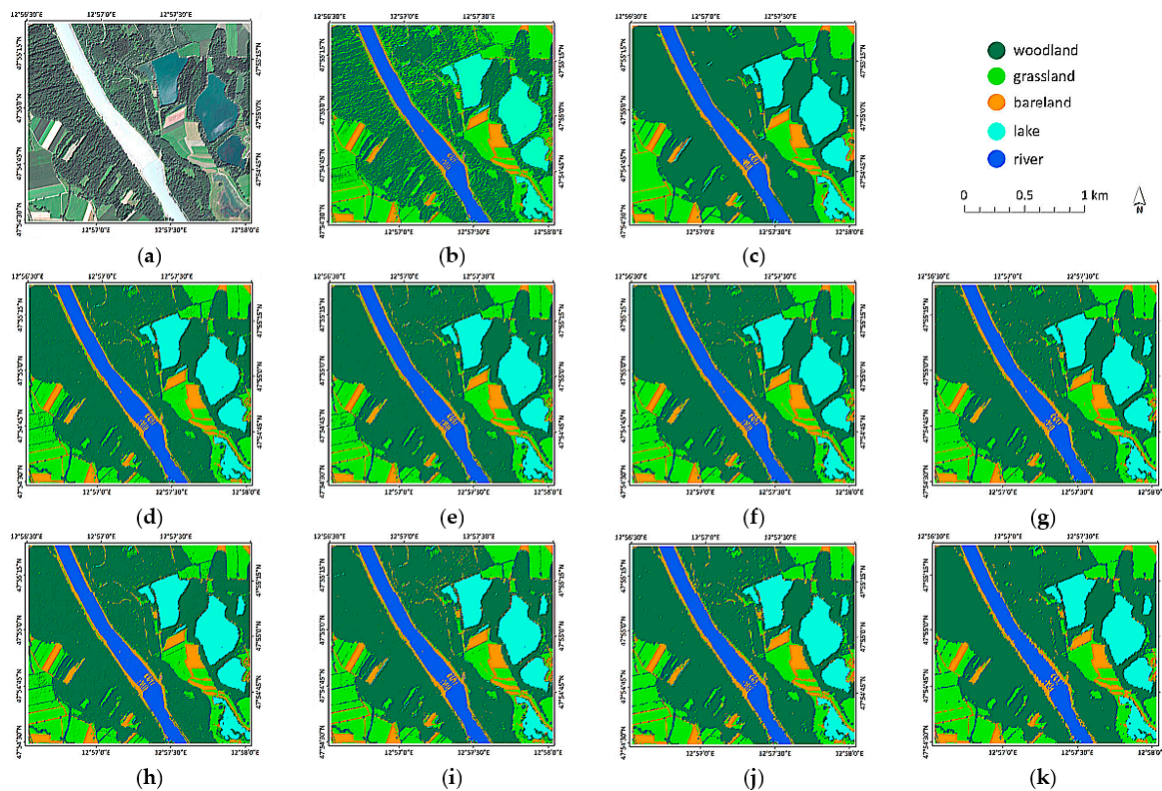
Most of the approaches presented in Section 2 have been applied for land cover mapping, i.e., assignment of each pixel of an HSI or other imaging modality to a certain class. This is essentially an image segmentation problem, which can be addressed by standard segmentation approaches, such as active contours or superpixels. It can also be posed as a clustering or classification problem at the pixel level, which can be addressed by standard clustering or classification approaches, as well as by deep neural networks. Some works address the classification of RS images as a whole. Still, such works are relevant to land cover mapping since an image classification-oriented framework can easily be modified to classify image patches.

Xia et al. [117] employed a multiscale AC for SAR image segmentation. Their method is based on the idea of integrating the nonlocal interactions between pairs of patches inside and outside each region. In addition, they introduce a multiscale strategy to speed up contour convergence and avoid local minima. Li et al. [118] proposed a semi-automated AC-based method for landslide inventory mapping (LIM) from bitemporal aerial orthophotos. Their method consists of two principal stages: thresholding based on change detection (see Section 3.5) and contour evolution defined by means of level-sets [52].

Tarabalka et al. [54] proposed another method for spectral-spatial classification of HSIs, combining SVMs and MRFs. First, a probabilistic SVM-based pixelwise classification of the HSI is applied. Second, spatial contextual information is used for refining the classification results obtained by means of MRF-based regularization. Yuan et al. [55] employed multitask joint sparse representation (MJSR) within the context of a stepwise MRF framework. The MJSR is aimed to reduce the spectral redundancy and retain necessary correlation in the spectral domain, whereas stepwise optimisation is aimed to further explore spatial correlation and enhance classification accuracy and robustness. Golipour et al. [59] integrated hierarchical segmentation with an MRF spatial prior in the context of the Bayesian framework. They extended statistical region merging to a hierarchical segmentation approach and extracted a multilevel fuzzy ‘border/no-border’ map, which is used to derive weighting coefficients. The latter are used to adjust the spatial prior of an MRF-based multilevel logistic model. Their method is aimed to address a common issue of MRF-based segmentation, i.e., over-smoothing.

Fang et al. [119] proposed a hybrid method for HSI classification, combining superpixels and SVMs. Their method starts with superpixel-based HSI oversegmentation and

separately employs three kernels utilizing both spectral and spatial information, within and between superpixels. The three kernels are combined and classified by an SVM. The same research group proposed another hybrid method for HSI classification, combining superpixels with dictionary-based representation [120]. Pixels within each superpixel are jointly represented by a set of common atoms from a dictionary via joint sparse regularization. The recovered sparse coefficients are utilized to determine the class label of each superpixel. Instead of directly using a large number of sampled pixels as dictionary atoms,  $k$ -singular value decomposition ( $k$ -SVD) learning is applied to simultaneously train a compact representation dictionary, as well as a classifier. Csillik [62] combined SLIC superpixels [63] with RFs, reducing the amount of time required to segment relatively very high resolution (VHR) RS images, while maintaining segmentation accuracy that is comparable to the one obtained by RF-based methods (Figure 4). Shi and Pun [121] used superpixels jointly with deep neural networks: they applied a 3D superpixel-based sample filling technique to cope with boundary misclassification, as well as a 3D recurrent convolutional network (3D RCNN) to further exploit spatial continuity and suppress noise.



**Figure 4.** Land cover mapping results obtained by the methods described by Csillik [62] (a) original WorldView-2 RGB image, (b) pixel-based classification, (c) multiresolution segmentation classification, (d–g) SLIC superpixel classification for various superpixel sizes, and (h–k) SLICO superpixel classification for various superpixel sizes.

Maulik et al. [122] used FCMs for land cover mapping in RS images. Their method is a variant of the differential evolution algorithm [123], which alters the mutation process, aiming to solve the optimisation problem more effectively. Qin et al. [124] proposed a variant of SLIC superpixels [65] for polarimetric SAR (PolSAR) image segmentation. Their method is aimed to guide the initialization of the cluster centers, associated with the  $k$ -means-derived superpixels in SLIC. Zhang et al. [125] proposed a subspace clustering method for land cover mapping in HSIs. Their method is based on the consideration that the pixels of each land cover class belong to a single subspace. In addition, they incorporated the limitations induced by the principle of the sparsity theory, in order to reduce data dimensionality. Wang et al. [126] presented a method for crop mapping in RS



images, using both supervised (RF) and unsupervised ( $k$ -means and GMMs) approaches. Reza et al. [127] proposed a clustering-based segmentation method for the estimation of rice yield from grain areas, using low altitude RGB aerial images collected by means of a rotary-wing type UAV. Image foreground and background are separated using graph-cuts. The extracted RGB image of the foreground is transformed into the Lab color space and pixels are clustered into regions with  $k$ -means, based on color information.

Apart from being directly employed for land cover mapping, some clustering-based methods have been proposed for spectral band selection in HSIs, which is aimed to aid the subsequent classification stage. Jia et al. [128] proposed such a spectral band selection method, named enhanced-fast density-peak-based clustering (E-FDPC). Their method is an extension of the more general fast density-peak-based clustering (FDPC), presented in [129] and is adapted on the hyperspectral band selection problem. Another spectral band selection method has been proposed by Yuan et al. [130], combining both the spectral and the spatial information of HSIs. Band selection is accomplished with clustering on the input HSI cube. In the band selection method of Wang et al. [131], the preliminary assumption is that the clusters of spectral bands consist of contiguous wavelengths. Accordingly, band selection is equivalent to finding the limits of the successive continuous intervals of wavelengths. This is achieved by means of a dynamic programming algorithm based on the maximization of an objective function. Zhai et al. [132] proposed a low-rank subspace clustering method for band selection in HSIs. Their method transforms the 3D input cube to a 2D matrix and determines a sparse representation coefficient matrix by the minimization of an objective function, which involves 3 weighted terms: its nuclear norm, the Frobenius norm of the noise component of the adopted affine data representation model and an 1D Laplacian regularizer. The minimization problem is solved using the alternating direction method of multipliers (ADMMs) [133].

Ham et al. [134] addressed several challenging issues in HSI classification including the problem of high dimensionality, the presence of multiple, potentially mixed, classes and the frequently limited quality of ground truth labelling. For these purposes, they proposed two RF-based classifiers, namely RF binary hierarchical classifier (RF-BHC) and RF-classification and regression tree (RF-CART). These classifiers are differentiated by the split criterion of nodes at the corresponding DTs. The best-basis binary hierarchical classifier (BB-BHC) [135] and random-subspace binary hierarchical classifier (RS-BHC) [136], are embedded in both classifiers. For the RF-BHC, the goal is to exploit the advantages of natural class affinity, while improving generalization in HSI classification using limited training samples. RF-CART is not directly affected by small sample size and potentially provides greater diversity within the forest, but typically produces large trees. Gislason et al. [137] proposed an RF-based method for land cover mapping, using 4 sources of data: one spectral (Landsat multispectral scanner, with four data channels) and three topographic (elevation, slope and aspect data, one channel each). Stumpf et al. [138] proposed a segmentation method for object-oriented mapping of landslides on VHR RS images. Initially, multi-scale segmentation based on region growing is performed to separate the various objects in the image. A number of different types of features, i.e., spectral, textural, geometrical, auxiliary (slope and hillshade) and combinations, is adopted for each object extracted. An RF classifier is trained to determine the importance of each such feature for landslides. Due to the rare appearance of the 'landslide' class, an iterative process is implemented to balance commission and omission errors of the classifications. Eisavi et al. [139] proposed an RF-based method for land cover mapping on spectral and thermal data. The classification results are derived using four combinations of images: spectral, thermal, combinations of spectral and thermal, and subsets of features extracted from the time-series of both thermal and spectral. Peerbhay et al. [140] proposed two RF-based methods for unsupervised land cover mapping on HSIs. Both methods derive an RF proximity matrix. Each HSI is divided to subimages and the two proposed classification methods are applied in each such subimage. In the first method, a reliable measure of outlier score, based on the RF proximity matrix, is evaluated for each pixel of the subimage. In the second method, the

eigen-decomposition of the RF proximity matrix is realized and an eigen-weight is assigned to each pixel of the subimage, according to [141]. In both methods, the resulting pixel classification is based on the sign of Anselin local Moran's I statistic [142]. Sun et al. [143] proposed an ensemble of RF and  $k$ -NN classifiers for land cover mapping using spectral and thermal images. Their experimentation leads to the conclusion that incorporating images of different thermal bands aid the end classification stage. Kalantar et al. [144] proposed an RF-based land cover mapping method based on UAV-acquired images. Their method employs the fuzzy unordered rule induction algorithm for object-based image analysis. The input images are preprocessed and segmented using a region growing variant. The segmentation parameters, including scale, shape, and compactness, are determined using combined feature space optimisation and the plateau objective function. Feature selection is performed by means of RF optimisation. This method outperforms both SVMs and DTs in a nine-class dataset acquired by a fixed-wing UAV.

Bazi et al. [145] proposed an SVM-based method for land cover mapping in HSIs, in which the SVM is parameterized by means of a genetic algorithm. Mantero et al. [146] addressed the problem of identification of samples drawn from unknown classes in RS images. The samples of unknown classes are distinguished from those of known classes, by means of a maximum a posteriori (MAP)-based classification rule. The estimation of the prior probabilities and probability distribution functions incorporated in this classification rule is performed by means of support vector regression. Foody and Mathur [147] proposed a novel approach to select informative training samples in the context of SVM-based land cover mapping. They considered that only support vectors affect the actual decision surface. In the context of the problem of distinguishing between Spring barley and Winter wheat, this consideration results in incorporating training samples from peat soil. The classification accuracy obtained is not affected significantly. The same authors [148] extended this work by using the same test area and an additional (third) crop type: sugar beet, in order to create alternative sets for SVM training. The same authors [149] proposed an SVM-based method, which, unlike other relevant methods, such as one-against-one or one-against-all, is only based on a single trained SVM classifier and the multiclass nature of the problem is addressed by the objective function adopted. Marconcini et al. [150] proposed a semi-supervised SVM variant for spectral-spatial classification of HSIs. Their method comprises three main stages: initialization, iterative semi-supervised learning and convergence. The final kernel matrix is derived as the weighted sum of two distinct kernel matrices, one for spectral and one for spatial data. Huang et al. [151] proposed a method for urban HSI classification using an SVM ensemble, which combines multiple spectral and spatial features at both pixel and object levels. Initially, principal component analysis (PCA) [152] is applied to multi/hyperspectral images for spectral feature extraction and only a small number of the corresponding principal components are maintained. Additional features extracted are GLCM [42], in order to exploit textural information, differential morphological profiles, which are defined as the successive differences of congener morphological profiles [153,154] and urban complexity index [155], which is based on the 3D Wavelet transform. Finally, three fusion algorithms, named C-voting, probabilistic fusion (P-fusion) and object-based semantic approach (OBSA), have been developed to optimally combine the different feature types for the final classification. Aiming at land cover mapping, Xu et al. [156] proposed a BoVW/SVM-based scheme for land cover mapping. Pasolli et al. [157] used SVMs in the context of an active learning-based framework, which combines spatial and spectral information for land cover mapping in VHR images.

Cheng et al. [158] proposed an LR-based method for feature selection and land cover mapping on RS Images. They extracted the deviance metric by subtracting the log-likelihood obtained for each feature and the log-likelihood of the corresponding saturated model. The selected features are used to estimate the parameters of an LR model. Li et al. [159] proposed another LR-based method for land cover mapping on HSIs. They adopted the Bayesian framework, employing the MAP criterion to extract the correspond-

ing a posteriori PDF. The conditional PDFs are estimated by means of the multinomial regression model. The parameters of the latter are estimated using the MAP criterion once more. For the conditional PDF of the second MAP criterion, an iterative algorithm, namely generalized expectation maximization (GEM) [160], is used. The mathematical formulation of this algorithm is based on the implementation of a block Gauss–Seidel iterative procedure, where the regression parameters of each class are the blocks of the matrices involved. In parallel, the prior PDF is considered for multilevel logistic MRF, which follows the Gibb’s distribution, providing spatial information supplied by the images. After substituting these results to the first MAP criterion, a combinatorial problem arises, which is solved following a graph-cut-based approach. The same group proposed a similar method for land cover mapping on HSIs [161]. Compared to [159], the main difference lies in the computation of the conditional PDFs. Starting from the consideration that the samples of a class lie in the same subspace, which is linearly independent from the subspaces related to other classes, it has been proved that the conditional PDFs follow a Gaussian distribution. Following some additional algebraic manipulations and considerations, it has also been proved that the problem of approximation of the conditional PDFs can be transformed to a multinomial logistic regression one.

Bruzzzone et al. [162] proposed an MLP-based method for land cover mapping, which follows a Bayesian framework and uses multi-temporal and multi-source RS images. Hu et al. [163] investigated the performance of two alternative methods for the estimation of impervious surfaces: (1) an MLP-based method, and (2) a SOM-based method. Their results show that the second method is slightly superior for this problem. D’Alimonte et al. [164] proposed an MLP-based method for phytoplankton determination in optically complex coastal waters, via estimating the Chla concentration, as well as the absorption of pigmented matter.

A large body of work addresses land cover mapping applications with the use of CNN-based architectures. One of the first CNN-based methods for land cover mapping was proposed by Makantasis et al. [165]. Their method hierarchically constructs high-level features, employs a CNN to encode spectral and spatial information, and an MLP for the end classification task. Hu et al. [166] proposed a CNN-based method, which embeds randomized PCA (R-PCA) in the network. Maggiori et al. [167] proposed a CNN-based method, adopting a two-stage training process: the first stage uses data from open street map (OSM), whereas the second stage fine-tunes the network with a few manually labelled images. Volpi et al. [168] proposed a CNN-based method for dense semantic labelling of sub-decimeter resolution images. The proposed architecture is called ‘full patch labelling by learned upsampling’ and consists of downsampling and upsampling blocks. This approach allows to densely label each pixel at the original resolution. Zhang et al. [169] proposed a complex-valued CNN (CV-CNN) for PolSAR image classification. Their CNN architecture is typical, yet every layer is extended to the complex domain, whereas training is performed with complex backpropagation. Introducing complex values in the network enables the use of both amplitude and phase information of complex SAR images. Scott et al. [170] tested CaffeNet [171], GoogleNet [31], and ResNet50 [30] for land cover mapping in high-resolution RS images. They explored transfer learning with and without fine-tuning, as well as data augmentation specialized for RS images. Xu et al. [172] proposed a bimodal land cover mapping method based on a two-branch CNN architecture. The first branch is a dual-channel CNN extracting spectral-spatial features from an HSI input. The second branch is a cascade-block CNN extracting features from a LiDAR or a high-resolution image. Li et al. [173] proposed a strategy for integrating multilayer features of CNNs for scene classification in RS images. Their strategy requires the synergy of two different CNN architectures, which both use a pre-trained model as a feature extractor. The first architecture passes the extracted feature maps into a series of fully connected layers. The result of each fully connected layer is passed into a dimensionality reduction module, which combines PCA and spectral regression kernel discriminant analysis (SRKDA) [174]. The second architecture receives multiscale images generated by means of Gaussian pyramids.

The feature map of each convolutional layer of the pre-trained model used in the second architecture is passed into a multiscale variant of the Fisher kernel framework [50]. This way, each feature map is represented by a Fisher vector. Each Fisher vector is passed to the dimensionality reduction module, as is the case with the feature maps of the first architecture. The two reduced feature maps are concatenated and used as input on an SVM. Cheng et al. [175], combined an autoencoder with a CNN (AECNN) for HSI classification. They used the autoencoder to enhance the non-linear features of the HSI [176] and reduce complexity. The enhanced image forms the input to a shallow CNN that consists of two convolutional layers, each followed by a dropout layer. Chen et al. [177] proposed a neural architecture search algorithm to enable the automatic design of CNNs for land cover mapping in HSIs. Cao et al. [178] proposed a CNN-based framework that employs active learning (AL) for land cover mapping in HSIs. MRFs are used, aiming to smooth class labels. Wu et al. [179] proposed cross-channel reconstruction network (CCR-Net), a CNN-based framework for classification of multimodal RS images. CCR-Net is a two-stream CNN, with both streams composed of the same CNN architecture that operates as a feature extractor. Each stream is dedicated to a different image modality. CCR-Net fuses the extracted multimodal features using an encoder-decoder scheme. Their method has been tested on datasets of LiDAR/SAR pairs of images. Mei et al. [180] proposed a quantization method for accelerating CNN-based HSI classification. The main idea is to replace the 32-bit single-precision floating-point numbers that are used for mathematical operations in a typical CNN, with low-bit integers. This becomes feasible by using two quantization methods: the first method, named step activation quantification (SAQ), restrains the input to the convolutional and fully connected layers. The second method, which is adopted from [181], is used to quantize network weights. Lin et al. [182] proposed an attention-aware pseudo-3D (AP3D) CNN for HSI classification. Each 3D convolutional layer is decomposed into three 2D convolutional layers that form a pseudo-3D (P3D) block. Different weights are assigned to each dimension of the extracted 3D features. In addition, two attention learning processes are employed for local and global feature learning. Dong et al. [183] proposed a weighted feature fusion of CNNs and graph attention networks for HSI classification. Their model consists of two branches: the first branch performs superpixel-level feature extraction using the graph attention network. SLIC-based superpixels [63] are extracted to form graph nodes, which are used by an encoder/decoder. The second branch is a CNN that consists of two pairs of a position attention module (PAM) with a channel attention module (CAM), as well as of two sets of convolutional layers. The extracted features of each branch are weighted and fused. Lu et al. [184] proposed an evolving block-based CNN (EB-CNN) for HSI classification. Their method employs a genetic algorithm (GA) to optimize the width of each CNN layer, as well as the CNN architecture depth.

Ienco et al. [185] evaluated the ability of an LSTM model to perform land cover mapping considering multitemporal spatial data derived from a time-series of satellite images. They carried out experiments on two different data sets considering both pixel-based and object-based classifications. Their results show that LSTMs are competitive compared with state-of-the-art classifiers and may outperform classical approaches in the presence of under-represented and/or highly mixed classes. Their results also show that the alternative feature representation generated by the LSTM can enhance the performance of standard classifiers. Mou et al. [186] proposed an RNN-based method for HSI classification, which analyzes hyperspectral pixels as sequential data and then determines information categories via network reasoning. Maggiori et al. [187] proposed an RNN-based method, which learns an iterative process in order to refine the results obtained by CNNs for land cover mapping on satellite images. Rußwurm et al. [188] proposed an RNN-based method for land cover mapping of satellite images. They adapted an encoder structure with convolutional recurrent layers in order to approximate a phenological model for vegetation classes based on a temporal sequence of Sentinel 2 images. They visualized internal activations over a sequence of cloudy and non-cloudy images and found several recurrent cells that reduce the input activity for cloudy observations. This indicated that their net-



work has learned cloud-filtering schemes solely from input data, alleviating the need for tedious cloud-filtering as a preprocessing step for many earth observation approaches. Ndikumana et al. [189] and Ho Tong Minh et al. [190] demonstrated that two RNN-based classifiers outperform three classical methods:  $k$ -NN, RF and SVM with respect to classification accuracy, when tested for agricultural land cover on Sentinel 1 images. Hang et al. [191] proposed a cascaded RNN-based method for HSI classification, using GRUs to explore the redundant and complementary information of HSIs. Their method mainly consists of two RNN layers. The first layer is used to eliminate redundant information between adjacent spectral bands, whereas the second one aims to learn the complementary information from nonadjacent spectral bands. In addition, taking into account the rich spatial information of HSIs, they extended their method to its spectral-spatial counterpart by incorporating extra convolutional layers.

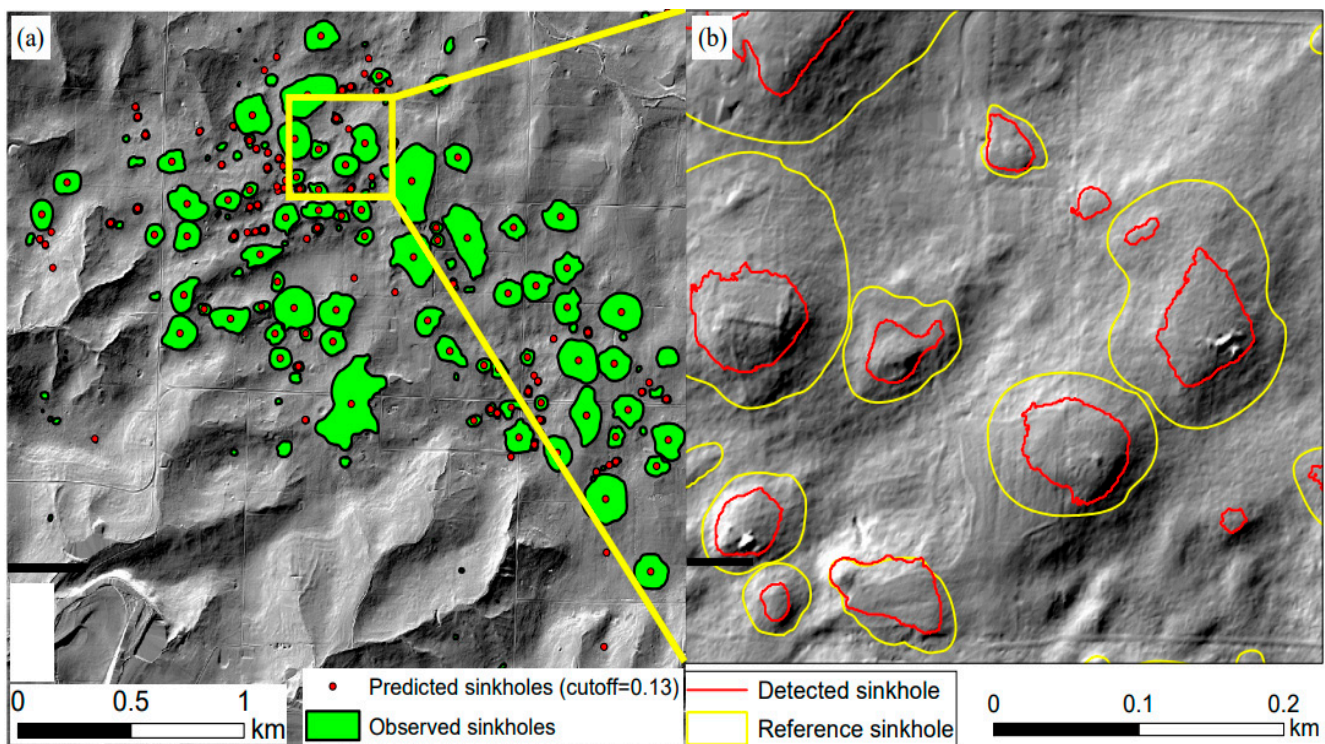
### 3.2. Target Detection

Object detection is important for a wide range of RS and geoscience-related applications, such as intelligent monitoring, precision agriculture and geographic information system (GIS) updating. This motivated intensive research for the development of object detection methods in various imaging modalities, which has been reinvigorated with the breakthrough of CNN-based architectures.

Lee et al. [192] proposed a clustering-based method for tree detection and tree parameter estimation in pine managed forests, using airborne LiDAR data. Tree tops are determined as seed points via an algorithm that compares the heights of each point, and a region growing method is applied to determine tree boundaries. This region growing method is a variation of the watershed segmentation algorithm [193] adapted to the raw LiDAR data. An agglomerative hierarchical clustering algorithm is applied in order to cope with oversegmentation.

Kim et al. [194] proposed an SVM-based method for human detection and activity classification on Doppler radar. Their method is tested separately for human detection and human activity classification. Kim et al. [195] proposed an LR-based method for sinkhole detection and characterization on LIDAR-derived digital elevation model (DEM) data. Initially, they employed a data preparation stage using GIS software. Exploiting this topographic representation, a set of 16 features is extracted. An LR model is used to determine the existence of sinkholes, prior to the generation of a probabilistic sinkhole susceptibility map (Figure 5).

Martorella et al. [196] proposed an MLP-based method for automatic target recognition on fully polarimetric inverse synthetic aperture radar (Pol-ISAR) images. As a first stage, they apply a technique named Pol-CLEAN to extract the brightest scattering centers. As a second stage, each extracted scattering center is decomposed with the use of Cameron's decomposition [197]. As a third stage, the decomposed scattering centers are 3D feature vectors, which become the input of the MLP. Tatavat et al. [198] proposed an MLP-based method for cloud detection on satellite images in various weather conditions. Their method uses the temperature and the reflectance of the region of interest and employs an MLP with one hidden layer and the hyperbolic tangent activation function.



**Figure 5.** Sinkhole detection results obtained by the method of Kim et al. [195] (a) map of sinkhole susceptibility, (b) reference sinkholes (yellow) and detected sinkholes (red).

Chen et al. [199] proposed a hybrid deep neural network for vehicle detection in satellite images. The last convolutional and max pooling layers are divided into multiple blocks, where each block has a different filter size or max pooling field size, resulting in multiscale features. Cheng et al. [200] proposed a rotation invariant CNN-based method for detecting objects in VHR images. They inserted a rotation invariant layer into AlexNet [27] and optimized a rotation-invariant objective function. Ding et al. [201] proposed a CNN-based method for target recognition in SAR images, employing data augmentation by means of translation, speckle noise and pose synthesis. Long et al. [202] proposed a CNN-based framework for object localization in RS images. Their framework consists of three stages: region proposal, feature extraction/classification, and object localization. In the first stage a selective search algorithm [203] is used to propose possible, category-independent ROIs. The second stage extracts image patches from the proposed ROIs. The patches extracted are used as inputs to AlexNet [27] and GoogleNet [31]. Both these models are used and combined. The second stage of the framework is aimed to remove redundant bounding boxes and combines non-maxima suppression and a method for bounding box optimisation. Cheng et al. [204] proposed a cascaded end-to-end CNN for road detection and centerline extraction on VHR RS images. Their method employs two cascaded NNs: the road detection network and the centerline extraction network. The road detection network is an autoencoder that performs semantic segmentation between two classes: road and background. The centerline extraction network is also an autoencoder, which takes as input the extracted feature maps of the road detection network and produces an image with the extracted centerline, which is refined by means of a thinning algorithm. Shao et al. [205] proposed a CNN-based method using multiscale features (MF-CNN) for cloud detection. Their method can detect and distinguish pixels with cloud, thick cloud or no cloud. Hsieh et al. [206] proposed an object detection and counting method based on UAV-acquired data. Their method encompasses layout proposal network (LPN), which is similar to region proposal networks (RPNs) [207] and uses spatial layout information. Kellenberger et al. [208] proposed a CNN-based method for detecting mammals in UAV-

acquired images. Their work addresses class imbalance, intra-class heterogeneity, inter-class homogeneity and background heterogeneity. Detection performance is further enhanced by means of curriculum learning [209] and hard negative mining [210]. Zhang et al. [211] developed a UAV-based object tracking and 3D localization system. Their system uses TrackletNetTracker (TNT) [212], which is a multi-object tracking method. TNT uses both spatial and temporal information and allows for the realization of the continuous trajectory of each detected object across frames. The UAV camera is self-calibrated using a monocular semi-direct visual odometry algorithm. The 3D localization of the detected objects is enabled by computed camera parameters, determined by the multi-view stereo (MVS) [213] method. Zhang et al. [214] proposed a CNN-based anomaly detection framework for the identification of multivariate geochemical anomalies. Their method adopts pixel-pair feature (PPF) [215] to augment the available dataset by recombining pixel pairs of each labelled sample.

Since 2016, several YOLO-based (see Section 2) methods have been proposed for target detection in geoscience-related applications. Xu and Wu [216] proposed multi-receptive fields fusion YOLO (MRFF-YOLO), based on the YOLO-v3 variant [217]. Their architecture encompasses four detection layers, instead of three in the classical YOLO-v3, whereas the problem of gradient vanishing is addressed by replacing convolutional layers with dense blocks. The same authors later proposed feature-enhanced YOLO (FE-YOLO) [218], a YOLO-v3-based single-stage detector, aiming to: (1) improve detection accuracy for small remote sensing targets, (2) detect densely distributed targets, and (3) realize real-time performance. Qing et al. [219] proposed RepVGG-YOLO, aiming to address some inherent difficulties for target detection in RS, such as the complex background, the large differences in target sizes and the uneven distribution of rotating objects. Their method employs: (1) RepVGG [220] as the backbone for feature extraction, (2) an improved feature pyramid network (FPN) and a path aggregation network (PANet) in order to reprocess feature output, and (3) circular smooth label (CSL) in order to enhance detection accuracy for objects in various angles. Wang et al. [221] proposed a variant of YOLO-v3 to facilitate the inspection of the opium poppy illegal cultivation through UAVs. Their method employs the ResNext module [222] and group convolutions to reduce model parameters, as well as atrous spatial pyramid pooling (ASPP) [223] to enhance local feature extraction, as well as to aid the use of contextual information. Jamali et al. [224] introduced a YOLO-v4 [225] variant for target detection in RS images. They use non-maximum suppression (NMS) thresholds in order to improve the detection accuracy of overlapping horizontal bounding boxes, whereas they address the anchor frame allocation problem in YOLO-v4 with two allocation schemes.

Ke et al. [226] proposed the global context boundary-aware network (GCBANet), aiming to improve SAR ship instance segmentation. Their network incorporates two blocks: a global context information modeling block (GCIM-Block) and a boundary-aware box prediction block (BABP-Block). GCIM-Block is used to capture spatial global long-range dependencies of ship contextual surroundings, whereas BABP-Block is used to estimate ship boundaries, improving the cross-scale box prediction. Li et al. [227] combined a CNN and a multiple-layer Transformer (see Section 2). The Transformer is used to aggregate global spatial features on multiple scales and model the interactions between pairwise instances. The pre-trained CNN is used as the backbone for RS image feature extraction, whereas the attention mechanism is employed for feature reweighting, in order to reduce the differences between source and target datasets. Xiao et al. [228] combined a shifted window (SWIN) Transformer-based encoding booster with a U-Net for building detection in RS images. Their architecture is called SWIN Transformer-based encoding booster U-shaped network (STEB-UNet). An important characteristic of STEB-UNet is that the features obtained at different levels by the encoding booster are fused with U-Net features to compensate for the lack of large-scale semantic extraction in U-Net. Chen et al. [229] proposed a method combining SWIN Transformer and MAP-Net [230] for building extraction from satellite

images. Their method uses SWIN Transformer to extract multiscale features and MAP-Net as a head network to fuse and refine features.

### 3.3. Pattern Mining in Geoscience Imaging Data

Mining spatiotemporal patterns in multiple scales, to understand physical, chemical and biological processes, which affect solid Earth, oceans and atmosphere, provide an obvious opportunity for pattern recognition applications. The availability of large amounts of data from various imaging modalities, along with other types of data (physical/chemical measurements, demographic data etc), create application opportunities, more so if we consider that the multidimensional nature of such patterns often complicates discovery that is guided by mere observation.

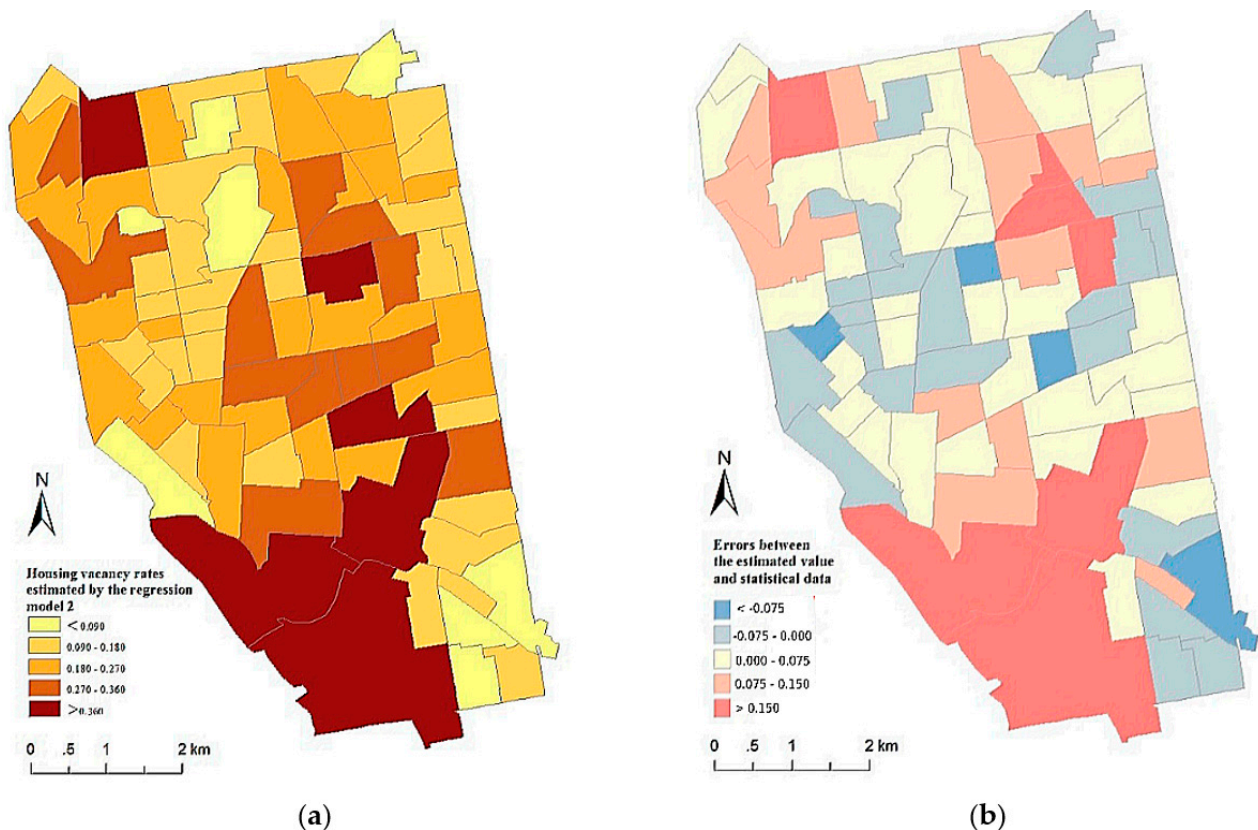
Joseph et al. [231] derived subpixel vegetation–impervious surface–soil (VIS) fractions from the Landsat ETM+ multispectral bands, and then used the geographically weighted regression (GWR) model to investigate the variation of population density with VIS variables and their derivatives. Unlike the ordinary least square (OLS) model, their model accounts for spatial non-stationarity. The study reveals that three VIS variables are significant in explaining population density: the mean values of houses fraction image and vegetation fraction image, as well as the standard deviation of vegetation fraction image. Hengl et al. [232] proposed a method for predicting soil properties utilizing information on agricultural management. For this purpose, they use two prediction models: a linear regression-kriging model [233,234] and an RF. Stevens et al. [235] proposed an RF-based semi-automated dasymetric method for gridded population density prediction, which coevaluates country-specific census data, land cover data and geospatial data. A feature importance procedure and the final prediction of a country-wide, pixel-level map of log population densities are implemented via an RF predictor. Georganos et al. [236] developed an expanded implementation of RFs, namely geographical RFs (GRFs), to be used as a predictive and exploratory tool to model populations as a function of RS data. The major goal of GRFs is to exploit the spatial information supplied by RS data, consisting of several local RFs. These local RFs, which constitute local predictors, operate in an adaptively determined neighborhood of the training data points. In parallel, a classical RF is trained using all the available training data, operating as global predictor. The final prediction for a data point is based on the fusion of the global prediction and the corresponding results of its nearest local predictor. The population modeling as a function of RS data is achieved by using the land cover mapping imagery, as well as demographic data associated with the actual population distribution. The results show that the inclusion of spatial information can enhance the prediction accuracy provided by a classical RF.

Sun et al. [237] employed an SVM variant, named  $v$ -SVM, which is used as a regressor to estimate Chlorophyll *a* (Chl<sub>a</sub>) concentration in inland turbid lake waters. The regression process is based on measurements of lake water quality parameters obtained within a period of approximately 2 weeks.

Kokaly et al. [238] proposed a method for determining leaf biochemistry by means of spectroscopic measurements. Two spectral analysis techniques are adopted: continuum-removal [239] and band-depth normalization. Continuum removal is applied to dry leaf spectra in order to broad their absorption features. The band depths from the continuum-removed spectra are normalized, in order to minimize the sensitivity to factors affecting RS measurements such as atmospheric absorptions. The normalized band depths are used by a stepwise LR model in order to identify chemistry-correlated wavelengths. Lee et al. [240] proposed an LR-based method for landslide susceptibility mapping. Their method uses 8 landslide occurrence factors: topographic slope, topographic aspect, topographic curvature, distance from drainage, lithology, distance from lineament, land use, and vegetation index. These factors are used in two different ways: (1) a different regression model is created for each factor, (2) one regression model encompasses all factors. Dardel et al. [241] proposed an LR-based method for trend analysis in normalized difference vegetation index (NDVI). Their analysis addresses the desertification of Sahel and is performed on data



of up to 30 years. Du et al. [242] employed LR to investigate the relationship between the housing vacancy rate (HVR) and the census tract level. They used data acquired by the Jilin1-03 satellite, which consist of high spatial resolution night-time light images, as well as digital orthoimages and parcel data. From each modality, a factor is determined, including human-activities, landuse structure, and physical environment. These factors are used as explanatory variables in two stepwise multivariate LR models (Figure 6). Tien Bui et al. [243] combined the evidential belief function (EBF) and LR for predicting flood susceptibility. Three ensemble variants were investigated, taking into account different subsets of 10 conditioning factors: altitude, slope angle, plan curvature, topographic wetness index (TWI), stream power index (SPI), distance from river, rainfall, geology, land use and NDVI, along with different weighting coefficients.



**Figure 6.** Human vacancy rate (HVR) estimations obtained by the LR-based method of Du et al. [242] (a) estimated HVR in Buffalo, (b) differences between the estimated value and the Census Bureau statistical data.

D’Alimonte et al. [164] proposed an MLP-based method for phytoplankton determination in optically complex coastal waters, via estimating the Chla concentration as well as the absorption of pigmented matter. Corsini et al. [244] proposed an MLP-based method for the estimation of optically active parameters (OAP) in case II waters on RS data. Specifically, the OAPs which are estimated are: (1) the chlorophyll and the other pigments contained in phytoplankton, (2) the non-chlorophyllous particles and (3) the yellow substance (or dissolved organic matter, DOM). Ozturk et al. [245], employed MLPs and Markov chain (MC) models for urban growth simulation using satellite images.

Al Najjar et al. [246] explored the ability of GANs to improve the performance of other machine learning models in the context of landslide susceptibility mapping applications. At first, 156 slide locations along with 15 conditioning factors were provided as input to five different machine learning models, which included DTs, RFs and SVMs. Synthetic data were generated with the use of a GAN, and combined with the original data. The newly

formed dataset was used by the models, resulting in improved performance for all models, with the exception of RF.

### 3.4. Boundary Extraction

Boundary extraction in RS and geoscience-related images is naturally associated with image segmentation and most frequently addresses coastline extraction. Standard image segmentation approaches, such as ACs or clustering-based segmentation, along with more recent CNN-based methods, have been applied in this direction.

Sukcharoenpong et al. [247] proposed a method for coastline extraction, which fuses HSIs with LiDAR-generated DEMs. Their method reaches an initial solution based on object spectra and a knowledge-based segmentation scheme. An AC refines this initial solution at subpixel level. Liu et al. [248] combined edge-based and region-based ACs for coastline detection in PolSAR images. Modava et al. [249] proposed a multi-stage pipeline for coastline extraction in high resolution SAR images. In the first stage, they perform fuzzy clustering with spatial constraints. This is followed by the application of Otsu's binarization algorithm [250] and morphological filtering. In the final stage, they refine the segmentation results by means of a level-set AC. Sun et al. [251] combined ACs and CNNs for building boundary extraction (Figure 7). They investigated two variants: the first integrates ACs into the CNN construction process, whereas the second builds footprint detection with a CNN and uses AC for post processing.

### 3.5. Change Detection

The increasing amount of SAR image data for Earth observation triggered research for the computational analysis of time-series of images acquired on the same geographical area. This can be carried out either with supervised classification (e.g., for producing thematic maps or maps of land cover transitions) or with unsupervised change detection (e.g., for generating change detection maps associated with damages caused by natural disasters or with land cover modifications) [1–5,252].

Bazi et al. [253] formulated unsupervised change-detection in RS images as a segmentation problem. In this context, the discrimination between changed and unchanged regions in the difference image is achieved by defining an energy functional minimized by means of a level-set. The difference image is analysed at multiple resolutions to enhance robustness against noise and initialization.

Gong et al. [254] proposed an MRF-based method for change detection in SAR images. Their method classifies changed and unchanged regions by means of FCM clustering using an MRF energy functional. The latter is defined to encompass an extra term modifying pixel membership. This term is contingent upon different situations and is ultimately established utilizing the least-square method.





**Figure 7.** Building boundary extraction results obtained by 5 different methods in 5 test scenes, as described in the work of Sun et al. [251]. Areas in green denote true positive, areas in blue denote false negative and areas in red denote false positive, at the object level.

Zheng et al. [255] applied probabilistic patch-based (PPB) [256] filtering to suppress speckle noise. The difference image of the given time-series, which consists of two images, is extracted using absolute difference, as well as absolute logarithmic difference. The first resulting image is mean-filtered, whereas the second one is median-filtered. The weighted sum of the two filtered images forms a final image, in which the  $k$ -means clustering is applied to separate pixels in two clusters: 'changed' and 'not changed'. Another clustering-based method for unsupervised change detection in multitemporal SAR images has been proposed by Ghosh et al. [257]. A pseudotraining set is created using CVA, in a similar fashion to the previously mentioned SVM-based method of Bovolo et al. [252]. Aiming to exploit the spatio-contextual information in the difference image, the gray-level as well as the mean gray-level of the 8 intermediate neighbors of each pixel, are used as features. Two fuzzy clustering algorithms, FCM and Gustafson-Kessel, are used to separate pixels in two clusters: 'changed' and 'not changed'. Two stochastic optimisation techniques, i.e., annealing and genetic algorithms, are employed to guide the convergence of the utilized clustering algorithms. Leichtle et al. [258] proposed a clustering-based change detection method for buildings in VHR images. They utilized both spectral and textural domains, as well as RGB and near-infrared (NIR) channels. The final, PCA-derived feature vectors are clustered by simple  $k$ -means.

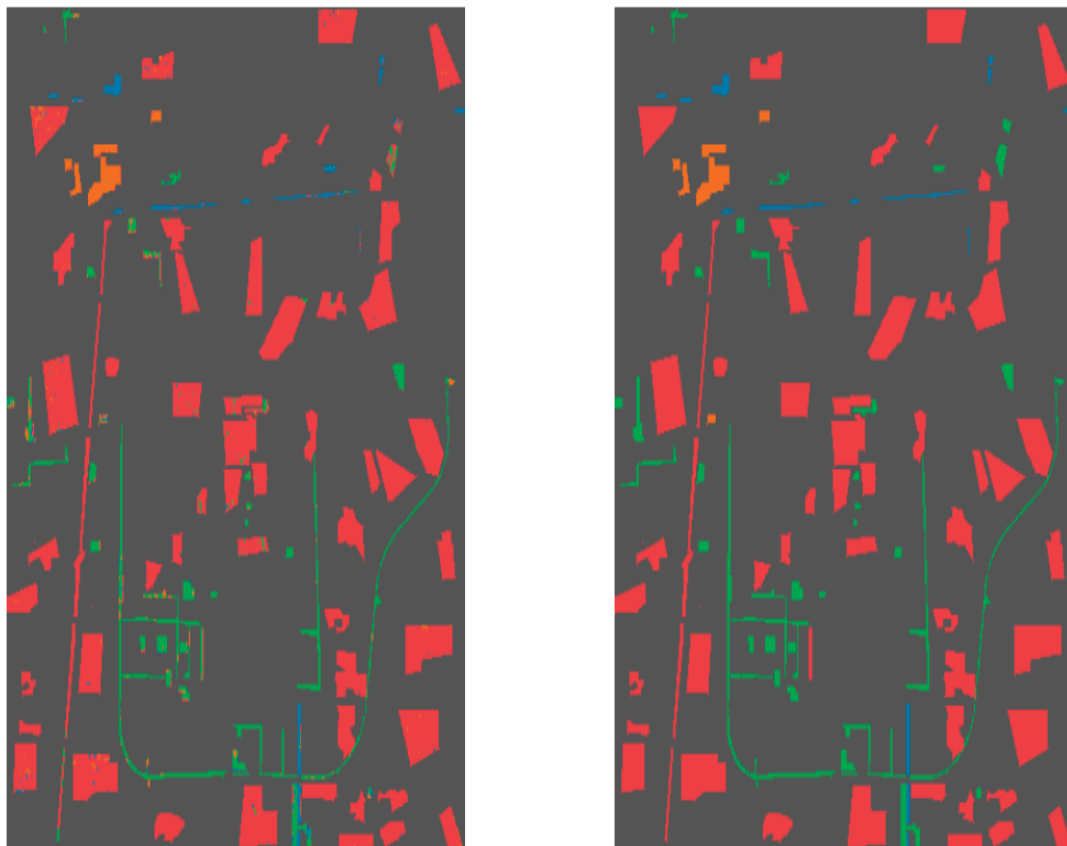
Bovolo et al. [252] addressed the problem of unsupervised change detection of geographical areas, using time-series of multispectral RS images of the same area. Their method comprises three main stages. A pseudotraining set is created in the first stage by means of change vector analysis (CVA) [259]. Given two images illustrating the same area at different times, CVA subtracts the spectral feature vectors of these images, and marks pixels related to the resulting vectors as 'changed' or 'not changed', comparing vector magnitudes with a threshold. The latter can be determined by the Bayes decision rule [252], considering the statistical parameters of vector magnitude distribution. Only those pixels associated with vector magnitudes far from this threshold are selected for the pseudotraining set, so as to mitigate ambiguity. In the second stage, a binary SVM is trained in a semi-supervised fashion, using the pseudotraining set. The third stage aims to determine the optimal combination of SVM parameters, considering two criteria, which are based on: (1) the data of the pseudotraining set, and (2) a similarity measure between all potential solutions.

Khurshid et al. [260] proposed a multinomial LR-based method for segmentation and change detection, aiming at assessing damages on multi-temporal images, available in 4 different spectral bands. The segmentation algorithm aims to extract the built-up areas using a series of transformations and a binomial LR model. For damage assessment, 6 change detection techniques are applied 4 times, one for each of the 4 pairs of corresponding spectral bands of the transformed images, before and after a damage. The extracted feature vectors are used to classify the damages in three classes: high, moderate and low, by means of a multinomial LR model. Tan et al. [261] introduced an LR-based ensemble of classifiers for change detection in high-resolution RS images. Spectral, texture, and morphological features are used to create difference images. The ensemble combines the extreme learning machine (ELM), multinomial logistic regression (MLR), and  $k$ -NN classifiers. Molin et al. [262] proposed an LR-based method for change detection in SAR images. Initially, they subtract a reference image from a monitored one and normalize the resulting image. They classified pixels as 'changed' or 'not changed' by means of an LR model.

Pacifici et al. [263] proposed a method for change detection in VHR images, using pulse-coupled neural networks (PCNNs). Any change is detected by quantifying the similarity between two PCNN signals, associated with two image instances. Salmon et al. [264] proposed an MLP-based method for the detection of new human settlements. Their method uses a sliding window operating as a feature extractor over hyper-temporal, multi-spectral images. The extracted features are multi-spectral time-series. Roy et al. [265] proposed an MLP-based change detection method. For this purpose, the difference image (DI) is produced using CVA and an ensemble of MLPs is used for the classification.



Zhao et al. [266] developed SiamCRNN, a deep siamese convolutional multilayer RNN for change detection in multisource VHR images. Their model consists of a deep siamese CNN (DSCNN), a stack of multiple LSTM units [33], and a series of fully connected layers. Lyu et al. [267] proposed an LSTM variant to acquire and record the change information of long-term sequence RS data (Figure 8). In particular, a core memory cell is utilized to learn the change rule from information on binary changes or multiclass changes. Three gates are utilized to control the input, output and update of the LSTM model for optimisation. In addition, the learned rule can be applied to detect changes and transfer the change rule from one learned image to another. Mou et al. [268] developed an end-to-end trainable recurrent CNN (ReCNN) for change detection in multispectral imagery. Their architecture receives two images and each image is passed through three main components. The first component is a convolutional sub-network that extracts spectral-spatial features. The second component is a recurrent sub-network that receives the feature maps extracted and calculates hidden states. The feature maps along with the calculated hidden states of the first image are the input to a second recurrent sub-network, which analyses the temporal dependence of the two images. The last component, consists of fully connected layers that loop through the sequence of the second recurrent sub-network.



**Figure 8.** Change map obtained by the land cover change detection method of Lyu et al. [267] and the ground truth map, where the unchanged areas are shown in red, the changed region of the city expansion is shown in green, the changed soil region is shown in orange and the changed water areas are shown in blue.

### 3.6. Image Preprocessing

Image preprocessing, including denoising and enhancement, is a part of most computer vision pipelines. In this Subsection, we present some relatively recent works, dedicated to the preprocessing of geoscience-related images. It is no surprise that CNNs and GANs are also prevalent in this area.

Yuan et al. [269] proposed a spectral-spatial deep residual CNN (HSID-CNN) for HSI denoising. Their architecture consists of a 2D CNN and a 3D CNN, which extract spatial and joint spectral-spatial features, respectively. Both types of features extracted are multiscale. A residual learning strategy is introduced to the network to ensure training stability, as well as reduce degradation [30]. Multilevel feature representation is enabled by skip connections. Li et al. [270] proposed a random-drop data augmentation method in order to aid mineral prospectivity mapping. Their method generates Gaussian-distributed random loci from the entire study area and generates a balanced dataset adopting an approach combining [271,272] in order to train a custom 12-layer CNN. Molini et al. [273] proposed Speckle2Void, a self-supervised blind-spot Bayesian framework for SAR image despeckling. Their framework adopts and improves the blind-spot CNN architecture of Laine et al. [274]. It consists of four branches, with each one processing four rotated input image versions and calculating a receptive field in a specific direction. Each receptive field is shifted, rotated back to its original orientation and concatenated with the other receptive fields. The concatenated receptive fields are connected to a series of 2D convolutions to generate inverse gamma parameters for each pixel. Unlike [274], the weights of the branches are shared in pairs, instead of being shared for all four branches. There are two variants of Speckle2Void: in the first variant, each branch follows the pattern of a typical CNN architecture, whereas in the second variant, non-local layers are added and operate as a dynamic weighted function of the feature vectors.

Liu et al. [275] proposed a pan-sharpening GAN (PSGAN) for RS images. The term pan-sharpening refers to mapping a high-resolution (HR) panchromatic (PAN) image and a low-resolution (LR) multispectral image to a HR multispectral image. For this purpose, they adopt two alternative approaches: the first approach is to feed each image into a different subnetwork, which extracts hierarchical features. The features extracted are used as input to an autoencoder-based network. The latter generates the final HR multispectral image. The second approach provides the two images in stacked form to the generator. The discriminator is common for both approaches and predicts the probability of each input being an unprocessed HR multispectral image or a pan-sharpened multispectral image. Pan et al. [276] proposed a GAN-based method for cloud removal in satellite images, named spatial attention GAN (SpA-GAN). They introduced a spatial attention mechanism, which enables the network to focus more on semantically substantial regions, as the clouds. The generator consists of four spatial attentive blocks, which aid cloud recognition, and two residual blocks, which aid image reconstruction without the clouds.

#### 4. Discussion

In this Section, we discuss overall trends and aspects of the applications presented, as well as the main challenges raised. Table 1 summarizes the articles presented in Section 3. It can be noted that the DL paradigm has recently been prevalent in computer vision applications in geoscience. However, the successful results obtained by DL-based methods come with the price of several issues accompanying DL. The most important issue is the availability of large datasets, more so in the case of supervised learning methods, which require labelled data. Another issue of DL-based methods comes at the price of computational cost, especially in the case of training. For CNNs in particular, which is the most frequently employed DL approach, the bottleneck in terms of processing time comes from convolution operations. The availability of dedicated hardware, such as GPUs, is critical. Cloud-based platforms such as Google's Dynamic World are important in this sense. Pre-DL methods, such as superpixels, often provide an alternative of low computational cost. Still, there are differences between pre-DL methods: SVMs demonstrate better generalization than shallow MLPs. Active contours and superpixels usually provide an unsupervised framework which works well in certain segmentation settings. As one could expect, there are several hybrid methods, combining SVMs with MRFs, CNNs with superpixels, CNNs with RNNs, etc.

**Table 1.** Summary of computer vision applications presented in Section 3.

| Application        | Article                                                   | Core Methodology                                                                                                                                                                                                     | Data Type                                                                                                              | Supervision Type                                                                                   | DL-Based                                          |
|--------------------|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|---------------------------------------------------|
| Land Cover Mapping | [117,118]                                                 | AC                                                                                                                                                                                                                   | MSI: [118]<br>SAR: [117]                                                                                               | Unsupervised                                                                                       | No                                                |
|                    | [57,61]                                                   | MRF                                                                                                                                                                                                                  | HSI: [55,59]                                                                                                           | Supervised                                                                                         | No                                                |
|                    | [54,62,119,121,124,126,173,183,187]                       | SVM/MRF: [54]<br>SVM/Superpixels: [119]<br>RF/Superpixels: [62]<br>CNN/Superpixels: [183]<br>CNN/RNN/Superpixels: [121]<br>Clustering/Superpixels: [124]<br>RF/Clustering: [126]<br>SVM/CNN: [173]<br>CNN/RNN: [187] | MSI: [62,126,173,187]<br>HSI: [54,119,121,183]<br>SAR: [124]                                                           | Supervised: [54,62,119,121,173,183,187]<br>Unsupervised: [124]<br>Supervised & Unsupervised: [126] | No: [54,62,119,124,126]<br>Yes: [122,174,184,188] |
|                    | [120]                                                     | Superpixels                                                                                                                                                                                                          | HIS                                                                                                                    | Supervised                                                                                         | No                                                |
|                    | [122,125,127,128,130–132]                                 | Clustering                                                                                                                                                                                                           | MSI: [122,127]<br>HSI: [125,128,130–132]                                                                               | Unsupervised                                                                                       | No                                                |
|                    | [134,137–140,143,144]                                     | RF                                                                                                                                                                                                                   | MSI: [137–139,143,144]<br>HSI: [134,140]<br>Other: [137–139,143]                                                       | Supervised                                                                                         | No                                                |
|                    | [145–151,156,157]                                         | SVM                                                                                                                                                                                                                  | MSI: [146,148,149,156,157]<br>HSI: [145,147,150,151]<br>Other: [151]                                                   | Semisupervised: [150]<br>Supervised: [145–151,156,157]                                             | No                                                |
|                    | [158,159,161]                                             | LR                                                                                                                                                                                                                   | MSI: [158]<br>HSI: [158,159,161]                                                                                       | Supervised                                                                                         | No                                                |
|                    | [162–164]                                                 | MLP                                                                                                                                                                                                                  | MSI: [162,163]<br>SAR: [162]<br>Other: [164]                                                                           | Supervised                                                                                         | No                                                |
|                    | [165–170,172,175,177–182,184]                             | CNN                                                                                                                                                                                                                  | MSI: [168,170,172,181]<br>HSI: [165,166,172,175,177–180,182,184]<br>SAR: [169,179]<br>LiDAR: [172,179]<br>Other: [167] | Supervised                                                                                         | Yes                                               |
|                    | [185,186,188–191]                                         | RNN                                                                                                                                                                                                                  | MSI: [185,188]<br>HSI: [186,191]<br>SAR: [189,190]                                                                     | Supervised                                                                                         | Yes                                               |
| Target Detection   | [192]                                                     | Clustering                                                                                                                                                                                                           | LiDAR                                                                                                                  | Supervised                                                                                         | No                                                |
|                    | [194]                                                     | SVM                                                                                                                                                                                                                  | Other                                                                                                                  | Supervised                                                                                         | No                                                |
|                    | [195]                                                     | LR                                                                                                                                                                                                                   | LiDAR, Other                                                                                                           | Supervised                                                                                         | No                                                |
|                    | [196,198]                                                 | MLP                                                                                                                                                                                                                  | SAR: [196]<br>Other: [198]                                                                                             | Supervised                                                                                         | No                                                |
|                    | [199–202,204–206,208,211,214,216,218,219,221,224,226–229] | CNN                                                                                                                                                                                                                  | MSI: [199,200,202,204–206,208,211,214,216,218,219,221,224,227–229]<br>SAR: [201,226]                                   | Supervised                                                                                         | Yes                                               |
| Pattern Mining     | [232,235,236]                                             | RF                                                                                                                                                                                                                   | Other                                                                                                                  | Supervised                                                                                         | No                                                |
|                    | [237]                                                     | SVM                                                                                                                                                                                                                  | Other                                                                                                                  | Supervised                                                                                         | No                                                |
|                    | [238,240–243]                                             | LR                                                                                                                                                                                                                   | MSI: [242]<br>Other: [238,240,241,243]                                                                                 | Supervised                                                                                         | No                                                |
|                    | [164,244,245]                                             | MLP                                                                                                                                                                                                                  | MSI: [245]<br>Other: [164,244,245]                                                                                     | Supervised                                                                                         | No                                                |
|                    | [246]                                                     | GAN                                                                                                                                                                                                                  | Other                                                                                                                  | Supervised                                                                                         | Yes                                               |



Table 1. Cont.

| Application          | Article               | Core Methodology                                                                                  | Data Type                                            | Supervision Type                         | DL-Based                            |
|----------------------|-----------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------|------------------------------------------|-------------------------------------|
| Boundary Extraction  | [247,248]             | AC                                                                                                | HSI: [247]<br>SAR: [248]<br>LiDAR: [247]             | Supervised: [247]<br>Unsupervised: [248] | No                                  |
|                      | [249,251,254,266,268] | Multiple:<br>Clustering/AC: [249]<br>CNN/AC: [251]<br>Clustering/MRF: [254]<br>CNN/RNN: [266,268] | MSI: [251,266,268]<br>SAR: [249,254]<br>LiDAR: [251] | Supervised: [251]<br>Unsupervised: [249] | No: [249,254]<br>Yes: [251,266,268] |
| Change Detection     | [253]                 | AC                                                                                                | MSI                                                  | Unsupervised                             | No                                  |
|                      | [255,257,258]         | Clustering                                                                                        | MSI: [263]<br>SAR: [255,257]                         | Unsupervised                             | No                                  |
|                      | [252]                 | SVM                                                                                               | MSI                                                  | Semisupervised                           | No                                  |
|                      | [260–262]             | LR                                                                                                | MSI: [260,261]<br>SAR: [262]                         | Supervised                               | No                                  |
|                      | [263–265]             | MLP                                                                                               | MSI                                                  | Supervised                               | No                                  |
|                      | [267]                 | RNN                                                                                               | MSI                                                  | Supervised                               | Yes                                 |
| Image Pre-processing | [269,270,273]         | CNN                                                                                               | HSI: [269]<br>SAR: [273]<br>Other: [270]             | Supervised                               | Yes                                 |
|                      | [275,276]             | GAN                                                                                               | MSI: [275,276]                                       | Supervised                               | Yes                                 |

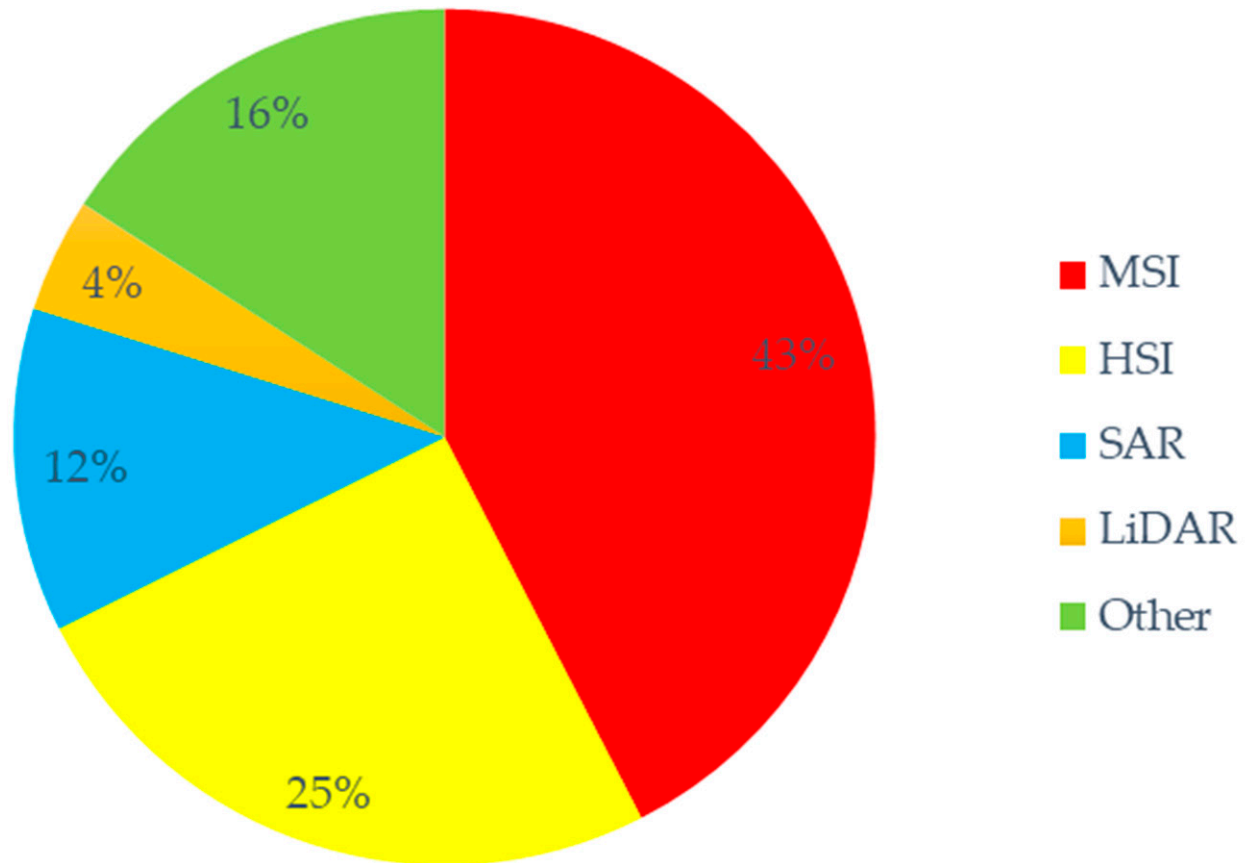
Apart from issues related with the methodologies applied, there are issues with respect to the quality of geoscience data, including noise, missing parts and the lack of precise boundaries defining geoscience-related objects and processes. An interesting attribute of several geoscience phenomena, which rather than being an issue could aid pattern recognition applications, is the existence of long-range spatial and temporal dependencies.

#### 4.1. Geoscience-Related Imaging Data Availability

Beyond supervised learning, the availability of imaging data is important for properly evaluating each method. Figure 9 illustrates modality representation in the works presented in this survey. It can be observed that multispectral images are the most frequently used, whereas HSIs come second. With respect to the datasets that are regularly employed there is no uniformly adopted benchmarking set. Different works perform benchmarking experiments on different subsets of data and direct comparisons are frequently not feasible. This confusion is extended to dataset naming conventions. Landsat [2] datasets are named by the satellite. The same holds for SPOT [277], ERS [278], RADARSAT [279], IRS [280], WorldView [281], QuickBird [282], Pleiades [283]. Still, the Ottawa and Yellow River datasets comprise RADARSAT images [254]. Other datasets are named by the sensor used, as is the case with AVIRIS [284], ROSIS [125], HYDICE [151] and AISA Eagle [140]. In a similar fashion to RADARSAT-derived datasets, the so-called Indian Pines and Salinas datasets [125] are part of the AVIRIS dataset.

Standardized datasets have appeared in order to promote ways for uniformly adopted benchmarking, facilitate comparisons of competing methods, as well as reproducibility. Recently, benchmark datasets, such as SAT-6 [285], DeepGlobe-2018 [286], EuroSAT [287], BigEarthNet [288], and SEN12MS [289], have been proposed for land cover mapping to meet the demand of DL methods for large sample data [290]. In parallel, several agencies produce free land cover products that are mapped and regularly updated to meet the global demand for land cover data applications. Examples of these products include ESA global 10 m land cover mapping product [291], Esri global 10 m land cover mapping product [292], Tsinghua University FROM-GLC10 land cover product [293], and Aerospace Information Research Institute GlobeLand30 product [294]. There are several benchmarking platforms [11], including the data and algorithm standard evaluation website [295] of the

IEEE Geoscience and Remote Sensing Society (IEEE GRSS), the IEEE GRSS annual data fusion contest [296], and the target detection blind test website of the Rochester Institute of Technology [297].



**Figure 9.** Modality representation in the works presented in this survey.

Despite these efforts, in several geoscience-related applications, there is an inherent difficulty in obtaining large amounts of data. For satellite observations, the number of samples is often limited, in both spatial and temporal dimensions. In some cases, there is also a difficulty in obtaining ground truth labels, often associated with the cost of high-quality measurements of geoscience variables. This cost is high for the analysis of chemical and physical material properties in HSIs, as well as when low-flying airplanes or field-based surveys are involved. For very complex systems, in which the exact state cannot be accurately inferred, ground truth is completely out of reach. In other cases, there are processes and events, such as cyclones, flash floods and heat waves, which although occur rarely, they significantly affect Earth's ecosystem [1].

A promising direction in DL research that addresses data and label scarcity is the generation of synthetic images by means of GANs. Abady et al. [298] proposed two GAN-based methods for image generation. The first method, named modified progressive GAN (proGAN) is used for multispectral satellite image generation. The data used for this task are multispectral Sentinel-2 image patches from SEN12MS [289]. The dataset contains images from all four seasons with different meteorological conditions. The spectrum analysis performed confirms that the generated data manage to retain the relationship between the bands for every terrain. Although the obtained image quality is lower than the quality of the original images, it is considered acceptable. The second method of Abady et al. [298], named no-independent component-for-encoding GAN (NICE-GAN), is used for generating bare land images from vegetation land images, and vice versa. For this, they collected multispectral satellite images of vegetation and bare land from ESA

Copernicus hub [299]. Jiang et al. [300] proposed a GAN-based method, named edge-enhanced GAN (EEGAN), for super resolution reconstruction of satellite images. Their network architecture consists of two subnetworks, named ultra-dense subnetwork (UDSN) and edge-enhancement subnetwork (EESN). The UDSN operates as a feature extractor, whereas the EESN operates as a contour extractor and enhancer.

Other machine learning approaches, which have been adopted in the context of geoscience to address data scarcity, are transfer learning, data augmentation, AL and unsupervised learning. It is tempting to speculate that few-shot learning [301], although not yet adopted, provides an interesting alternative.

Another issue related to data and label scarcity is the ‘curse of dimensionality’ or Hughes phenomenon. In the case of HSIs, spectral dimensionality is equal to the total number of bands and ranges in hundreds [11]. Moreover, in some applications, multiple extra variables (e.g., measurements obtained from multiple layers in the atmosphere or groundwater) are employed, contributing to the increase of dimensionality [1]. When the number of dimensions is linearly increased, the volume of the feature space increases exponentially and hence a large amount of data is required [91]. This requirement is not always satisfied due to the data and label scarcity discussed above, leading to overfitting. This is even more prominent in DL-based methods, whereas the issue is amplified by the increase of the intra-class variance and decrease of the inter-class variance in high resolution images, which leads to a decrease of the separability in the spectral domain, particularly for the spectrally similar classes.

A standard approach to alleviate the effects of Hughes phenomenon is dimensionality reduction via feature extraction. This approach is not modality-specific and aims to transform a high-dimensional feature space into a low-dimensional one, using linear or non-linear projections, such as PCA, locality preserving projection (LPP), projection pursuit, or local discriminant embedding [302]. In the case of HSIs, another option is band selection, which aims to preserve the physical meaning of data by selecting the most relevant and informative spectral bands. Supervised band selection utilizes the class separability of labelled training samples. Unsupervised band selection utilizes ranking or clustering. Semi-supervised band selection employs both labelled and unlabelled samples. Most methods in the latter category are based either on manifold learning or hypergraph models [302].

As a final remark with respect to the availability data, we may note that standard, non-learning-based approaches, as those presented in Section 2.1, often provide an alternative to bypass data scarcity, either as standalone approaches or in the context of hybrid methods that also employ learning-based components.

#### 4.2. Inherent Issues in Geoscience Imaging Data

Beyond the quantity of available samples and labels, there are several modality-specific issues in geoscience imaging data. Multispectral images raise difficulties for many image processing algorithms, due to their high resolution, the requirement for GPU and multiprocessing or parallel processing and the increasing computing complexity when the number of frequency bands increases [303]. Another issue related to multispectral images, as well as to HSIs, is the trade-off between spatial and spectral resolution [304]. The quality of SAR images is degraded by a number of factors [305]. Geometric distortions, system nonlinear effects, range migrations are inherent issues on SAR images. A special mention on SAR speckle noise is necessary. Speckle noise is a type of a multiplicative noise which is generated by the random interference of many elementary reflectors within one resolution cell [306]. The process of elimination of the speckle noise from an image is named image despeckling and in the case of SAR images has attracted intensive research interest [307–311]. LiDAR data are also affected by a number of factors [312], including the high cost, the difficulty in processing huge LiDAR datasets, the inability to penetrate water bodies, as the LiDAR system laser beam is absorbed by the water, as well as the difficulty in classifying ground from non-ground data (for DEM generation).

The complementarity between modalities has been exploited in order to enhance the performance in various tasks. The fusion of multispectral, hyperspectral or SAR imagery with LiDAR data [313–317] as well as various combinations of 2D representations [318–322], have been extensively investigated. Some review papers which summarize the corresponding works are [323–325]. Another notable technique is pansharpening, which is the fusion of multispectral and panchromatic images, aiming to exploit the spectral resolution of the former and the spatial resolution of the latter [326].

Some issues extend beyond specific modalities. Several geoscience datasets, such as those collected by Earth observing satellites, contain images with noise, often attributed to atmospheric or surface interference. Imaging quality is also affected by missing values in cases of temporary sensor failures. Even synthetic data, as those generated with GANs, have uncertainties associated with the approximations used. Image time-series analysis may be affected by alterations in imaging equipment, as is satellite switching or the replacement of a damaged camera [1].

In addition, several geoscience processes are non-stationary in space or time. E.g., geographies, vegetation types, rock formations and climatic conditions are non-stationary in the spatial domain, whereas glaciation, polarity reversals and climate phenomena are non-stationary in the temporal domain [1]. This non-stationarity undermines the generalization capability of learning-based approaches.

An issue specific to land cover mapping applications is the lack of precise boundaries in geoscience objects and processes. Cyclones, atmospheric rivers and ocean eddies generally have amorphous boundaries in space and time. Moreover, the form, structure, and patterns of geoscience objects and processes are very complex. E.g., storms and hurricanes dynamically deform in complex fashion over very short periods of time [1]. This complexity poses an extra challenge for the development of target detection and image segmentation methods.

An aspect of geoscience processes, which rather than being an issue could aid pattern recognition applications, is the existence of long-range spatial and temporal dependencies, such as teleconnections [327], where two distant regions show strongly coupled activity in climate variables such as temperature or pressure. Geoscience processes also demonstrate long-term memory in time. Related examples include the effect of the El Nino southern oscillation (ENSO) and Atlantic multidecadal oscillation (AMO) on global floods, droughts, and forest fires [1]. These spatial and temporal dependencies could be taken into account to constrain the actual solution space of a prediction model.

## 5. Conclusions

This work provides an overview of computational methods aiding geoscientists in the analysis of 2D or 3D imaging data, including HSIs, SAR images, point clouds or DEMs. Naturally, advances in such methods follow the progress in computer vision and pattern recognition. This progress is defined by classical computer vision and pattern recognition approaches, such as active contours, superpixels, MRFs, descriptor-guided classification with RFs, DTs, SVMs or MLPs, as well as by DL approaches: CNNs and CNN-based architectures, deep RNNs and the generation of synthetic images with GANs. A critical analysis of related research leads to the following main conclusions:

- There are several widely adopted geoscience datasets. Still, most works involve model training or benchmarking with ad hoc subsets of these datasets. Aiming to cope with this issue, there is a number of organized efforts towards the standardization of geoscience datasets, including Microsoft's effort for AI on Earth, Google Earth and various benchmarking platforms [11].
- There are inherent difficulties in obtaining labelled geoscience data. Satellite observations are often limited, in both spatial and temporal dimensions, whereas ground truth labelling is often associated with the cost of high-quality measurements. For very complex systems, in which the exact state cannot be accurately inferred, ground truth is completely out of reach. In other cases, there are processes and events that



occur rarely. This labelled data paucity can be partially addressed with standard data augmentation approaches, as well as with synthetic data generation by means of GANs. These neural network architectures provide a promising tool for synthetic data generation. In addition, some machine learning approaches, such as active learning or few shot learning, aid model training in cases of limited availability of labelled data. Still, active learning has not been widely adopted, whereas few shot learning has not been applied at all in geoscience.

- Another issue related to data and label scarcity is the ‘curse of dimensionality’ or Hughes phenomenon, which ultimately leads to overfitting. This issue is even more intense in the case of HSIs, due to the large number of correlated spectral bands. A standard remedy for Hughes phenomenon is dimensionality reduction via feature extraction.
- DL-based methods, especially CNNs and CNN-based architectures, are prevalent in recent developments. Still, the successful application of such approaches depends on labelled data availability. Standard pre-DL computer vision approaches remain relevant and often provide an alternative to bypass the aforementioned difficulties in data labelling, either as standalone approaches or in the context of hybrid methods that also employ learning-based components.

Machine learning methods, including DL-based ones, depend on hyperparameters and hyperparameter optimisation (HPO) is crucial for successful RS applications. Recently, Yang and Shami [328] surveyed a diverse range of HPO methods, starting from the adjustment of the  $k$  parameter in the  $k$ -NN classifier, up to HPO of DL models. HPO in DL is also the focus of other recent surveys [329,330]. A challenge of machine learning that has recently attracted intensive research is explainability. Deep neural networks may perform well in various tasks, yet very often in a ‘black-box’ fashion, hindering the understanding of their decisions and concealing any bias and other shortcomings in model performance and datasets. There is a need for explainable machine learning methods in RS multi-label classification tasks towards producing human-interpretable explanations and improve transparency. Recently, Kakogeorgiou and Karantzalos [331] showed that Occlusion [332], Grad-CAM [333] and Lime [334] were interpretable and reliable in RS-related tasks, providing valuable insights for the decisions and performance of DL-based methods, as well for the composition and shortcomings of benchmark datasets. However, none of these approaches delivers high-resolution outputs, whereas both Lime and Occlusion are computationally expensive [331]. Abdollahi and Pradhan [335] proposed an explainable method for Urban vegetation mapping from aerial imagery, whereas Temenos et al. [336] obtained novel insights in spatial epidemiology utilizing explainable AI and RS. A recent survey of Gevaert [337] reviews explainable AI for earth observation, including societal and regulatory perspectives.

Other recent geoscience-related applications focus on the use of point clouds. Vasilakis and Konsolaki [338] combined point cloud data covering an entire cave, which were acquired by a handheld laser scanner, with UAV-acquired data covering the open-air surface above the cave. The absolute and exact placement of the point cloud within a geographic reference frame allow 3D measurements, detailed visualization and quantitative analysis of the subsurface structures. DL-based methods for point cloud analysis, such as PointNet [339,340], could advance this field, as is the case with the work of Ding et al. [341], which employed PointNet for high spatial resolution land cover mapping.

Multimodal approaches provide another promising direction. Hong et al. [342] proposed a multimodal deep learning method for remote sensing (MDL-RS), which combines two subnetworks: Ex-Net and Fu-Net. Ex-Net operates on different modalities as a feature extractor and Fu-Net undertakes the fusion task. MDL-RS has been applied on HSI/LiDAR, as well as on multispectral/SAR data. Audebert et al. [343] investigated the use of CNNs for semantic labelling of multimodal and multiscale VHR urban RS images. At first, they considered an architecture that is based on SegNet [344] and accounts of large spatial

context and high-resolution data. They also used FuseNet [345] for early and late fusion of multispectral and LiDAR data.

Finally, several works addressing the computational cost of deep architectures are important in the context of RS applications [346]. Local connectivity and weight sharing, can reduce the number of parameters and increase processing speed [347]. Mathieu et al. [348] speedup CNN training and testing by employing the fast Fourier transform (FFT) in convolution operations. Jaderberg et al. [349] speedup CNN testing by decomposing layers.

**Author Contributions:** Conceptualization, M.A.S., C.N.V. and T.K.B.; Data curation, M.A.S., C.N.V. and T.K.B.; Formal analysis, M.A.S., C.N.V. and T.K.B.; Writing—review & editing, M.A.S., C.N.V. and T.K.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Tran. Knowl. Dat. Eng.* **2019**, *31*, 1544–1554. [CrossRef]
2. NASA; USGS. Landsat Data Archive. Available online: <https://landsat.gsfc.nasa.gov/data/> (accessed on 24 June 2022).
3. Jafarbiglu, H.; Pourreza, A. A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. *Comput. Electron. Agric.* **2022**, *197*, 106844. [CrossRef]
4. Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Madrigal, V.P.; Mallinis, G.; Ben Dor, E.; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the Use of Unmanned Aerial Systems for Environmental Monitoring. *Remote Sens.* **2018**, *10*, 641. [CrossRef]
5. Peckham, S.D. The CSDMS standard names: Cross-domain naming conventions for describing process models, data sets and their associated variables. In Proceedings of the International Congress on Environmental Modelling and Software, San Diego, CA, USA, 15–19 June 2014.
6. Microsoft, AI for Earth. Available online: <https://www.microsoft.com/en-us/ai/ai-for-earth> (accessed on 24 June 2022).
7. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Env.* **2017**, *202*, 18–27. [CrossRef]
8. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, near real-time global 10 m land use land cover mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
9. O'Connor, J.; Smith, M.J.; James, M.R. Cameras and settings for aerial surveys in the geosciences. *Prog. Phys. Geogr. Earth Environ.* **2017**, *41*, 325–344. [CrossRef]
10. Eismann, M.T. *Hyperspectral Remote Sensing*; SPIE Press: Bellingham, WA, USA, 2012.
11. Gewali, U.B.; Monteiro, S.T.; Saber, E. Machine learning based hyperspectral image analysis: A survey. *arXiv* **2018**, arXiv:1802.08701.
12. Yan, W.Y.; Shaker, A.; El-Ashmawy, N. Urban land cover classification using airborne LiDAR data: A review. *Remote Sens. Environ.* **2015**, *158*, 295–310. [CrossRef]
13. Li, W.; Chen, C.; Su, H.; Du, Q. Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3681–3693. [CrossRef]
14. Hsu, P.-H. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 78–92. [CrossRef]
15. Dalla Mura, M.; Villa, A.; Benediktsson, J.A.; Chanussot, J.; Bruzzone, L. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geosci. Remote Sens. Lett.* **2010**, *8*, 542–546. [CrossRef]
16. Azar, S.G.; Meshgini, S.; Rezaii, T.Y.; Beheshti, S. Hyperspectral image classification based on sparse modeling of spectral blocks. *Neurocomputing* **2020**, *407*, 12–23. [CrossRef]
17. Johnson, A.; Hebert, M. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1999**, *21*, 433–449. [CrossRef]
18. Rusu, R.B.; Blodow, N.; Marton, Z.C.; Beetz, M. Aligning point cloud views using persistent feature histograms. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 3384–3391.
19. Tombari, F.; Salti, S.; Stefano, L.D. Unique signatures of histograms for local surface description. In Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; pp. 356–369. [CrossRef]
20. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [CrossRef]

21. Csurka, G.; Dance, C.; Fan, L.; Willamowski, J.; Bray, C. Visual categorization with bags of keypoints. In Proceedings of the European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 1–22.
22. Hu, F.; Xia, G.-S.; Wang, Z.; Huang, X.; Zhang, L.; Sun, H. Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2015–2030. [[CrossRef](#)]
23. van Gemert, J.C.; Veenman, C.J.; Smeulders, A.W.; Geusebroek, J.-M. Visual Word Ambiguity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1271–1283. [[CrossRef](#)]
24. Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vis.* **1988**, *1*, 321–331. [[CrossRef](#)]
25. Blake, A.; Kohli, P.; Rother, C. *Markov Random Fields for Vision and Image Processing*; The MIT Press: Cambridge, MA, USA, 2011.
26. Stutz, D.; Hermans, A.; Leibe, B. Superpixels: An evaluation of the state-of-the-art. *Comput. Vis. Image Underst.* **2018**, *166*, 1–27. [[CrossRef](#)]
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–8 December 2012.
28. Le Cun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten digit recognition with a back-propagation network. In Proceedings of the International Conference on Neural Information Processing Systems, Denver, CO, USA, 27–30 November 1989.
29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
31. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
32. Schmidhuber, J. *Network Architectures, Objective Functions, and Chain Rule*; Institut für Informatik, Technische Universität München: Munich, Germany, 1993.
33. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neur. Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
34. Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
35. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
36. Bergen, K.J.; Johnson, P.A.; de Hoop, M.V.; Beroza, G.C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **2019**, *363*, eaau0323. [[CrossRef](#)]
37. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]
38. Ioannidou, A.; Chatzilari, E.; Nikolopoulos, S.; Kompatsiaris, I. Deep learning advances in computer vision with 3D data: A survey. *ACM Comp. Surv.* **2018**, *50*, 1–38. [[CrossRef](#)]
39. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [[CrossRef](#)]
40. Zhang, L.; Xia, G.-S.; Wu, T.; Lin, L.; Tai, X.-C. Deep Learning for Remote Sensing Image Understanding. *J. Sens.* **2016**, *2016*, 7954154. [[CrossRef](#)]
41. Beroza, G.C.; Segou, M.; Mousavi, S.M. Machine learning and earthquake forecasting—Next steps. *Nat. Commun.* **2021**, *12*, 4761. [[CrossRef](#)]
42. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
43. Ojala, T.; Pietikainen, M.; Maenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
44. Shen, L.; Jia, S. Three-Dimensional Gabor Wavelets for Pixel-Based Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 5039–5046. [[CrossRef](#)]
45. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comp. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
46. Frome, A.; Huber, D.; Kolluri, R.; Bülow, T.; Malik, J. Recognizing Objects in Range Data Using Regional Point Descriptors. In Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 224–237.
47. Rusu, R.B.; Bradski, G.; Thibaux, R.; Hsu, J. Fast 3D recognition and pose using the viewpoint feature histogram. In Proceedings of the IEEE/RSSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014.
48. Sivic, J.; Zisserman, A. Video Google: A text retrieval approach to object matching in videos. In Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 14–18 October 2004.

49. Jegou, H.; Perronnin, F.; Douze, M.; Sanchez, J.; Perez, P.; Schmid, C. Aggregating Local Image Descriptors into Compact Codes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 1704–1716. [[CrossRef](#)]
50. Perronnin, F.; Liu, Y.; Sanchez, J.; Poirier, H. Large-scale image retrieval with compressed Fisher vectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010.
51. Wu, Q.; An, J. An Active Contour Model Based on Texture Distribution for Extracting Inhomogeneous Insulators From Aerial Images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3613–3626. [[CrossRef](#)]
52. Osher, S.; Sethian, J.A. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formula-tions. *J. Comp. Phys.* **1988**, *79*, 12–49. [[CrossRef](#)]
53. Chan, T.F.; Vese, L.A. Active contours without edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277. [[CrossRef](#)]
54. Tarabalka, Y.; Fauvel, M.; Chanussot, J.; Benediktsson, J.A. SVM- and MRF-based method for accurate classification of hyperspec-tral images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 736–740. [[CrossRef](#)]
55. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral Image Classification via Multitask Joint Sparse Representation and Stepwise MRF Optimization. *IEEE Trans. Cybern.* **2016**, *46*, 2966–2977. [[CrossRef](#)]
56. Solberg, A.; Taxt, T.; Jain, A. A Markov random field model for classification of multisource satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 100–113. [[CrossRef](#)]
57. Wang, C.; Komodakis, N.; Paragios, N. Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey. *Comput. Vis. Image Underst.* **2013**, *117*, 1610–1627. [[CrossRef](#)]
58. Kolmogorov, V.; Zabini, R. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 147–159. [[CrossRef](#)]
59. Golipour, M.; Ghassemian, H.; Mirzapour, F. Integrating Hierarchical Segmentation Maps with MRF Prior for Classification of Hyperspectral Images in a Bayesian Framework. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 805–816. [[CrossRef](#)]
60. Moser, G.; Serpico, S.B.; Benediktsson, J.A. Land-Cover Mapping by Markov Modeling of Spatial-Contextual Information in Very-High-Resolution Remote Sensing Images. *Proc. IEEE* **2013**, *101*, 631–651. [[CrossRef](#)]
61. Neubert, P.; Protzel, P. *Superpixel Benchmark and Comparison*; Karlsruher Instituts für Technologie (KIT) Scientific Publishing: Karlsruhe, Germany, 2012.
62. Csillik, O. Fast Segmentation and Classification of Very High Resolution Remote Sensing Data Using SLIC Superpixels. *Remote Sens.* **2017**, *9*, 243. [[CrossRef](#)]
63. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [[CrossRef](#)] [[PubMed](#)]
64. Moore, A.P.; Prince, J.; Warrell, J.; Mohammed, U.; Jones, G. Superpixel lattices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.
65. Levinstein, A.; Stere, A.; Kutulakos, K.N.; Fleet, D.J.; Dickinson, S.J.; Siddiqi, K. TurboPixels: Fast Superpixels Using Geometric Flows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2290–2297. [[CrossRef](#)] [[PubMed](#)]
66. Vedaldi, A.; Soatto, S. Quick shift and kernel methods for mode seeking. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008.
67. Webb, A.R.; Copsey, K.D. *Statistical Pattern Recognition*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2011.
68. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neur. Net.* **2005**, *16*, 645–678. [[CrossRef](#)]
69. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [[CrossRef](#)]
70. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comp. Surv.* **1999**, *31*, 264–323. [[CrossRef](#)]
71. Madhulatha, T.S. An Overview on Clustering Methods. *IOSR J. Eng.* **2012**, *2*, 719–725. [[CrossRef](#)]
72. Fahad, A.; Alshatri, N.; Tari, Z.; Alamri, A.; Khalil, I.; Zomaya, A.Y.; Foufou, S.; Bouras, A. A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Trans. Emerg. Top. Comput.* **2014**, *2*, 267–279. [[CrossRef](#)]
73. Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview. *WIREs Data Min. Knowl. Discov.* **2012**, *2*, 86–97. [[CrossRef](#)]
74. Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview, II. *WIREs Data Min. Knowl. Discov.* **2017**, *7*, e1219. [[CrossRef](#)]
75. Baraldi, A.; Blonda, P. A survey of fuzzy clustering algorithms for pattern recognition. I. *IEEE Trans. Sys. Man Cybern. Part B (Cybern.)* **1999**, *29*, 778–785. [[CrossRef](#)]
76. Chiou, Y.-C.; Lan, L.W. Genetic clustering algorithms. *Eur. J. Oper. Res.* **2001**, *135*, 413–427. [[CrossRef](#)]
77. Yang, M.-S.; Wu, K.-L. Unsupervised possibilistic clustering. *Pattern Recognit.* **2006**, *39*, 5–21. [[CrossRef](#)]
78. Kriegel, H.-P.; Kröger, P.; Sander, J.; Zimek, A. Density-based clustering. *WIREs Dat. Min. Knowl. Disc.* **2011**, *1*, 231–240. [[CrossRef](#)]
79. Vidal, R. Subspace Clustering. *IEEE Sig. Proc. Mag.* **2011**, *28*, 52–68. [[CrossRef](#)]
80. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]



81. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybern.* **1973**, *3*, 32–57. [[CrossRef](#)]
82. Gustafson, D.E.; Kessel, W.C. Fuzzy clustering with a fuzzy covariance matrix. In Proceedings of the IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, San Diego, CA, USA, 10–12 January 1979.
83. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland, Oregon, 2–4 August 1996.
84. Fukunaga, K.; Hostetler, L. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* **1975**, *21*, 32–40. [[CrossRef](#)]
85. Ankerst, M.; Breunig, M.M.; Kriegel, H.-P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. *SIGMOD Rec.* **1999**, *28*, 49–60. [[CrossRef](#)]
86. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: An efficient data clustering method for very large databases. In Proceedings of the 1996 ACM SIGMOD international conference on Management of Data—SIGMOD'96, Montreal, QC, Canada, 4–6 June 1996.
87. Vidal, R.; Ma, Y.; Sastry, S. Generalized principal component analysis (GPCA). *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1945–1959. [[CrossRef](#)]
88. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
89. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
90. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [[CrossRef](#)]
91. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 4th ed.; Academic Press, Inc.: Cambridge, MA, USA, 2008.
92. Dietterich, T.G.; Bakiri, G. Solving multi-class learning problems via error-correcting output codes. *J. Art. Intell. Res.* **1995**, *2*, 263–286.
93. Theodoridis, S. *Machine Learning, a Bayesian and Optimization Perspective*; Academic Press: New York, NY, USA, 2015.
94. Chong, E.K.P.; Zak, S.H. *An Introduction to Optimization*; Wiley: New York, NY, USA, 2001.
95. Hassoun, M.H.; Intrator, N.; McKay, S.; Christian, W. Fundamentals of Artificial Neural Networks. *Comput. Phys.* **1995**, *10*, 137. [[CrossRef](#)]
96. Gurney, K. *An Introduction to Neural Networks*; Taylor & Francis, Inc.: Florence, KY, USA, 1997.
97. Jain, A.; Mao, J.; Mohiuddin, K. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31–44. [[CrossRef](#)]
98. Svozil, D.; Kvasnicka, V.; Pospichal, J. Introduction to multi-layer feed-forward neural networks. *Chemom. Intell. Lab. Syst.* **1997**, *39*, 43–62. [[CrossRef](#)]
99. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2001.
100. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [[CrossRef](#)]
101. McCulloch, W.S.; Pitts, W.H. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
102. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
103. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
104. Chung, J.; Cho, C.G.K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
105. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches. In Proceedings of the Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014.
106. Graves, A.; Mohamed, A.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013.
107. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2016**, arXiv:1409.0473.
108. Savelonas, M.; Vernikos, I.; Mantzakis, D.; Spyrou, E.; Tsakiri, A.; Karkanis, S. Hybrid Representation of Sensor Data for the Classification of Driving Behaviour. *Appl. Sci.* **2021**, *11*, 8574. [[CrossRef](#)]
109. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [[CrossRef](#)]
110. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *1*. [[CrossRef](#)]
111. Aleissae, A.A.; Kumar, A.; Anwer, R.M.; Khan, S.; Cholakkal, H.; Xia, G.-S.; Khan, F.S. Transformers in remote sensing: A survey. *arXiv* **2022**, arXiv:2209.01206.
112. Metz, L.; Poole, B.; Pfau, D.; Sohl-Dickstein, J. Unrolled generative adversarial networks. *arXiv* **2016**, arXiv:1611.02163.

113. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
114. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
115. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
116. Isola, P.; Zhu, J.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
117. Xia, G.-S.; Liu, G.; Yang, W.; Zhang, L. Meaningful object segmentation from SAR images via a multiscale nonlocal active contour model. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1860–1873. [\[CrossRef\]](#)
118. Li, Z.; Shi, W.; Myint, S.W.; Lu, P.; Wang, Q. Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sens. Env.* **2016**, *175*, 215–230. [\[CrossRef\]](#)
119. Fang, L.; Li, S.; Duan, W.; Ren, J.; Benediktsson, J.A. Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6663–6674. [\[CrossRef\]](#)
120. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–Spatial Classification of Hyperspectral Images with a Superpixel-Based Discriminative Sparse Model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201. [\[CrossRef\]](#)
121. Shi, C.; Pun, C.-M. Superpixel-based 3D deep neural networks for hyperspectral image classification. *Pattern Recognit.* **2018**, *74*, 600–616. [\[CrossRef\]](#)
122. Maulik, U.; Saha, I. Automatic fuzzy clustering using modified differential evolution for image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3503–3510. [\[CrossRef\]](#)
123. Storn, R.; Price, K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **1997**, *11*, 341–359. [\[CrossRef\]](#)
124. Qin, F.; Guo, J.; Lang, F. Superpixel Segmentation for Polarimetric SAR Imagery Using Local Iterative Clustering. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 13–17. [\[CrossRef\]](#)
125. Zhang, H.; Zhai, H.; Zhang, L.; Li, P. Spectral–Spatial Sparse Subspace Clustering for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3672–3684. [\[CrossRef\]](#)
126. Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [\[CrossRef\]](#)
127. Reza, N.; Na, I.S.; Baek, S.W.; Lee, K.-H. Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosyst. Eng.* **2019**, *177*, 109–121. [\[CrossRef\]](#)
128. Jia, S.; Tang, G.; Zhu, J.; Li, Q. A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 88–102. [\[CrossRef\]](#)
129. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496. [\[CrossRef\]](#)
130. Yuan, Y.; Lin, J.; Wang, Q. Dual-Clustering-Based Hyperspectral Band Selection by Contextual Analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [\[CrossRef\]](#)
131. Wang, Q.; Zhang, F.; Li, X. Optimal Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [\[CrossRef\]](#)
132. Zhai, H.; Zhang, H.; Zhang, L.; Li, P. Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1723–1740. [\[CrossRef\]](#)
133. Afonso, M.V.; Bioucas-Dias, J.M.; Figueiredo, M.A.T. An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems. *IEEE Trans. Image Process.* **2011**, *20*, 681–695. [\[CrossRef\]](#)
134. Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 492–501. [\[CrossRef\]](#)
135. Morgan, J.T.; Henneguelle, A.; Ham, J.; Ghosh, J.; Crawford, M.M. Adaptive feature spaces for land cover classification with limited ground truth. *Int. J. Pattern Recognit. Art. Intell.* **2004**, *18*, 777–799. [\[CrossRef\]](#)
136. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [\[CrossRef\]](#)
137. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [\[CrossRef\]](#)
138. Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* **2011**, *115*, 2564–2577. [\[CrossRef\]](#)
139. Eisavi, V.; Homayouni, S.; Yazdi, A.M.; Alimohammadi, A. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environ. Monit. Assess.* **2015**, *187*, 291. [\[CrossRef\]](#) [\[PubMed\]](#)
140. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Random Forests Unsupervised Classification: The Detection and Mapping of *Solanum mauritianum* Infestations in Plantation Forestry Using Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3107–3122. [\[CrossRef\]](#)

141. Scott, G.L.; Longuet-Higgins, H.C. Feature grouping by relocalisation of eigenvectors of proximity matrix. In Proceedings of the British Machine Vision Conference, Oxford, UK, September 1990.
142. Anselin, L. The Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [\[CrossRef\]](#)
143. Sun, L.; Schulz, K. The Improvement of Land Cover Classification by Thermal Remote Sensing. *Remote Sens.* **2015**, *7*, 8368–8390. [\[CrossRef\]](#)
144. Kalantar, B.; Mansor, S.B.; Sameen, M.I.; Pradhan, B.; Shafri, H.Z.M. Drone-based land-cover mapping using a fuzzy unordered rule induction algorithm integrated into object-based image analysis. *Int. J. Remote Sens.* **2017**, *38*, 2535–2556. [\[CrossRef\]](#)
145. Bazi, Y.; Melgani, F. Toward an Optimal SVM Classification System for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3374–3385. [\[CrossRef\]](#)
146. Mantero, P.; Moser, G.; Serpico, S.B. Partially Supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 559–570. [\[CrossRef\]](#)
147. Foody, G.M.; Mathur, A. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sens. Environ.* **2004**, *93*, 107–117. [\[CrossRef\]](#)
148. Foody, G.M.; Mathur, A. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sens. Environ.* **2006**, *103*, 179–189. [\[CrossRef\]](#)
149. Mathur, A.; Foody, G.M. Multiclass and Binary SVM Classification: Implications for Training and Classification Users. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 241–245. [\[CrossRef\]](#)
150. Marconcini, M.; Camps-Valls, G.; Bruzzone, L. A Composite Semisupervised SVM for Classification of Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 234–238. [\[CrossRef\]](#)
151. Huang, X.; Zhang, L. An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 257–272. [\[CrossRef\]](#)
152. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comp. Stat.* **2010**, *2*, 433–459. [\[CrossRef\]](#)
153. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3804–3814. [\[CrossRef\]](#)
154. Chini, M.; Pacifici, F.; Emery, W.J. Morphological operators applied to X-band SAR for urban land use classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.
155. Yoo, H.Y.; Lee, K.; Kwon, B.-D. Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery. *Int. J. Remote Sens.* **2009**, *30*, 6219–6239. [\[CrossRef\]](#)
156. Xu, S.; Fang, T.; Li, D.; Wang, S. Object Classification of Aerial Images with Bag-of-Visual Words. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 366–370. [\[CrossRef\]](#)
157. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J. SVM Active Learning Approach for Image Classification Using Spatial Information. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2217–2233. [\[CrossRef\]](#)
158. Cheng, Q.; Varshney, P.K.; Arora, M.K. Logistic Regression for Feature Selection and Soft Classification of Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 491–494. [\[CrossRef\]](#)
159. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised Hyperspectral Image Segmentation Using Multinomial Logistic Regression with Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4085–4098. [\[CrossRef\]](#)
160. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Series B (Methodol.)* **1977**, *39*, 1–38.
161. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–823. [\[CrossRef\]](#)
162. Bruzzone, L.; Prieto, D.F.; Serpico, S.B. A neural-statistical approach to multitemporal and multisource remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1350–1359. [\[CrossRef\]](#)
163. Hu, X.; Weng, Q. Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sens. Environ.* **2009**, *113*, 2089–2102. [\[CrossRef\]](#)
164. D’Alimonte, D.; Zibordi, G. Phytoplankton determination in an optically complex coastal region using a multilayer perceptron neural network. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2861–2868. [\[CrossRef\]](#)
165. Makantasis, K.; Karantzalos, K.; Doulamis, A.; Doulamis, N. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 4959–4962.
166. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sensors* **2015**, *2015*, 258619. [\[CrossRef\]](#)
167. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 645–657. [\[CrossRef\]](#)
168. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images with Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 881–893. [\[CrossRef\]](#)

169. Zhang, Z.; Wang, H.; Xu, F.; Jin, Y.-Q. Complex-Valued Convolutional Neural Network and Its Application in Polarimetric SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7177–7188. [\[CrossRef\]](#)
170. Scott, G.J.; England, M.R.; Starns, W.A.; Marcum, R.A.; Davis, C.H. Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 549–553. [\[CrossRef\]](#)
171. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014.
172. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 937–949. [\[CrossRef\]](#)
173. Li, E.; Xia, J.; Du, P.; Lin, C.; Samat, A. Integrating Multilayer Features of Convolutional Neural Networks for Remote Sensing Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5653–5665. [\[CrossRef\]](#)
174. Cai, D.; He, X.; Han, J. SRDA: An Efficient Algorithm for Large-Scale Discriminant Analysis. *IEEE Trans. Knowl. Data Eng.* **2007**, *20*, 1–12. [\[CrossRef\]](#)
175. Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [\[CrossRef\]](#)
176. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [\[CrossRef\]](#)
177. Chen, Y.; Zhu, K.; Zhu, L.; He, X.; Ghamisi, P.; Benediktsson, J.A. Automatic Design of Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7048–7066. [\[CrossRef\]](#)
178. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral Image Classification with Convolutional Neural Network and Active Learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4604–4616. [\[CrossRef\]](#)
179. Wu, X.; Hong, D.; Chanussot, J. Convolutional Neural Networks for Multimodal Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5517010. [\[CrossRef\]](#)
180. Mei, S.; Chen, X.; Zhang, Y.; Li, J.; Plaza, A. Accelerating convolutional neural network-based hyperspectral image classification by step activation quantization. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5502012.
181. Rastegari, M.; Ordonez, V.; Redmon, J.; Farhadi, A. XNOR-Net: ImageNet classification using binary convolutional neural networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.
182. Lin, J.; Mou, L.; Zhu, X.X.; Ji, X.; Wang, Z.J. Attention-Aware Pseudo-3-D Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 7790–7802. [\[CrossRef\]](#)
183. Dong, Y.; Liu, Q.; Du, B.; Zhang, L. Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 1559–1572. [\[CrossRef\]](#) [\[PubMed\]](#)
184. Lu, Z.; Liang, S.; Yang, Q.; Du, B. Evolving block-based convolutional neural network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5525921. [\[CrossRef\]](#)
185. Ienco, D.; Gaetano, R.; Dupaquier, C.; Maurel, P. Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1685–1689. [\[CrossRef\]](#)
186. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [\[CrossRef\]](#)
187. Maggiori, E.; Charpiat, G.; Tarabalka, Y.; Alliez, P. Recurrent Neural Networks to Correct Satellite Image Classification Maps. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4962–4971. [\[CrossRef\]](#)
188. Rußwurm, M.; Körner, M. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [\[CrossRef\]](#)
189. Ndikumana, E.; Minh, D.H.T.; Baghdadi, N.; Courault, D.; Hossard, L. Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sens.* **2018**, *10*, 1217. [\[CrossRef\]](#)
190. Ho Tong Minh, D.; Lalande, N.; Ndikumana, E.; Osman, F.; Maurel, P. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR Sentinel-1. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 464–468. [\[CrossRef\]](#)
191. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [\[CrossRef\]](#)
192. Lee, H.; Slatton, K.C.; Roth, B.E.; Cropper, W.P., Jr. Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *Int. J. Remote Sens.* **2010**, *31*, 117–139. [\[CrossRef\]](#)
193. Beucher, S.; Lantuéjoul, C. Use of watersheds in contour detection. In Proceedings of the International Workshop on Image Processing: Real-Time Edge and Motion Detection/Estimation, Rennes, France, 17–21 September 1979.
194. Kim, Y.; Ling, H. Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1328–1337. [\[CrossRef\]](#)
195. Kim, Y.J.; Nam, B.H.; Youn, H. Sinkhole detection and characterization using LiDAR-derived DEM with logistic regression. *Remote Sens.* **2019**, *11*, 1592. [\[CrossRef\]](#)



196. Martorella, M.; Giusti, E.; Capria, A.; Berizzi, F.; Bates, B. Automatic Target Recognition by Means of Polarimetric ISAR Images and Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3786–3794. [\[CrossRef\]](#)
197. Cameron, W.L.; Rais, H. Conservative polarimetric scatterers and their role in incorrect extensions of the Cameron decomposition. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3506–3516. [\[CrossRef\]](#)
198. Taravat, A.; Proud, S.; Peronaci, S.; Del Frate, F.; Oppelt, N. Multilayer Perceptron Neural Networks Model for Meteosat Second Generation SEVIRI Daytime Cloud Masking. *Remote Sens.* **2015**, *7*, 1529–1539. [\[CrossRef\]](#)
199. Chen, X.; Xiang, S.; Liu, C.L.; Pan, C.H. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1797–1801. [\[CrossRef\]](#)
200. Cheng, G.; Zhou, P.; Han, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [\[CrossRef\]](#)
201. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional Neural Network with Data Augmentation for SAR Target Recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [\[CrossRef\]](#)
202. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [\[CrossRef\]](#)
203. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [\[CrossRef\]](#)
204. Cheng, G.; Wang, Y.; Xu, S.; Wang, H.; Xiang, S.; Pan, C. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3322–3337. [\[CrossRef\]](#)
205. Shao, Z.; Pan, Y.; Diao, C.; Cai, J. Cloud Detection in Remote Sensing Images Based on Multiscale Features-Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4062–4076. [\[CrossRef\]](#)
206. Hsieh, M.R.; Lin, Y.L.; Hsu, W.H. Drone-based object counting by spatially regularized regional proposal network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
207. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015.
208. Kellenberger, B.; Marcos, D.; Tuia, D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sens. Environ.* **2018**, *216*, 139–153. [\[CrossRef\]](#)
209. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 14–18 June 2009.
210. Shrivastava, A.; Gupta, A.; Girshick, R. Training region-based object detectors with online hard example mining. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
211. Zhang, H.; Wang, G.; Lei, Z.; Hwang, J.N. Eye in the sky: Drone-based object tracking and 3D localization. In Proceedings of the ACM International Conference on Multimedia, Nice, France, 21–25 October 2019.
212. Wang, G.; Wang, Y.; Zhang, J.N.; Gu, R.; Hwang, J.N. Exploit the connectivity: Multi-object tracking with TrackletNet. In Proceedings of the ACM International Conference on Multimedia, Nice, France, 21–25 October 2019.
213. Seitz, S.M.; Curless, B.; Diebel, J.; Scharstein, D.; Szeliski, R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006.
214. Zhang, C.; Zuo, R.; Xiong, Y. Detection of the multivariate geochemical anomalies associated with mineralization using a deep convolutional neural network and a pixel-pair feature method. *Appl. Geochem.* **2021**, *130*, 104994. [\[CrossRef\]](#)
215. Li, W.; Wu, G.; Zhang, F.; Du, Q. Hyperspectral Image Classification Using Deep Pixel-Pair Features. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 844–853. [\[CrossRef\]](#)
216. Xu, D.; Wu, Y. MRFF-YOLO: A Multi-Receptive Fields Fusion Network for Remote Sensing Target Detection. *Remote Sens.* **2020**, *12*, 3118. [\[CrossRef\]](#)
217. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
218. Xu, D.; Wu, Y. FE-YOLO: A Feature Enhancement Network for Remote Sensing Target Detection. *Remote Sens.* **2021**, *13*, 1311. [\[CrossRef\]](#)
219. Qing, Y.; Liu, W.; Feng, L.; Gao, W. Improved YOLO Network for Free-Angle Remote Sensing Target Detection. *Remote Sens.* **2021**, *13*, 2171. [\[CrossRef\]](#)
220. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. RepVGG: Making VGG-style convnets great again. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021.
221. Wang, C.; Wang, Q.; Wu, H.; Zhao, C.; Teng, G.; Li, J. Low-Altitude Remote Sensing Opium Poppy Image Detection Based on Modified YOLOv3. *Remote Sens.* **2021**, *13*, 2130. [\[CrossRef\]](#)
222. Xie, S.; Girshick, R.; Dollár, P. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

223. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
224. Zakria, Z.; Deng, J.; Kumar, R.; Khokhar, M.S.; Cai, J.; Kumar, J. Multiscale and Direction Target Detecting in Remote Sensing Images via Modified YOLO-v4. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1039–1048. [[CrossRef](#)]
225. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
226. Ke, X.; Zhang, X.; Zhang, T. GCBA-Net: A Global Context Boundary-Aware Network for SAR Ship Instance Segmentation. *Remote Sens.* **2022**, *14*, 2165. [[CrossRef](#)]
227. Li, Q.; Chen, Y.; Zeng, Y. Transformer with Transfer CNN for Remote-Sensing-Image Object Detection. *Remote Sens.* **2022**, *14*, 984. [[CrossRef](#)]
228. Xiao, X.; Guo, W.; Chen, R.; Hui, Y.; Wang, J.; Zhao, H. A Swin Transformer-Based Encoding Booster Integrated in U-Shaped Network for Building Extraction. *Remote Sens.* **2022**, *14*, 2611. [[CrossRef](#)]
229. Chen, X.; Qiu, C.; Guo, W.; Yu, A.; Tong, X.; Schmitt, M. Multiscale Feature Learning by Transformer for Building Extraction From Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 2503605. [[CrossRef](#)]
230. Zhu, Q.; Liao, C.; Hu, H.; Mei, X.; Li, H. MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction From Remote Sensed Imagery. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6169–6181. [[CrossRef](#)]
231. Joseph, M.; Wang, L.; Wang, F. Using Landsat Imagery and Census Data for Urban Population Density Modeling in Port-au-Prince, Haiti. *GIScience Remote Sens.* **2012**, *49*, 228–250. [[CrossRef](#)]
232. Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; de Jesus, J.M.; Tamene, L.; et al. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE* **2015**, *10*, e0125814. [[CrossRef](#)]
233. Odeh, I.O.A.; McBratney, A.B.; Chittleborough, D.J. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma* **1995**, *67*, 215–226. [[CrossRef](#)]
234. Hengl, T.; Heuvelink, G.B.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315. [[CrossRef](#)]
235. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* **2015**, *10*, e0107042. [[CrossRef](#)] [[PubMed](#)]
236. Georganos, S.; Grippa, T.; Gadiaga, A.N.; Linard, C.; Lennert, M.; VanHuyse, S.; Mboga, N.; Wolff, E.; Kalogirou, S. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* **2021**, *36*, 121–136. [[CrossRef](#)]
237. Sun, D.; Li, Y.; Wang, Q. A Unified Model for Remotely Estimating Chlorophyll a in Lake Taihu, China, Based on SVM and In Situ Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2957–2965. [[CrossRef](#)]
238. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Env.* **1999**, *67*, 267–287. [[CrossRef](#)]
239. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res. Solid Earth* **1984**, *89*, 6329–6340. [[CrossRef](#)]
240. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [[CrossRef](#)]
241. Dardel, C.; Kergoat, L.; Hiernaux, P.; Mougin, E.; Grippa, M.; Tucker, C.J. Re-greening Sahel: 30years of remote sensing data and field observations (Mali, Niger). *Remote Sens. Environ.* **2014**, *140*, 350–364. [[CrossRef](#)]
242. Du, M.; Wang, L.; Zou, S.; Shi, C. Modeling the Census Tract Level Housing Vacancy Rate with the Jilin1-03 Satellite and Other Geospatial Data. *Remote Sens.* **2018**, *10*, 1920. [[CrossRef](#)]
243. Tien Bui, D.; Khosravi, K.; Shahabi, H.; Daggupati, P.; Adamowski, J.F.; Melesse, A.M.; Pham, B.T.; Pourghasemi, H.R.; Mahmoudi, M.; Bahrami, S.; et al. Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sens.* **2019**, *11*, 1589. [[CrossRef](#)]
244. Corsini, G.; Diani, M.; Grasso, R.; De Martino, M.; Mantero, P.; Serpico, S. Radial Basis Function and Multilayer Perceptron neural networks for sea water optically active parameter estimation in case II waters: A comparison. *Int. J. Remote Sens.* **2003**, *24*, 3917–3931. [[CrossRef](#)]
245. Ozturk, D. Urban Growth Simulation of Atakum (Samsun, Turkey) Using Cellular Automata-Markov Chain and Multi-Layer Perceptron-Markov Chain Models. *Remote Sens.* **2015**, *7*, 5918–5950. [[CrossRef](#)]
246. Al-Najjar, H.A.; Pradhan, B. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. *Geosci. Front.* **2021**, *12*, 625–637. [[CrossRef](#)]
247. Sukcharoenpong, A.; Yilmaz, A.; Li, R. An Integrated Active Contour Approach to Shoreline Mapping Using HSI and DEM. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1586–1597. [[CrossRef](#)]
248. Liu, C.; Xiao, Y.; Yang, J. A Coastline Detection Method in Polarimetric SAR Images Mixing the Region-Based and Edge-Based Active Contour Models. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3735–3747. [[CrossRef](#)]

249. Modava, M.; Akbarizadeh, G. Coastline extraction from SAR images using spatial fuzzy clustering and the active contour method. *Int. J. Remote Sens.* **2017**, *38*, 355–370. [\[CrossRef\]](#)
250. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
251. Sun, Y.; Zhang, X.; Zhao, X.; Xin, Q. Extracting Building Boundaries from High Resolution Optical Images and LiDAR Data by Integrating the Convolutional Neural Network and the Active Contour Model. *Remote Sens.* **2018**, *10*, 1459. [\[CrossRef\]](#)
252. Bovolo, F.; Bruzzone, L.; Marconcini, M. A Novel Approach to Unsupervised Change Detection Based on a Semisupervised SVM and a Similarity Measure. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2070–2082. [\[CrossRef\]](#)
253. Bazi, Y.; Melgani, F.; Al-Sharari, H.D. Unsupervised Change Detection in Multispectral Remotely Sensed Imagery with Level Set Methods. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3178–3187. [\[CrossRef\]](#)
254. Gong, M.; Su, L.; Jia, M.; Chen, W. Fuzzy Clustering with a Modified MRF Energy Function for Change Detection in Synthetic Aperture Radar Images. *IEEE Trans. Fuzzy Syst.* **2014**, *22*, 98–109. [\[CrossRef\]](#)
255. Zheng, Y.; Zhang, X.; Hou, B.; Liu, G. Using combined difference image and k-means clustering for SAR image change detection. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 691–695. [\[CrossRef\]](#)
256. Deledalle, C.-A.; Denis, L.; Tupin, F. Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights. *IEEE Trans. Image Process.* **2009**, *18*, 2661–2672. [\[CrossRef\]](#)
257. Ghosh, A.; Mishra, N.S.; Ghosh, S. Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Inf. Sci.* **2011**, *181*, 699–715. [\[CrossRef\]](#)
258. Leichtle, T.; Geiß, C.; Wurm, M.; Lakes, T.; Taubenböck, H. Unsupervised change detection in VHR remote sensing imagery—An object-based clustering approach in a dynamic urban environment. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 15–27. [\[CrossRef\]](#)
259. Singh, A. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [\[CrossRef\]](#)
260. Khurshid, H.; Khan, M.F. Segmentation and Classification Using Logistic Regression in Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 224–232. [\[CrossRef\]](#)
261. Tan, K.; Jin, X.; Plaza, A.; Wang, X.; Xiao, L.; Du, P. Automatic Change Detection in High-Resolution Remote Sensing Images by Using a Multiple Classifier System and Spectral–Spatial Features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3439–3451. [\[CrossRef\]](#)
262. Molin, R.D.; Rosa, R.A.S.; Bayer, F.M.; Pettersson, M.I.; Machado, R. A change detection algorithm for SAR images based on logistic regression. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
263. Pacifici, F.; Del Frate, F. Automatic Change Detection in Very High Resolution Images with Pulse-Coupled Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2009**, *7*, 58–62. [\[CrossRef\]](#)
264. Salmon, B.P.; Olivier, J.C.; Kleynhans, W.; Wessels, K.J.; Van den Bergh, F.; Steenkamp, K.C. The use of a multilayer perceptron for detecting new human settlements from a time series of MODIS images. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 873–883. [\[CrossRef\]](#)
265. Roy, M.; Routaray, D.; Ghosh, S.; Ghosh, A. Ensemble of Multilayer Perceptrons for Change Detection in Remotely Sensed Images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 49–53. [\[CrossRef\]](#)
266. Zhao, B.; Zhong, Y.; Xia, G.-S.; Zhang, L. Dirichlet-Derived Multiple Topic Scene Classification Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2108–2123. [\[CrossRef\]](#)
267. Lyu, H.; Lu, H.; Mou, L. Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection. *Remote Sens.* **2016**, *8*, 506. [\[CrossRef\]](#)
268. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning Spectral–Spatial–Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 924–935. [\[CrossRef\]](#)
269. Yuan, Q.; Zhang, Q.; Li, J.; Shen, H.; Zhang, L. Hyperspectral Image Denoising Employing a Spatial–Spectral Deep Residual Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1205–1218. [\[CrossRef\]](#)
270. Li, T.; Zuo, R.; Xiong, Y.; Peng, Y. Random-drop data augmentation of deep convolutional neural network for mineral prospectivity mapping. *Nat. Res. Res.* **2021**, *30*, 27–38. [\[CrossRef\]](#)
271. Zuo, R.; Wang, Z. Effects of Random Negative Training Samples on Mineral Prospectivity Mapping. *Nat. Resour. Res.* **2020**, *29*, 3443–3455. [\[CrossRef\]](#)
272. Nykänen, V.; Lahti, I.; Niiranen, T.; Korhonen, K. Receiver operating characteristics (ROC) as validation tool for prospectivity models—A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geol. Rev.* **2015**, *71*, 853–860. [\[CrossRef\]](#)
273. Molini, A.B.; Valsesia, D.; Fracastoro, G.; Magli, E. Speckle2Void: Deep Self-Supervised SAR Despeckling with Blind-Spot Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–17. [\[CrossRef\]](#)
274. Laine, S.; Karras, T.; Lehtinen, J.; Aila, T. High-quality self-supervised deep image denoising. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.

275. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSAN: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpener. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 10227–10242. [CrossRef]
276. Pan, H. Cloud removal for remote sensing imagery via spatial attention generative adversarial network. *arXiv* **2020**, arXiv:2009.13015.
277. Spot. Available online: <https://earth.esa.int/eogateway/missions/spot> (accessed on 27 September 2022).
278. ERS. Available online: <https://earth.esa.int/eogateway/missions/ers> (accessed on 27 September 2022).
279. RADARSAT. Available online: <https://earth.esa.int/eogateway/missions/radarsat> (accessed on 27 September 2022).
280. IRS. Available online: <https://earth.esa.int/eogateway/missions/irs-1d> (accessed on 27 September 2022).
281. WorldView. Available online: <https://earth.esa.int/eogateway/missions/worldview-3> (accessed on 27 September 2022).
282. QuickBird. Available online: <https://earth.esa.int/eogateway/catalog/quickbird-full-archive> (accessed on 27 September 2022).
283. Pleiades. Available online: <https://earth.esa.int/eogateway/catalog/pleiades-esa-archive> (accessed on 27 September 2022).
284. AVIRIS. Available online: [https://aviris.jpl.nasa.gov/data/free\\_data.html](https://aviris.jpl.nasa.gov/data/free_data.html) (accessed on 27 September 2022).
285. Basu, S.; Ganguly, S.; Mukhopadhyay, S.; DiBiano, R.; Karki, M.; Nemani, R. DeepSat: A learning framework for satellite imagery. In Proceedings of the SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, DC, USA, 3–6 November 2015.
286. Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A challenge to parse the earth through satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
287. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [CrossRef]
288. Sumbul, G.; Charfuelan, M.; Demir, B.U.M.; Markl, V. Big Earth Net: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.
289. Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion. *arXiv* **2019**, arXiv:1906.07789.
290. Xu, G.; Fang, Y.; Deng, M.; Sun, G.; Chen, J. Remote Sensing Mapping of Build-Up Land with Noisy Label via Fault-Tolerant Learning. *Remote Sens.* **2022**, *14*, 2263. [CrossRef]
291. ESA World Cover 10 m 2020 v100. Available online: <https://doi.org/10.5281/zenodo.5571936> (accessed on 27 September 2022).
292. Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global land use/land cover with Sentinel 2 and deep learning. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021.
293. Gong, P.; Liu, H.; Zhang, M.; Li, C.; Wang, J.; Huang, H.; Clinton, N.; Ji, L.; Li, W.; Bai, Y.; et al. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* **2019**, *64*, 370–373. [CrossRef]
294. Jun, C.; Ban, Y.; Li, S. Open access to Earth land-cover map. *Nature* **2014**, *514*, 434. [CrossRef]
295. Dell’Acqua, F.; Iannelli, G.C.; Kerekes, J.; Moser, G.; Pierce, L.; Goldoni, E. The IEEE GRSS data and algorithm standard evaluation (DASE) website: Incrementally building a standardized assessment for algorithm performance. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017.
296. IEEE GRSS Data Fusion Contest. Available online: <https://www.grss-ieee.org/community/technical-committees/2022-ieee-grss-data-fusion-contest/> (accessed on 5 September 2022).
297. Target Detection Blind Test. Available online: <http://dirtsapps.cis.rit.edu/blindtest/> (accessed on 5 September 2022).
298. Abady, L.; Barni, M.; Garzelli, A.; Tondi, B. GAN generation of synthetic multispectral satellite images. In Proceedings of the SPIE 11533, Image and Signal Processing for Remote Sensing XXVI, Online, 21–25 September 2020.
299. Copernicus Open Access Hub. Available online: <https://scihub.copernicus.eu/dhus/#/home> (accessed on 5 September 2022).
300. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-Enhanced GAN for Remote Sensing Image Superresolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5799–5812. [CrossRef]
301. Wang, Y.; Yao, Q.; Kwok, J.; Ni, L.N. Generalizing from a few examples: A survey on few-shot learning. *arXiv* **2019**, arXiv:1904.05046. [CrossRef]
302. Sellami, A.; Ben Abbes, A.; Barra, V.; Farah, I.R. Fused 3-D spectral-spatial deep neural networks and spectral clustering for hyperspectral image classification. *Pattern Recognit. Lett.* **2020**, *138*, 594–600. [CrossRef]
303. Thyagarajan, K.K.; Vignesh, T. Soft Computing Techniques for Land Use and Land Cover Monitoring with Multispectral Remote Sensing Images: A Review. *Arch. Comput. Methods Eng.* **2019**, *26*, 275–301. [CrossRef]
304. Kwan, C. Methods and challenges using multispectral and hyperspectral images for practical change detection applications. *Information* **2019**, *10*, 353. [CrossRef]



305. Singh, P.; Diwakar, M.; Shankar, A.; Shree, R.; Kumar, M. A Review on SAR Image and its Despeckling. *Arch. Comput. Methods Eng.* **2021**, *28*, 4633–4653. [\[CrossRef\]](#)
306. Liu, S.; Wu, G.; Zhang, X.; Zhang, K.; Wang, P.; Li, Y. SAR despeckling via classification-based nonlocal and local sparse representation. *Neurocomputing* **2017**, *219*, 174–185. [\[CrossRef\]](#)
307. Wang, G.; Bo, F.; Chen, X.; Lu, W.; Hu, S.; Fang, J. A collaborative despeckling method for SAR images based on texture classification. *Remote Sens.* **2022**, *14*, 1465. [\[CrossRef\]](#)
308. Choi, H.; Jeong, J. Speckle Noise Reduction Technique for SAR Images Using Statistical Characteristics of Speckle Noise and Discrete Wavelet Transform. *Remote Sens.* **2019**, *11*, 1184. [\[CrossRef\]](#)
309. Dalsasso, E.; Yang, X.; Denis, L.; Tupin, F.; Yang, W. SAR Image Despeckling by Deep Neural Networks: From a Pre-Trained Model to an End-to-End Training Strategy. *Remote Sens.* **2020**, *12*, 2636. [\[CrossRef\]](#)
310. Mullissa, A.G.; Marcos, D.; Tuia, D.; Herold, M.; Reiche, J. DeSpeckNet: Generalizing deep learning-based SAR image despeckling. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5200315. [\[CrossRef\]](#)
311. Zhao, Y.; Liu, J.G.; Zhang, B.; Hong, W.; Wu, Y.-R. Adaptive Total Variation Regularization Based SAR Image Despeckling and Despeckling Evaluation Index. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2765–2774. [\[CrossRef\]](#)
312. Muhadi, N.A.; Abdullah, A.F.; Bejo, S.K.; Mahadi, M.R.; Mijic, A. The use of LiDAR-derived DEM in flood applications: A Review. *Remote Sens.* **2020**, *12*, 2308. [\[CrossRef\]](#)
313. Rasti, B.; Ghamisi, P.; Plaza, J.; Plaza, A. Fusion of hyperspectral and LiDAR data using sparse and low-rank component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6354–6365. [\[CrossRef\]](#)
314. Zhou, L.; Geng, J.; Jiang, W. Joint classification of hyperspectral and LiDAR data based on position-channel cooperative attention network. *Remote Sens.* **2022**, *14*, 3247. [\[CrossRef\]](#)
315. Luo, S.; Wang, C.; Xi, X.; Zeng, H.; Li, D.; Xia, S.; Wang, P. Fusion of airborne discrete-return LiDAR and hyperspectral data for land cover classification. *Remote Sens.* **2016**, *8*, 3. [\[CrossRef\]](#)
316. Millard, K.; Richardson, M. Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Can. J. Remote Sens.* **2013**, *39*, 290–307. [\[CrossRef\]](#)
317. Pourshamsi, M.; Garcia, M.; Laval, M.; Balzter, H. A Machine-Learning Approach to PolInSAR and LiDAR Data Fusion for Improved Tropical Forest Canopy Height Estimation Using NASA AfriSAR Campaign Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3453–3463. [\[CrossRef\]](#)
318. Seo, D.K.; Kim, Y.H.; Eo, Y.D.; Lee, M.H.; Park, W.Y. Fusion of SAR and Multispectral Images Using Random Forest Regression for Change Detection. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 401. [\[CrossRef\]](#)
319. Zhang, H.; Shen, H.; Yuan, Q.; Guan, X. Multispectral and SAR image fusion based on Laplacian pyramid and sparse representation. *Remote Sens.* **2022**, *14*, 870. [\[CrossRef\]](#)
320. Hu, J.; Hong, D.; Wang, Y.; Zhu, X.X. A Comparative Review of Manifold Learning Techniques for Hyperspectral and Polarimetric SAR Image Fusion. *Remote Sens.* **2019**, *11*, 681. [\[CrossRef\]](#)
321. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 639–643. [\[CrossRef\]](#)
322. Sun, W.; Ren, K.; Meng, X.; Xiao, C.; Yang, G.; Peng, J. A Band Divide-and-Conquer Multispectral and Hyperspectral Image Fusion Method. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [\[CrossRef\]](#)
323. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.M.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.M.; Anders, K.; Gloaguen, R.; et al. Multisource and multitemporal data fusion in remote sensing a comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 6–39. [\[CrossRef\]](#)
324. Dalla Mura, M.; Prasad, S.; Pacifici, F.; Gamba, P.; Chanussot, J.; Benediktsson, J.A. Challenges and opportunities of multi-modality and data fusion in remote sensing. *Proc. IEEE* **2015**, *103*, 1585–1601. [\[CrossRef\]](#)
325. Kahraman, S.; Bacher, R. A comprehensive review of hyperspectral data fusion with LiDAR and SAR data. *Ann. Rev. Contr.* **2021**, *51*, 236–253. [\[CrossRef\]](#)
326. Vivone, G.; Alparone, L.; Chanussot, J.; Mura, M.D.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [\[CrossRef\]](#)
327. Kawale, J.; Liess, S.; Kumar, A.; Steinbach, M.; Snyder, P.; Kumar, V.; Ganguly, A.R.; Samatova, N.F.; Semazzi, F. A graph-based approach to find teleconnections in climate data. *Stat. Anal. Data Min. ASA Data Sci. J.* **2013**, *6*, 158–179. [\[CrossRef\]](#)
328. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [\[CrossRef\]](#)
329. Yu, T.; Zhu, H. Hyper-parameter optimization: A review of algorithms and applications. *arXiv* **2020**, arXiv:2003.05689.
330. Hernández, A.M.; Nieuwenhuys, I.V.; Rojas-Gonzalez, S. A survey on multi-objective hyperparameter optimization algorithms for machine learning. *arXiv* **2021**, arXiv:2111.13755.
331. Kakogeorgiou, I.; Karantzalos, K. Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102520. [\[CrossRef\]](#)

332. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference of Computer Vision, Zurich, Switzerland, 5–12 September 2014.
333. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
334. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier; Association for Computing Machinery: New York, NY, USA, 2016.
335. Abdollahi, A.; Pradhan, B. Urban Vegetation Mapping from Aerial Imagery Using Explainable AI (XAI). *Sensors* **2021**, *21*, 4738. [\[CrossRef\]](#)
336. Temenos, A.; Tzortzis, I.N.; Kaselimi, M.; Rallis, I.; Doulamis, A.; Doulamis, N. Novel Insights in Spatial Epidemiology Utilizing Explainable AI (XAI) and Remote Sensing. *Remote Sens.* **2022**, *14*, 3074. [\[CrossRef\]](#)
337. Gevaert, C.M. Explainable AI for earth observation: A review including societal and regulatory perspectives. *Int. J. Appl. Earth Obs. Geoinf. ITC J.* **2022**, *112*, 102869. [\[CrossRef\]](#)
338. Vassilakis, E.; Konsolaki, A. Quantification of cave geomorphological characteristics based on multi source point cloud data interoperability. *Zeitschr. Geomorphol.* **2022**, *63*, 265–277. [\[CrossRef\]](#)
339. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
340. Aoki, Y.; Goforth, H.; Srivatsan, R.A.; Lucey, S. PointNetLK: Robust & efficient point cloud registration using PointNet. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
341. Ding, L.; Cai, Y.; Zhang, J.; Gao, Y.; Wang, J.; Zheng, C.; Lei, L.; Ma, A. PointNet: Learning point representation for high-resolution remote sensing imagery land-cover classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021.
342. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4340–4354. [\[CrossRef\]](#)
343. Audebert, N.; Le Saux, B.; Lefèvre, S. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 20–32. [\[CrossRef\]](#)
344. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [\[CrossRef\]](#)
345. Hazirbas, C.; Ma, L.; Domokos, C.; Cremers, D. FuseNet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016.
346. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
347. Zhang, Q.; Zhang, M.; Chen, T.; Sun, Z.; Ma, Y.; Yu, B. Recent advances in convolutional neural network acceleration. *Neurocomputing* **2019**, *323*, 37–51. [\[CrossRef\]](#)
348. Mathieu, M.; Henaff, M.; Le Cun, Y. Fast training of convolutional networks through FFTs. *arXiv* **2013**, arXiv:1312.5851.
349. Jaderberg, M.; Vedaldi, A.; Zisserman, A. Speeding up convolutional neural networks with low rank expansions. *arXiv* **2014**, arXiv:1405.3866.