Beyond Supervised Learning in Remote Sensing: A Systematic Review of Deep Learning Approaches

Benyamin Hosseiny, Masoud Mahdianpari, *Senior Member, IEEE*, Mohammadali Hemati, *Student Member, IEEE*, Ali Radman, Fariba Mohammadimanesh, *Member, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—An increasing availability of remote sensing data in the era of geo big-data makes producing well-represented, reliable training data to be more challenging and requires an excessive amount of human labor. In addition, the rapid increase in graphics processing unit processing power has enabled the development of advanced deep learning algorithms, which achieve impressive results in the field of satellite image processing. However, they require a huge and comprehensive training dataset to avoid overfitting problems and to represent a generalizable model. Thus, moving toward the development of nonsupervised deep learning (NSDL) models in different remote sensing applications is an inevitable need. To provide an initial response to that need, this article performs a comprehensive review and systematic meta-analysis of recently published research articles focusing on the applications of NSDL for remote sensing data processing. In order to identify future research directions and formulate recommendations, we extract trends and highlight interesting approaches from this large body of literature. Consequently, current challenges, prospects, and recommendations are also discussed to uncover the trend. According to the results, there is a sharp increasing trend in the applicability of NSDL methods during these few years particularly, with the advent of new deep architectures, such as adversarial, graph, and transformer models. As a result, this review article discusses different remote sensing data processing applications and challenges that can be addressed using NSDL approaches.**

***Index Terms*—Self-supervised, semisupervised, training data, transfer learning (TL), unsupervised, weakly supervised.**

I. Introduction

RECENT technological advances in platforms, sensors, and data management infrastructures have made the various, frequent, and high-quality remote sensing datasets more accessible, leading the research focus into the "geo Big Data." Therefore, developing processing tools to effectively use remote sensing dataset have been always one of the hot topics in the research community. On the one hand, the main processing tasks can be divided into classification, object detection, change detection (CD), and image fusion. On the other hand, many remote sensing applications require preprocessing steps, such as noise reduction, data fusion, image registration, and super-resolution (SR). Hence, diverse methods from statistical parametric approaches to machine learning (ML) and recent deep learning (DL) methods have been developed or evaluated to address remote sensing data processing challenges [1].

Despite the significant improvements of DL algorithms on various remote sensing processing tasks, they require extensive labeled training data. In other words, the excellence of DL models is highly dependent on the quality and amount of training samples. However, acquiring high-quality annotations on a large scale and in some applications could be laborious and costly. In particular, for Earth observation (EO) applications, several field works are required to collect proper training samples. In addition, the large volume and variety of remote sensing datasets and the number of input data layers directly increase the requirement of training samples due to the Hughes phenomenon in higher dimensions [2]. Furthermore, some imaging scenes, such as complex and heterogeneous landscapes, could add various challenges and, thus, affecting the generalization capability of the DL model due to their mixed distribution. To sum up, DL models' performance and generalization capability extremely relies on the large-scale training samples.

Despite annotated training data, unlabeled data are more accessible and easier to collect. Accordingly, exploiting unlabeled data in order to avoid expensive training data collection and achieving the highest performance of DL models can be considered an attractive solution. Hence, nonsupervised learning methods can be adapted to maximize the applicability of unlabeled data for training the DL model. However, due to the inherent complexity of remote sensing data and existing challenges in nonsupervised learning, precise learning schemes and specific model designs are required. This article provides a meta-analysis and a systematic review of the latest deep nonsupervised learning methods offered by researchers in the field of remote sensing. There are various approaches for categorizing nonsupervised

Manuscript received 26 April 2023; revised 9 August 2023; accepted 2 September 2023. Date of publication 19 September 2023; date of current version 6 December 2023. The work of Dr. Masoud Mahdianpari was supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery under Grant RGPIN-2022-04766. *(Corresponding author: Masoud Mahdianpari.)*

Benyamin Hosseiny is with the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1417935840, Iran (e-mail: ben.hosseiny@ut.ac.ir).

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

Mohammadali Hemati and Ali Radman are with the Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: mhemati@mun.ca; aradman@mun.ca).

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

Jocelyn Chanussot is with the Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France (e-mail: jocelyn.chanussot@grenoble-inp.fr).

Digital Object Identifier 10.1109/JSTARS.2023.3316733

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

learning methods. However, this study divides them into five main categories: transfer learning (TL) or finetuning, weakly supervised learning (WSL), semisupervised learning (Semi-SL), self-supervised learning (Self-SL), and unsupervised learning (USL).

While considering that supervised learning is performed in an ideal condition with abundant and accurate labeled samples, the main capability of nonsupervised learning is leveraging unlabeled or weakly labeled data to improve the trained model's generalization. TL addresses the problem of limited training data and accelerates the learning process by using a pretrained model from another domain and finetuning it with labeled samples from the main dataset. Unlike TL, other learning methods train a base model without pretrained weights. The key idea behind WSL is training a model with low quality of training data, while the quantity can be sufficient, such as, semantic segmentation with image-level labeled data. Semi-SL algorithms, generally, take advantage of unlabeled data to increase the size of training dataset by iteratively generating pseudosamples. Self-SL uses unlabeled data for unsupervised representation learning and then fine-tunes the model by small number labeled data. Despite these learning methods, USL fully utilizes unlabeled data for modeling.

Several review papers have been published concerning DL applications in remote sensing data processing. However, these studies mainly focus on a few topics, such as general applications of the DL models to remote sensing data [3], [4], reviewing a specific DL architecture [5], [6], [7], or a specific application or dataset [8], [9]. Hence, there is no attention on either learning methods or training strategies in most of these studies. One of the very first comprehensive reviews of DL in remote sensing was published by Zhu et al. [4]. They reviewed well-known models, applications, and benchmark remote sensing datasets for developing DL algorithms. Ma et al. [3] gathered a systematic review and meta-analysis of DL in various remote sensing applications and reviewed related papers in the topic, created a meta-analysis by categorizing the major interconnections between DL and remote sensing data, such as DL architecture, study target, and data resolution. Hoeser et al. [10], [11] published a comprehensive overview of DL models focusing on object detection and segmentation using remotely sensed images.

Apart from a general review of DL in remote sensing, a few reviews have been published focusing on narrow subjects, such as specific remotely sensed data or specific DL architecture. Paoletti et al. [8] provided a review focused on evaluating and comparing the performance of state-of-the-art DL models applied for hyperspectral (HS) image classification. Zhu et al. [9] reviewed the applications and existing challenges of processing synthetic aperture radar (SAR) data with DL algorithms. A comprehensive review of various DL models developed for CD in remote sensing can be found in [12]. A systematic review and meta-analysis of recent trends on the applicability of convolutional neural networks on remote sensing tasks are provided in [5]. Effects and improvements obtained by adding attention mechanism to DL models have been reviewed and discussed in [7].

As shown, many review studies have been published so far, yet none of them focused on long-lasting challenges that exist in applying DL to remote sensing data: well-represented, sufficient annotated training data. Although few studies reviewed applications of nonsupervised deep learning (NSDL) algorithms in remote sensing, they are mostly focused on a small and specific case. For instance, a review by Dong et al. [13] compared different autoencoder models for unsupervised target detection (TD) from SAR imagery. Applicability and performance of graph neural networks (GNN) for HS image classification in a semisupervised manner has been studies in [14].

There are a number of review papers in other fields that are closely related to the main topics of this article, where they discuss the challenges of supervised DL and review the cutting-edge algorithms and applications of NSDL. For example, a review on WSL and its various aspects in general field of data science was provided in [15]. A technical review on the applicability of GNNs in semisupervised problems in image classification was provided in [16]. Ohri and Kumar [17] gathered a review on deep Self-SL for image recognition. Chum et al. [18] reviewed and compared the supervised and unsupervised DL algorithms in the context of autonomous navigation. Peng and Wang [19] provided a comprehensive review on the utilization of DL algorithms in imperfect training data for medical image segmentation.

Despite growing advances of NSDL methods in remote sensing, there is no review paper that specifically focus on NSDL methods in this field to the best of our knowledge. Therefore, the main objective of this article is to provide a comprehensive review of NSDL methods in the context of remote sensing data processing. Accordingly, this study contributes to building a systematic review and meta-analysis on the recently published papers in the field of NSDL for remote sensing processing and discusses the trends and potential future directions in this field. The current research also provides a comprehensive and detailed overview of state-of-the-art NSDL methods using a standard categorization paradigm and, thus, making it easier to identify a suitable solution for remote sensing problems.

II. Method

A comprehensive review process in this article was performed by preparing a systematic literature search query using the SCOPUS, IEEEXplore, and Science Direct databases, which cover almost all the well-known scholarly literature in the field of remote sensing. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) [20] methodology was followed for the selection of articles. Accordingly, after a few trials, three groups of keywords, including the "field of study," "learning method," and "modeling method" were selected to extract relevant research articles from the databases. Thus, the query design for searching the databases is as follows:

("transfer learning" **OR** "fine tuning" **OR** "semi-supervised" **OR** "weakly supervised" **OR** "active learning" **OR** "self-supervised" **OR** "unsupervised") **AND** (Deep" **OR** "Deep learning") **AND** ("Remote sensing" **OR** "Earth observation").

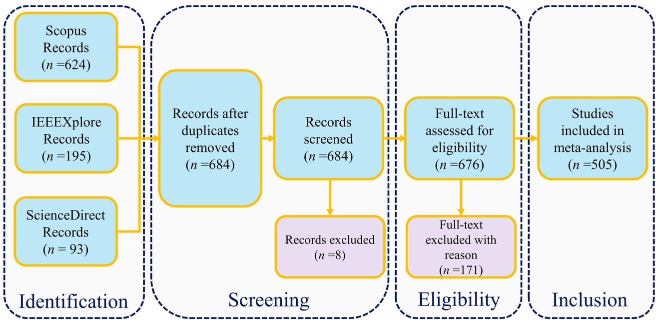


Fig. 1. PRISMA flowchart for meta-analysis.

The database searching and data extraction process were performed until 31 December, 2022. From the existing list of papers, only the research papers from the peer-reviewed journals written in English were selected. At first, 912 research papers were extracted from the combination of the three databases. After removing the duplicated or unrelated papers, 505 papers were eligible for the meta-analysis. Fig. 1 summarizes the database searching and literature meta-analysis process using the PRISMA methodology.

III. Results and Brief Overview

The generated database of the reviewed papers and their corresponding data is used for meta-analysis of the NSDL methods in the remote sensing data processing.

The essential information about a remote sensing dataset could be electromagnetic band, spatial resolution, and type of study area, while the main parts of the modeling process are learning method, training data, and model architecture. Finally, the trained model is evaluated with proper metrics, such as accuracy. Therefore, this section provides the results of meta-analysis of NSDL methods in remote sensing in the following subsections: first, a general characteristic of evaluated studies, such as a simple statistical analysis of published papers in different years and journals, is provided. The Next, deep architectures used within the nonsupervised learning scheme in the remote sensing studies are reviewed in the second subsection. This is followed by an analysis of results attained from different training sizes and datasets in the third and fourth sections, respectively. In the last part, the statistical results of NSDL methods applied to various study targets of remote sensing are shown and discussed.

A. General Characteristics of Studies

Fig. 2 shows the number of published journal papers in the investigated field from 2015 to 2022. Comparing the number of published papers in 2015 (3) and 2022 (232) demonstrates an increasing interest in applying nonsupervised learning for remote sensing problems within less than a decade. We also showed the number of open-access codes regarding the published papers in each year. The results show that as interest in developing NSDL methods increases, researchers are more willing to share their developed code publicly. Furthermore, a list of most frequently publishing journals, which issued more

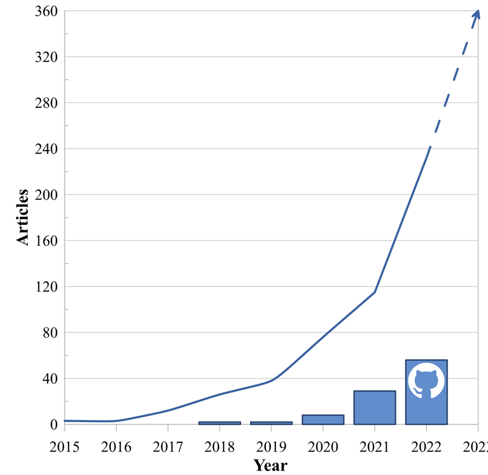


Fig. 2. Annual frequency of the published NSDL papers in remote sensing applications.

than two papers evaluated herein, is provided in Fig. 3. As demonstrated in this figure, *Remote Sensing MDPI* and IEEE Transactions on Geoscience and Remote Sensing are two top journals that have the largest number of publications in this field. However, the IEEE family journal contained the largest portion of papers and, thus, can be considered as the main publisher in this field.

As shown in Fig. 4, the majority of published papers were related to USL (33%) and TL (26.9%), respectively. This is followed by the Semi-SL methods with 21.6% of published papers, and only 10.1% and 7.3% were focused on the Self-SL and WSL methods in their DL model learning process. Notably, Self-SL and USL seem to have gained the highest attentions in the past two years, according to their sharp increase rate. Importantly, the use of WSL methods has not yet been well received by researchers. This could be due to the complexity of WSL methods for remote sensing applications compared to the other mentioned learning methods.

B. Deep Architectures

Developed deep models primarily employ a combination of multiple deep modules and base architectures to achieve state-of-the-art results. Fig. 5 presents the statistics on the usage of deep modules and base architectures for deep nonsupervised modeling of remote sensing data. We define a deep module as a set of operations or layers that transform input data into desired output representations, such as Residual and Attention mechanisms, while the base architectures consist of multiple modules, such as Transformer.

As depicted in Fig. 5(a), the Residual module is the most commonly used module in the construction of deep nonsupervised models, with a total of 111 records. On the other hand, although Attentions are relatively newer compared to other modules, they exhibit higher usage (62) than the Recurrent (17) and Inception (11) modules. This can be attributed to the Attention mechanism's ability to enhance focus on the most informative and relevant features within the data, thereby facilitating the

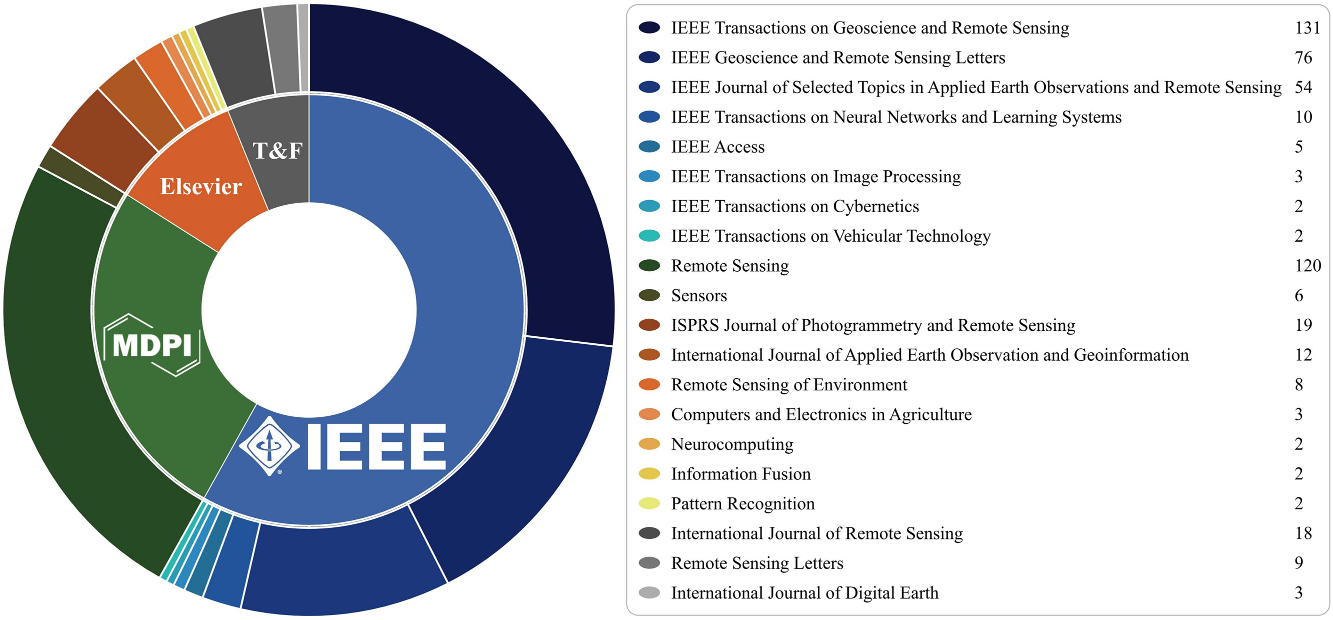


Fig. 3. Distribution of the published papers divided by journals and publishers.

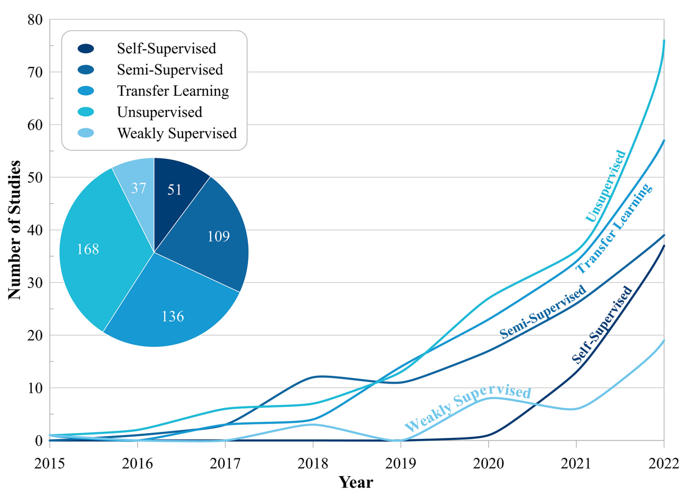
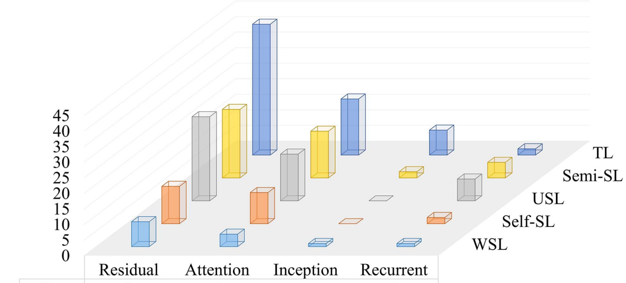


Fig. 4. Annual frequency of published NSDL papers in remote sensing applications, divided by learning methods.

learning process while reducing the reliance on large quantities or high-quality training data.

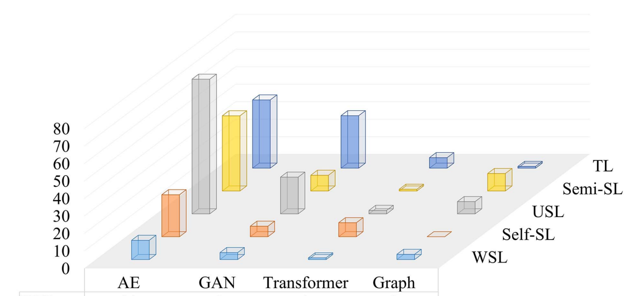
Regarding the deep architectures [see Fig. 5(b)], AE, followed by GAN, have been the most frequently employed architectures for deep nonsupervised modeling of remote sensing data, with 194 and 70 studies, respectively. In contrast, Graph (21) and Transformer (18) exhibit the lowest usage. The significant usage of AE in USL models and GAN in TL models is worth noting. This can be attributed to the automated feature learning capabilities of AE and the strong ability of GAN to generate synthetic data and address domain shift in TL problems through adversarial learning. Furthermore, although Transformer has shown the least usage among the studies, it is important to highlight that this architecture was introduced in 2021, and we anticipate an increase in the number of studies involving this architecture in the future years.

(a)



	Residual	Attention	Inception	Recurrent
WSL	8	4	1	1
Self-SL	12	10	0	2
USL	27	15	0	7
Semi-SL	22	15	2	5
TL	42	18	8	2

(b)



	AE	GAN	Transformer	Graph
WSL	11	4	1	3
Self-SL	24	6	8	0
USL	77	21	2	7
Semi-SL	43	9	1	10
TL	39	30	6	1

Fig. 5. Frequency of the published papers for each nonsupervised learning method using different (a) DL modules and (b) architectures.

C. Training Data

Amount or fraction of available labeled training data is an important factor in order to choose a suitable learning method in DL modeling process. According to the definition of USL in this review, we assume that USL does not use any annotated

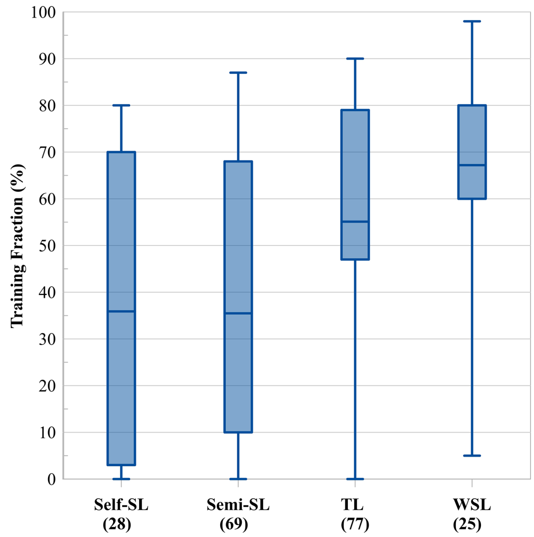


Fig. 6. Training fraction distribution versus nonsupervised learning method.

samples and, thus, the training fraction is zero. Fig. 6 shows the fraction of training samples used for training the model based on the other learning methods. Accordingly, Semi-SL and Self-SL algorithms have used the lowest training percentage from the available ground truth with 35.9% and 36.4%, while TL and WSL algorithms, respectively, have used 57.2% and 68.9% of the available ground-truth dataset for training their model. Thus, TL methods consumed more training data compared to Semi-SL and Self-SL, while the papers based on these methods cover a wider range of training data. Compared to other methods, it is noticeable that WSL seemingly requires higher fraction of training data. In particular, WSL methods increase the modeling accuracy by improving the low-quality training dataset. As such, data that have been used in the WSL studies may be high in quantity but not good enough in quality.

D. Data Type

Selecting a suitable data among the existing extensive datasets is a crucial step in solving a remote sensing problem. According to the database collected in this review study, until now, the main goal has been centered on the algorithm development without considering data type. However, this section categorizes the applied datasets in the reviewed articles based on the type of sensors and spatial resolution. Results in Fig. 7 show that three band RGB images were the most used dataset in all four learning methods with 162 studies, followed by HS and multispectral (MS) datasets with 103 and 105 studies, respectively. In contrast, only few studies dedicated to active sensors, where SAR have been used in 40 studies, and only seven papers used LiDAR as their data source. Notably, 41 papers used multiple sources of data in their research work. Fig. 8 depicts the frequency of used sensors based on their spatial resolutions, for each learning method. Accordingly, high portion of studies (232 papers) have used very high-resolution (VHR) images, while fewer studies have been dedicated to high-, medium-, or low-resolution datasets. This is aligned with the results obtained for

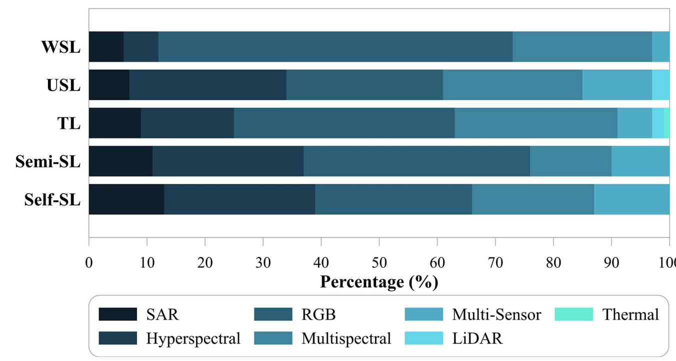


Fig. 7. Usage rate of data types for each learning method.

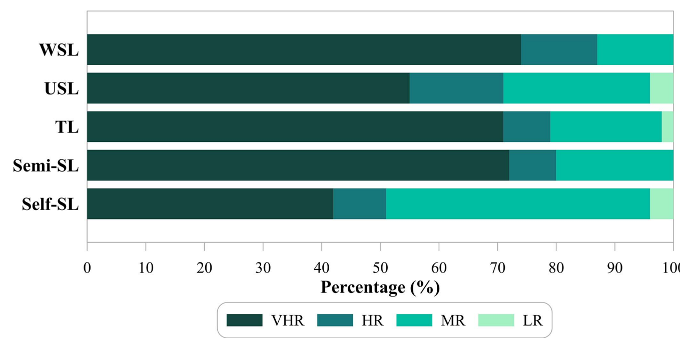


Fig. 8. Percentage of different data with various spatial resolution used in various learning method (VHR: <4 m, HR: 4–10 m, MR: 10–30 m, LR: >30).

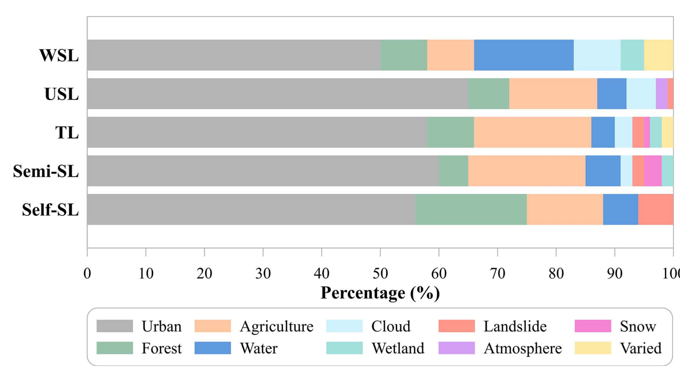


Fig. 9. Usage rate of study areas for each learning method.

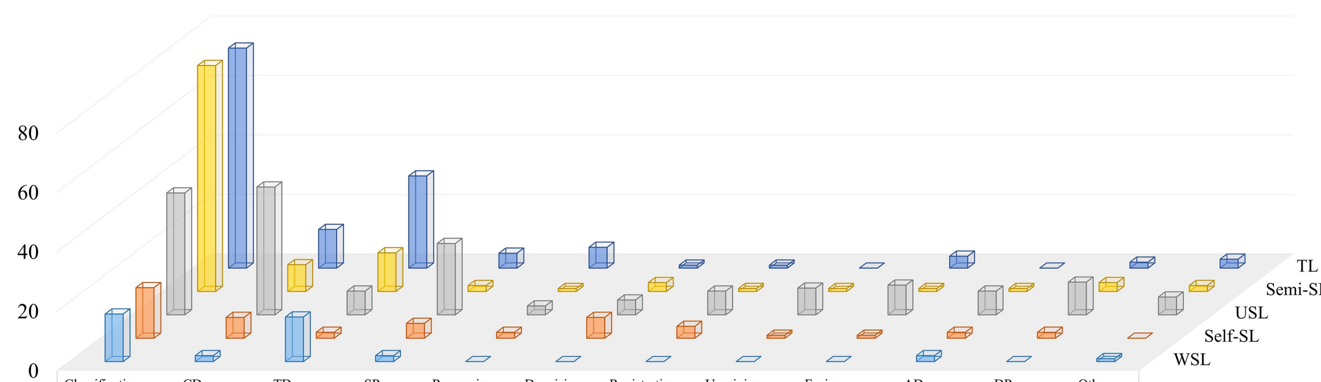
investigating the sensor type (see Fig. 7), where lower spectral discriminability can bring higher spatial resolution. Table I summarizes the most commonly used benchmark datasets in the field of NSDL in remote sensing, along with their characteristics and main applications.

E. Study Target

As shown in Fig. 9, urban, agriculture, and forests are the most common study areas investigated in this topic given their wide coverage and importance. Few studies used TL, Semi-SL, WSL methods for wetland mapping and classification. There were also few works on using TL in order to improve the modeling results on snow or land slide areas. Furthermore, few studies

TABLE I
SUMMARY OF PUBLIC BENCHMARK DATASETS SUITABLE FOR NSDL ALGORITHMS IN REMOTE SENSING

Dataset	Type	Categories	Images	Resolution	Image size	Applications
UC-Merced [139]	RGB	21	2100	0.3	256 × 256	Scene classification
NWPU-RESISC45 [140]	RGB	45	31 500	0.2-30	256 × 256	Scene classification
Bigearthnet-S2 [69]	MS	19–43	590 326	10	120 × 120	Scene classification
EuroSAT [141]	MS	10	27 000	10	64 × 64	Scene classification
SpaceNet-4 [142]	MS	1	28 728	0.5	900 × 900	Weakly supervised building footprint detection
OSCD [143]	MS	2	24	10	600 × 600	Urban change detection
Hermiston city [144]	HS	5	1	30	390 × 200	Multiple crop change detection
ABU [145]	HS	2	13	17	100 × 100, 150 × 150	Anomaly detection
MSTAR [146]	SAR	10	5950	0.3	128 × 128	Military object classification
NIST-SAR [38]	SAR	10	2746	0.3	128 × 128	Few-shot military object classification
Bigearthnet-S1 [70]	SAR	19–43	590 326	10	120 × 120	Scene classification
Bigearthnet-MM [70]	Multisensor (SAR & MS)	19–43	590 326	10	120 × 120	Scene classification
SEN12MS [147]	Multisensor (SAR & MS)	17	180 662	10	256 × 256	Scene classification
SEN12MS-CR-TS [148]	Multisensor (SAR & MS)	1	18 300	10	256 × 256	Cloud detection and removal



	Classification	CD	TD	SR	Regression	Denoising	Registration	Unmixing	Fusion	AD	DR	Other
WSL	16	2	15	2	0	0	0	0	0	2	0	1
Self-SL	17	7	2	5	2	7	4	1	1	2	2	0
USL	41	43	8	24	3	5	8	9	10	8	11	6
Semi-SL	76	9	13	2	1	3	1	1	1	1	3	2
TL	74	13	31	5	7	1	1	0	4	0	2	3

Fig. 10. Frequency of the published papers for each nonsupervised learning method and remote sensing study targets.

took advantage of USL in order to cope with the training data challenges in studying of atmosphere.

Fig. 10 shows the frequency of the published papers for different applications in each learning method. According to the results, majority of papers focused on classification (including segmentation and land-use land-cover classification) (224), followed by CD (75) and TD (69). There is a high number of papers that utilized TL or Semi-SL in scene classification and TD by remote sensing images using DL. On the other hand, USL demonstrates its usage in applications that having annotated data is not a common approach, such as anomaly detection (AD) and CD, where automated frameworks are favorable. Similarly, USL is also applied in data preprocessing or enhancement related applications such as denoising, dimension reduction, SR, spectral unmixing, data fusion, and registration. Although the WSL method has been less frequently used compared to other learning methods, it has mainly been employed in classification and TD.

IV. Nonsupervised Learning Methods

Deep supervised learning methods have obtained state-of-the-art results in many fields, including several remote sensing applications [3], [4]. However, these methods necessitate a considerably huge and comprehensive training dataset in order to present their best performance. Furthermore, they are prone to overfit on training data, and their performance extremely degrades on the new domain of test datasets [18]. In order to address the mentioned drawbacks, nonsupervised learning methods are proposed. There is no unique categorization of nonsupervised learning methods in the literature, and it is not feasible to find hard boundaries to separate these methods. However, in this article, we formulate them based on the most commonly used methods in remote sensing. To this end, we divide nonsupervised learning strategy into five broad subcategories: TL, WSL, Semi-SL, Self-SL, and USL.

A. Transfer Learning

In TL, a model is first trained on a dataset from one domain and then reused for a dataset from another domain. TL is also referred to finetuning or pretraining strategy in the literature [21], [22]. Generally, training a deep model with thousands of learnable parameters from scratch requires a large-scale training dataset. Hence, the main goal of TL is alleviating the problem of limited training data and accelerating the learning process by using the pretrained model from another domain and finetuning the model with the existing labeled samples of the main dataset. The most prevalent approach in TL is to utilize base models that have been pretrained on large-scale natural benchmark datasets, such as ImageNet [23] or COCO [24]. These models are then finetuned on the desired dataset. While this approach is effective in reducing the need for a large number of training samples in computer vision applications, it is not as effective in most scenarios for remote sensing due to the different geometry and nature of the acquired data, and still large number of training data is required for these applications. Therefore, it is important to note that we did not consider this approach as a nonsupervised TL method for remote sensing applications.

Data augmentation by generating synthetic or simulated data can be considered as another approach in TL [25], [26], [27], [28]. Data augmentation by generating synthetic or simulated data is another TL approach [25], [26], [27], [28]. One primary advantage of using synthetic data is that the precise annotation is become available for free. On the other hand, collecting and annotating data for a large number of problems is not only a tedious process but also prone to human errors [18]**.** Recent studies have shown that GAN models are effective in increasing the number and diversity of training data, and balancing the imbalanced labeled data by synthesizing new samples (see Fig. 11).

The main drawback of TL would be the domain shift between the synthesized training data and the test dataset [29], [30].

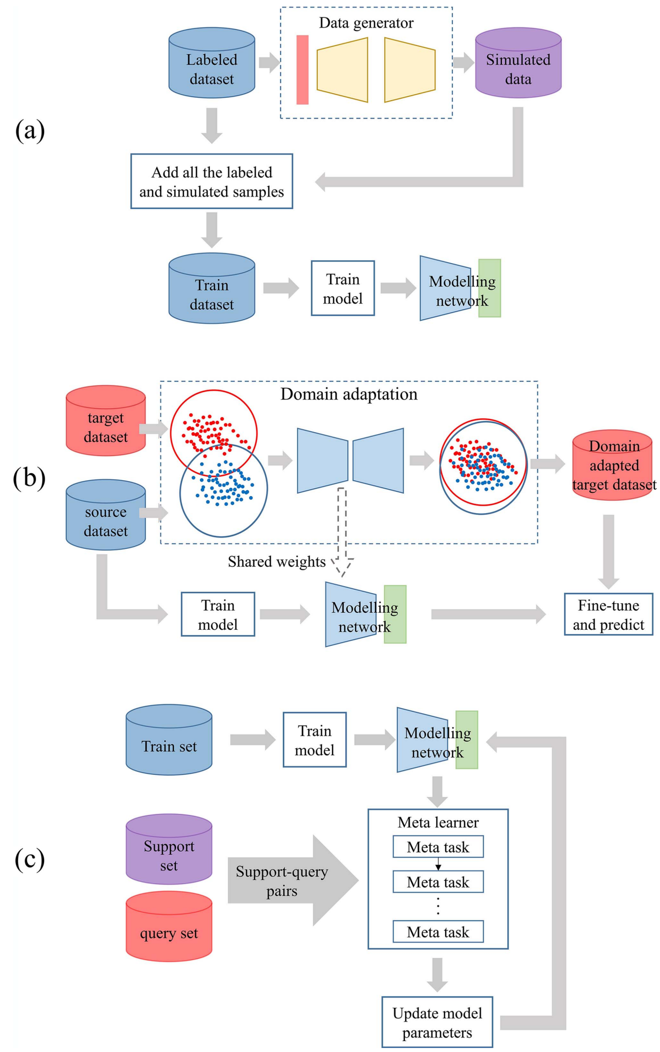


Fig. 11. TL approaches. (a) Data augmentation by synthetic data. (b) Domain adaptation. (c) Few-shot meta learning.

More specifically, in remote sensing problems, the existence of different sensors for multisensor or multitemporal problems or different illumination, reflectance, and topographic conditions on large scales can be known as the main sources of domain inconsistencies [31]**.** Consequently, this degrades the generalization capability of the trained model on test samples [6]. Domain adaptation can address the mentioned limitations in remote sensing data processing by minimizing the distribution gap between source and target domains [18], [31]. Diverse domain adaptation scenarios can be defined in the context of remote sensing data processing, but one categorization can be between number of source domains (single or multiple) as training data and number of target domains as test data [31]. For further details on domain adaptation in remote sensing, we refer to [32].

Another subdivision of TL is few-shot learning (FSL) methods. In this scenario, only a limited number of ground-truth samples are available, which may originate from different domains or even have distinct labels compared to the trained model [17], [18]. FSL mainly comes with meta-learning in literature. Despite conventional ML, which is sample training, meta-learning is a training task that updates its learning rules in multiple iterations and based on the knowledge of previous tasks. Meta-learning comprises multiple tasks, where each task contains a pair of support and query sets. The learning processing accomplishes in two different inner and outer levels. At the inner level, the base learner concerns with plenty of single tasks to learn from the new observations. In contrast, at the outer level, the meta learner adapts the knowledge gathered from previous tasks to generalize a new task [33], [34], [35], [36], [37], [38]. Accordingly, the model's learning improves by the time and with more experience. Thus, meta-learning learns the learning rule or update function, which helps to converge the model with very few training samples and improves the model's generalization by updating the learning function. However, as mentioned in [38] and [39], one limitation is that this method requires a meta-training dataset rich in class diversity to attain a generalizable meta-model. We refer to [39] for further details on few-shot and meta-learning.

B. Weakly Supervised Learning

Collecting high-quality and large-scale remote sensing annotation data is time-consuming and expensive, whereas small number of high-quality annotations may result in an overfitted model [40]. Hence, the WSL method aims to bridge between the low-quality training data and well-trained model. In other words, compared to the rest of nonsupervised algorithms, instead of quantity, the WSL's goal is to improve the model's performance by improving the training data quality. Generally, WSL algorithms deal with inaccurate, or inexact training data to attain a highly generalizable model [17], [18]. Thus, to make it more specific, we can say that WSL algorithms consist of two major subfields of inaccurate (noisy) and inexact supervision [15], [18], [19]. In inaccurate supervision, the training data are contaminated with noise, decreasing the ground-truth reliability. For example, in a classification problem, noisy labels can affect the model's performance and are the source of inconsistency [15], [19]. Inexact supervision refers to the situation where only a coarse level of it is available instead of the exact required supervision information [41], [42], [43], [44], [45], [46]. For instance, pixel-level image segmentation only has image (patch) or point level training data [15], [18].

C. Semisupervised Learning

The main idea behind the Semi-SL method is to improve the model's performance by making use of unlabeled data in the model's learning process [15], [47]. More specifically, Semi-SL mainly aims to leverage the power of thousands or millions of unannotated samples in a dataset to reinforce the training process and improve the generalization of the final model. So compared to the other nonsupervised approaches, Semi-SL's primary strategy is to generate pseudolabels for test samples [15], [47]. Mainly, few criteria or conditions are needed to select high-quality pseudolabels and ignore the uncertain labels to be evaluated in the subsequent iterations. Accordingly, apart from human interpretation and annotation of candidate samples [19],

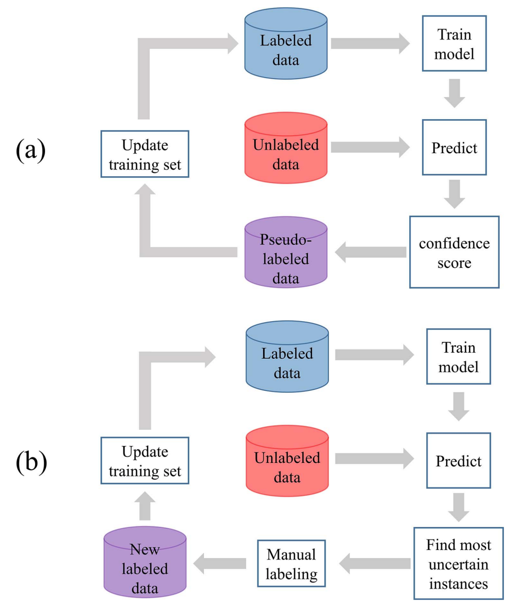


Fig. 12. (a) General semi-SL pipeline. (b) General AL pipeline.

some of the suggested approaches in the literature could be evaluating the predictions of ensemble models [48] or Bayesian modeling evaluation, which is estimating the class assignment confidence based on the posterior probabilities [49].

Active learning (AL) can also be regarded as a subdivision of Semi-SL. In essence, AL commences with training a learner using a limited-sized training dataset. Following the initial model training, unlabeled data are employed for uncertainty sampling. Samples exhibiting high levels of uncertainty as predicted by the model are chosen for human annotation. Subsequently, these newly labeled samples are incorporated into the training dataset in order to retrain the model [15], [18], [47], [48], [49], [50], [51], [52]. Fig. 12 Demonstrates the general pipeline for Semi-SL and AL.

D. Self-Supervised Learning

Self-SL is a relatively new approach that has gained much attention in recent studies, due to its capability in reduction of required labeled data with representative feature extraction from unlabeled data [17], [40]. General pipeline of Self-SL consists of two main steps: first unsupervised representation learning, and then supervised finetuning of the pretrained model for the corresponding main application [18] [see Fig. 13(a)]. In unsupervised representation learning, underlying high- or low-level features of the dataset are extracted by leveraging the huge number of unlabeled samples. This step is mainly performed based on a pretext learning task, which aims to learn generalizable representations of the under the test dataset. At this point, model output is mainly defined from the input data properties, such as signal reconstruction, predicting a pretext task label, or contrasting the similarity between the input signals [17], [40]. In the second stage, trained model is fine-tuned by the available training data in a supervised manner, for a

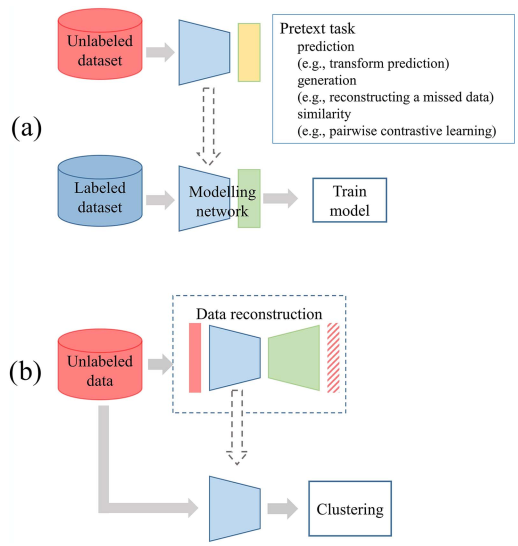


Fig. 13. (a) General pipeline of Self-SL. (b) General pipeline of USL.

corresponding application. The main difference between Self-SL and USL can be in term of their methodology. In USL, no external labels exist to aid the learning process [17]. Thus, USL mainly find patterns by utilizing the statistical properties of the data samples [40], whereas Self-SL tries to extract representative features by self-training. Thus, it uses far more feedback signals than USL during the training processing [17], [40]. Main steps of Self-SL might resemblance TL, where it finetunes a pretrained model. However, in contrast to TL, pretraining stage needs no training data in Self-SL, and the whole training process needs much smaller training samples than TL. Moreover, domain gap is much less than TL, due to using a same dataset for pretraining and finetuning [18], [40].

E. Unsupervised Learning

Contrary to supervised learning and other nonsupervised methods mentioned earlier that depend on labeled data for training, USL focuses on uncovering significant and distinguishing patterns without using any ground truth or prior knowledge [53], [54], [55], [56]. The main idea behind USL is to propose a suitable model that can discover latent representations from the data by reducing the reconstruction loss of the input signals or estimating pairwise similarity/dissimilarity. This allows providing more distinguishable latent representations from the data, enabling tasks such as clustering, dimensionality reduction, or AD [see Fig. 13(b)].

Basically, in DL, unsupervised algorithms are mainly proposed by stacked encoder–decoder (SAE) architecture to automatically extract distinctive features [57], [58], [59], [60]. In recent studies, adversarial models, such as GANs and contrastive learning models, were utilized to extract inherent data patterns or generate pseudotraining datasets by leveraging the min–max game of two contrastive loss functions in the training stage [17], [18], [19] [61], [62], [63], [64], [65].

With the provided overview of the most prominent NSDL methods in remote sensing, Table II highlights their main advantages and disadvantages. This aims to make easier to choose a suitable learning method for solving remote sensing problems.

V. NSDL in Remote Sensing

Applications and recent advances of the DL algorithms in remote sensing have been discussed in previous review papers [3], [4], [10]. Thus, following the taxonomy provided in the previous literature, this section aims to discuss and give some insights into the applications of NSDL algorithms in remote sensing data processing.

According to the results shown in Fig. 10 of this review, the primary focus of studies is on the development of NSDL algorithms for remote sensing image classification (including segmentation and land-use land-cover mapping) followed by TD and CD studies. Although small portion of recent studies are focused on other applications, such as AD, and data enhancement and preprocessing applications, such as fusion, denoising, registration, they are still worth mentioning in this section in order to draw more attentions for future studies.

A. Classification and Semantic Segmentation

Image classification and semantic segmentation have the most application in using DL models, due to their importance and also availability of various benchmark datasets. Remote sensing classification tasks can be object classification from small patches, such as classifying airplanes, buildings, or cloudy images or pixelwise scene classification or semantic segmentation, such as land use land cover, crop types, or wetlands. However, as mentioned earlier, collecting a rich ground truth in order to produce a generalizable model is an expensive task. According to the results, NSDL algorithms offer various solutions to these issues. As shown in Fig. 10, the most common approaches are TL and Semi-SL.

The basic idea of utilizing TL for remote sensing in the early stages of DL involved finetuning pretrained base DL models, such as the ResNet family [66], VGGs, and AlexNet [67], [68]. However, this approach relies on models that were originally pretrained on large-scale RGB computer vision datasets, such as ImageNet [23] or COCO [24], which are not well-suited for remote sensing datasets. Remote sensing datasets typically have different imaging geometry and may contain more than three channels, such as MS and HS images. As a result, a significant number of training samples are still required to obtain a reliable model using this approach, which does not differ significantly from supervised training.

Recent developments addressed the issues related to TL from computer vision pretrained datasets by releasing large remote sensing datasets such as BigEarthNet S2 [69], S1, and MM [70], which are labeled image patches from Sentinel-1 and 2 for image classification. Furthermore, Yang et al. [71] investigated the TL between two HS datasets, as source and target domains. Generating simulated or synthetic samples can also provide more customizable and diverse training data with a lower domain distance from the target dataset. The feasibility of this process has been increased significantly with the emergence of

TABLE II
SUMMARY OF THE NSDL METHODS IN TERMS OF THEIR STRENGTHS AND CHALLENGES

Learning method	Strengths	Limitations and challenges
Transfer-learning	• Leveraging knowledge from a model trained on large datasets, even when the target dataset is small. • Reducing the need for extensive computational resources and training. • Improving the model's performance by initializing the network with pre-trained weights.	• The pre-trained model may not be suitable for the target task, leading to suboptimal performance. • There might be a domain shift between the source and target domains, affecting the transferability of learned features. • Different preprocessing steps between the source and target datasets might affect the fine-tuning performance.
Semi-supervised learning	• Leverage a large amount of unlabeled data, which is often easier to obtain than labeled data. • Improving model performance by utilizing both labeled and unlabeled data. • Reducing the reliance on manual annotation, which can be expensive and time-consuming.	• Model's performance may vary depending on the quality and quantity of the available unlabeled data. • Designing effective approaches for scoring uncertainties and pseudo-labels can be complex and challenging. • Some approaches might require additional assumptions about the underlying data distribution and the relationship between labeled and unlabeled data.
Weakly-supervised learning	• Reducing the need for fine-grained annotations by using weaker forms of supervision, such as image-level, point-level or bounding boxes labels. • It enables training on larger datasets, as weak annotations are often easier and cheaper to obtain.	• The model's robustness and performance can be highly affected by the quality of the training data. • Designing effective high resolution and accurate model can be challenging with low-level annotations. • Model's generalization capability can be hindered by the quality of labels and uncertainties.
Self-supervised learning	• Leveraging large amounts of unlabeled data to learn useful representations. • It can pre-train models on auxiliary tasks and then fine-tune them on downstream tasks, leading to more robust and improved performance.	• Designing effective pretext task can be challenging. • Performance and effectiveness of the model is influenced by the pretext task and complexity of the unlabeled data.
Unsupervised learning	• It can discover hidden patterns and structures in unlabeled data. • Mapping data into latent subspace, making them more distinguishable. • It does not require manual annotation, making it more scalable and cost-effective.	• Can be more challenging than the other learning methods due to the lack of labels. • The quality of learned representations heavily depends on the underlying assumptions and algorithms used. • Evaluating the performance of unsupervised learning models can be difficult without a clear objective or ground truth.

generative models. Numerous recent studies have observed the effectiveness of GAN models for remote sensing data classification by addressing the imbalanced or scarce training datasets by generating synthetic data. This approach has been successfully implemented in classification tasks across diverse fields [25], [28], [72]. Recent studies in the past few years tried to develop novel methods in TL with very few annotated data. To this end, GNN model was presented in [73] for few-shot VHR RGB image classification, while Cao et al. [74] investigated 3-D convolution Siamese model with a contrastive loss for HS classification. Furthermore, the number of studies investigating the few-shot meta-learning to facilitate and improve the learning process in the presence of limited training samples is increasing. For instance, they are being implemented in target classification with

SAR imagery [38] and scene classification using Sentinel-1 and 2 [75].

Similar to TL, Semi-SL has demonstrated the highest frequency in the reviewed papers of this study, within the field of remote sensing image classification. In fact, many unannotated samples are available in a vast remote sensing image scene that can be effectively exploited in a semisupervised process for improving the model generalization and final classified results. This has resulted in numerous advancements over the past few years in the development of DL frameworks that are based on various methods of Semi-SL for classification of remote sensing imagery. Along these lines, deep AL was developed and investigated for different remote sensing data classification, such as RGB [76], HS [50], [51], and polarimetric SAR [48], or different applications, such as crop type mapping [52]. For instance, Liu et al. [51] proposed an AL deep belief network for HS classification, where their algorithm updated training samples and weighted their importance in every iteration of learning process based on uncertainty and similarity criteria.

Regarding the deep WSL, the number of studies for remote sensing data classification is considerably lower than other learning methods, but in overall, it is increasing. In fact, preparing reliable and accurate labels for large-scale multilabel image classification is a difficult and expensive process that also involves human errors. To reduce the effect of labeling uncertainties, WSL methods have been investigated in various fields of classification tasks, such as image-level annotated WSL for building segmentation using VHR RGB imagery [77], and tree species classification using VHR MS and LiDAR [45], and point-level annotated WSL for road network [44] and water body [78] segmentations. Furthermore, as shown by Wang et al. [79], even weak training samples in point-level or image-level annotations combined with a deep model, such as U-Net, can outperform supervised baseline methods, such as support vector machine (SVM).

The published papers investigating the Self-SL for remote sensing image classification are relatively newer than the rest. Similar to Semi-SL algorithms, Self-SL also requires few annotated data, but as mentioned earlier, the crucial part of Self-SL is unsupervised pretext learning, where the model aims to find nonlinear and distinctive patterns from the unannotated data. Thus, deep Self-SL models were developed in various remote sensing classification fields, including VHR imagery scene classification [80], crop type mapping using time-series of Sentinel-2 imagery [81], [82], HS classification [83], [84], PolSAR [85], and multimodal [86] land cover mapping. Among these, a BERT transformer deep Self-SL model was presented in [81] for multilabel crop classification from time-series of Sentinel-2 imagery. They manually added noise to some of input time-series and the model was trained to predict the noisy inputs as pretext task for unsupervised feature learning from a pool of unlabeled data. A Self-SL classification framework using GAN model was presented in [80], where a similarity loss between two different views of a same image was proposed as pretext task of this framework.

First USL studies on remote sensing classification used mainly shallow models such as RBM or AE architectures for automated feature representation learning using reconstruction or distance-base loss functions. In fact, early studies of NSDL algorithms used mainly different variations of AE architectures, such as convolutional, denoising, and sparse models, for unsupervised representation learning with a relatively simpler classifier or clustering algorithm on top of model. For instance, Tao et al. [87] used stacked sparse AE for unsupervised feature learning combined with SVM classifier for HS scene classification. A joint multiple parallel SAE and CAE was presented in [88] for unsupervised land cover classification using multimodal VHR HS, LiDAR, and digital surface model (DSM) dataset. Radman et al. [89] proposed an unsupervised burnt area mapping using Sentinel-1 SAR data, where a convolutional neural network (CNN) model was trained based on combination of a fuzzy c-means clustering and a saliency-guided segmentation.

B. Target, Anomaly, and CD

Anomalies, targets, and changes all follow a similar characteristic in remote sensing datasets, which is their scarce presence compared to the background samples. This makes modeling process so challenging due to the imbalanced distribution of object of interest and the surrounding background. NSDL algorithms can tackle this issue and prevent model failure.

Majority of NSDL studies for TD are focused on TL and Semi-SL, which quantitively is similar to what we observed in studies regarding classification. Various studies investigated TL from pretrained models on large datasets for TD, such as airplane detection from VHR RGB images using SSD [90] or R-CNN [91] deep object detectors, which their backbone deep model is a pretrained VGG16. To address the issues of high complexity and computation cost associated with pretrained models in TD in VHR imagery, due to their numerous learnable parameters, Chen et al. [92] presented a compressed deep model. This model aims to estimate the parameters of the complex model by minimizing the l_2 loss between the output of the compressed and complex models. Li et al. [93] proposed a CNN with attention transformer model in order to reduce the domain gap between pretrained model on ImageNet dataset and remote sensing dataset for TD. A TL framework between satellite MS imagery for cloud detection was presented in [94]. They proposed a convolutional AE model with domain adaptation between Landsat-8 and Proba-V imagery. As mentioned earlier, GANs can improve the model's generalization and robustness with adversarial learning and assist with data augmentation with synthesizing training samples. This capability was investigated in target generation and detection with VHR RGB [95], [96] and SAR [97]. More specifically, a conditional GAN framework for aircraft detection was presented in [98], while a vehicle detection framework based on GAN was presented in [99]. Simulated TL was also investigated in plant detection using UAV RGB imagery, where Hosseiny et al. [27] presented an image simulation scenario to generate a pool of simulated training data for their R-CNN plant detection framework.

The number of studies of deep Semi-SL models for TD is fewer than TL, but they cover a wide range of applications such as the detection of tree-crowns, bridges, landslides, buildings, vehicles, and roads using VHR RGB imagery. Xu et al. [100] proposed a multiscale hierarchical AL algorithm using CNN model for detecting vehicles from large-scale remote sensing dataset. Xia et al. [101] investigated the incorporation of Semi-SL and CNN for building edge detection. Liao et al. [102] proposed a joint Semi-SL and TL framework for TD from SAR imagery. They trained model with small number of SAR and abundant amount of optical imagery, and then updated the training samples during semi-SL process. The main focus of WSL studies in the field of remote sensing TD is pixel-level detection of target of interest using low-level annotated data. Li et al. [46] presented a WSL convolutional framework for cloud detection from high-resolution satellite imagery using image-level annotated data. A CNN model based on multiscale class activation maps was proposed for WSL building detection from VHR imagery in [103]. As first studies on USL TD, an unsupervised DBN model for VHR imagery was presented in [104]. Shao et al. [105] proposed a cloud detection framework utilizing Landsat imagery and unsupervised feature learning based on a fuzzy autoencoder model. An unsupervised training sample generation joint with CNN was investigated in [106] for urban tree detection using MS images and DSM. A few numbers of Self-SL studies were focused on TD. Nevertheless, a Self-SL with rotation pretext task was proposed for SAR target recognition [107]. In another study, an adversarial contrastive Self-SL model was developed for SAR TD [108]. For HS TD, Yao et al. [109] proposed a Self-SL framework with spectral similarity pretext task.

Although AD basically is an unsupervised process, very few studies could be detected through the reviewed database. Among these almost all the studies were focused on HS imagery. An investigation of convolutional AE for unsupervised HS AD can be found in [58]. Jiang et al. [110] proposed a GAN-assisted unsupervised HS AD framework. In this work, first a salient map of input data is generated, where the approximate anomalous areas are weakly extracted. Then, based on the approximate anomaly and background pixels, a weakly supervised spectral discriminative GAN model was trained for final AD. Rao et al. [111] designed an HS AD framework based on TL and contrastive Siamese network. First, model was trained on a dataset with known labels, where the input signals are similar and dissimilar pairs. Then, the model was finetuned using the dataset containing anomalous pixels in order to iteratively detect anomalies.

According to our database majority of the CD works were focused on USL, which demonstrate the high demand of automated CD systems during the time. Similar to the other applications, first DL unsupervised change detection studies were based on automated feature learning using different AE architectures. Zhang et al. [60] proposed a Siamese-like multilayer perceptron network, where the initial change map was obtained by overclustering, but the final change map was generated after iterative optimization of the labels. Ren et al. [64] proposes a GAN-based unsupervised framework for VHR image CD. The main focus of this work is leveraging generative capability of GAN to alleviate the misregistration error between objects in image pairs. Luppino et al. [112] presented a convolutional AE model aiming unsupervised multimodal CD, where they aligned the encoded image from one modality to be decoded based on another modality. TL was exploited in various CD studies, such as urban [113], deforestation [114], land cover [115], and buildings [116]. Fang et al. [117] proposed a TL framework, where at first, feature maps of each bitemporal data are generated using pretrained model and then change map is generated by image differencing change vector analysis in an unsupervised fashion. Soto et al. [118] proposed domain adaptation TL model, which was aimed for balancing the target domain samples as a prevalent issue in CD tasks. In the context of Semi-SL, the main idea was to propagate the number of annotated changed pixels during the training process, in order to decrease the imbalanced distribution between change and no-change annotations. Hao et al. [119] proposed a multimodal framework for SAR and MS CD, where they proposed a clustering label propagation method for generating pseudo-labeled pixel candidates, and after a denoising step, most confident pixels were added to the model learning process. A Semi-SL Siamese-Graph Attention guided framework was proposed in [120], where after training the model on a small training set, unlabeled data were predicted and the most confident samples were used for finetuning the model. Dong et al. [121] proposed a GAN deep model for reducing the domain difference between bitemporal dataset in a self-supervised fashion. They exploited GAN discriminator as pretext learner in order to predict the input patch's time, which is pre- or postchange in this example. A Self-SL ResUnet framework was presented in [122]. They used contrastive loss incorporated with dual branch AE in order to suppress effect of seasonal changes between bitemporal input images.

C. Data Preprocessing (Denoising, Fusion, and SR)

The previously discussed applications in the preceding sections mainly utilize learning methods and DL to process the available data, in order to generate classified maps or detect specific targets or anomalies. However, DL can also be used in the preprocessing step, such as signal (or image) SR, registration, denoising, spectral unmixing, and data fusion, which their main goal is to enhance data. In such scenarios, NSDL methods could potentially play more crucial roles than supervised methods, due to the reason that classic preprocessing algorithms are typically automated with the least reliance on annotations.

The aim of SR is to improve the signal resolution spectrally, or spatially in order to provide more details of the sensed signal. In this regard, various unsupervised DL methods were developed mainly based on AE, CNN, or GAN deep architectures. An unsupervised SR based on fully CNN AE was presented in [123], where the authors aim to generate HR image by downsampling and upsampling layers and optimizing the distance between the low-resolution input image and downsampled reconstructed high-resolution image. Nguyen et al. [124] proposed a multitask CNN model for unsupervised sharpening of Sentinel-2's 20 and 60m bands into full resolution. Ma et al. [125] proposed GAN architecture for MS SR, where they used weakly supervised saliency for increasing the focus and quality of the reconstructed HR image in the informative areas. Hong et al. [126] proposed an adversarial Self-SL model for HS SR using MS imagery. In their method, first they roughly combine low-resolution HS and high-resolution MS images using convolutional AE and adversarial loss to generate recombined HS and MS images with least domain gap, then the reconstructed images are fed to a self-SL model to compare and couple the estimated abundance maps and endmembers in order to create a HS image with high spatial resolution.

Image registration is the process of geometric alignment of different images captured from multiple sensors with different resolution, viewing angle, or capturing time. A review on recent advances in SAR and optical registration can be found in [127]. Accurate image registration is a crucial preprocessing step for various remote sensing applications, such as CD, image mosaicking, fusion, and 3-D reconstruction. Accordingly, recent studies investigated NSDL methods on different types of remote sensing data, such as Lidar point cloud, MS, RGB, and multimodal (e.g., SAR-optical) registration. Papadomanolaki et al. [128] proposed an unsupervised framework that utilizes a fully CNN model to iteratively reduce the distance between two sources of VHR images until they align. A GAN-based model for SAR-optical registration was presented in [129]. First, a set of synthetic SAR images were generated from their corresponding optical image. Then, the synthesized SAR images and real SAR images were used as training data for SAR-optical matching.

The presence of noise is inevitable in the remotely sensed data. While the power and amount of noise varies in different sensors, HS and SAR images generally show higher level of noise. Various studies utilized NSDL methods to address this issue with least dependency on annotated training data. An SAR despeckling framework, namely, SAR2SAR, was presented in [130]. The proposed method was based on the assumption that the denoised image should be close to the average of the images from the same area captured at different times. Accordingly, the method consisted of simulated training and then finetuning from the Sentinel-1 archive images. In another study, a self-SL speckle noise reduction from SAR images can be found in [131]. In this study, the authors aimed to model SAR noise by extending noise2noise method, where a pair of subsampled noisy images were generated from the input noisy image and the output denoised image was generated by optimizing a custom loss function, which calculates the pixelwise distance between the reconstructed denoised pairs. A Self-SL denoising network for HSI was presented in study [132]. In this study, Imamura et al. [132] proposed a fully CNN and trained it by manually adding noise to the HSI. The objective of the model was to generate the original image from the noisy one. In another study for HSI denoising, Faghih et al. [133] proposed an unsupervised method using a deep image prior (DIP) UNet architecture. They showed that DIP trained with Huber loss function can generate denoised HSI without needing training data.

VI. Conclusion and Prospects

This article provided a comprehensive systematic review and a meta-analysis of the applications and developments of NSDL methods in the remote sensing field. The results demonstrated the increasing trends in using NSDL methods, which is inevitable with increasing the number of satellites and data providers in the era of big EO data. In this study, we reviewed and analyzed 505 papers that have been published in well-known peer-reviewed journals since 2015. Subsequently, different features, including DL architectures, training data requirements, applications, and accuracies, were extracted from these reviewed articles. Eventually, deeper overview and literature review of NSDL methods and their applications in remote sensing was provided. The summary of the key conclusion including the remaining challenges and future directions is as follows.

1) This study confirmed that the utilization of NSDL methods is increasing in various fields of remote sensing applications. The results indicate that these methods offer various solutions for the different levels and scenarios of training data. The main reason can be the availability of a vast amount of remote sensing data from various sensors and satellites, making it nearly impossible to prepare suitable annotated data by humans. Notably, insufficient training data can increase the chance of overfitting in supervised DL models.
2) According to the meta-analysis, majority of studies were based on using VHR and RGB imagery. In addition to the rich spatial features of such data, this can be due to the high availability of VHR and HR images with the increasing development and accessibility of drones, surveillance, or even smart phones' cameras along with the availability of high-resolution satellites. This also has led to the availability of many of these type of datasets as benchmarks. However, field of remote sensing covers various sensors with wide range spatial and spectral resolutions. Recent technological advances in the high availability of remote sensing satellite imagery with very short revisit times and the advent of cloud-based processing platforms, such as Google Earth Engine, have increased the demand for remote sensing and geospatial data applications. Consequently, as the need for ground truth data is at its highest point, the number of NSDL studies on free datasets, such as Sentinel and Landsat constellations, need to be increased. In addition, there were considerably fewer studies addressing LiDAR or SAR processing using NSDL. LiDAR and SAR, as active sensors, follow different imaging principles than passive sensors. The data obtained by these sensors are often noisy and full of clutters, which makes their interpretation a cumbersome task. Consequently, preparing manual annotation data from active sensors would be a time-consuming task with a low confidence level. This emphasizes the importance and applicability of NSDL for such data for preprocessing such as denoising, despeckling, and registration and remote sensing applications such as anomaly or CD.
3) Presenting novel and diverse customized deep models have become more feasible with the growing variability of new DL architectures, such as the family of Attention mechanisms and Transformers, GANs, a variety of AE models, as well as various loss functions, such as

contrastive and adversarial losses. For instance, AEs due to their capability in automated feature extraction can increase the discriminability of complex remote sensing classes in latent subspace. Attentions showed their capability by highlighting the salient areas, especially, in solving WSL problems. Furthermore, GANs were useful for data augmentation and balancing the training data in highly imbalanced problems, such as CD. However, as also discussed in [6], improving the quality and reliability of the generative models in order to presenting diverse and accurate simulated data would be some of the remaining challenges that need to be carefully considered and addressed in future works. Furthermore, while designing a complex model for nonsupervised problems in remote sensing may be effective, it is imperative to pay attention to the efficiency and robustness of the developed model. For instance, as discussed in [39], model collapse in deep Semi-SL models on remote sensing data is a common challenge. Thus, it is essential for future studies to focus on this issue by probably providing a comprehensive analytical and theoretical perspective, as well as designing more efficient frameworks based on feature engineering, efficient architecture, or novel loss functions.

4) As shown in the meta-analysis, NSDL methods were well-received in remote sensing studies, such as classification and CD. Although these objectives are the main applications of remote sensing data, NSDL methods have high potentials to be explored in other applications. In general, there is still much more to explore in multimodal image registration, or fusion, image or point cloud denoising, HS unmixing, and time-series AD.
5) The majority of NSDL methods rely on a small number of annotated training data, and as such, the accuracy and reliability of this data is of utmost importance. Each sample can significantly affect the training process, which emphasizes the need for a careful annotation process. A recent study [134] looked into the effects of label noise on multilabel remote sensing classification. However, further research is required to provide frameworks that can remain robust in the face of data uncertainties.
6) As explainable AI is becoming more prevalent in remote sensing applications [135], [136], it can be linked to NSDL methods by opening two doors in improving the model training.
   a) Making the model simpler by removing irrelevant features and, thus, reducing the need for extra training data for high dimensional modeling. The most prominent approaches for this purpose are SHAP [137] and LIME [138].
   b) Highlighting the most influenced areas in the feature activation maps, which can improve the saliency-oriented training methods.
7) Currently, various sources of data are available, which can be leveraged for NSDL model training. Thanks to the increasing availability of free satellite datasets, a massive amount of unlabeled EO data are currently available. Another source of data can be open data with geo-tags from the internet and social media. While supervised methods do not effectively exploit this abundant pool of unlabeled data, the main challenge for NSDL modeling can be filling the domain gap in the presence of a pool of available multimodal data. Thus, domain adaptation from another time or spatial or electromagnetic domain for TL, and unsupervised representation learning and then finetuning in a Self-SL framework will play important roles in the majority of future developments.

References

[1] Y. Bouslihim, M. H. Kharrou, A. Miftah, T. Attou, L. Bouchaou, and A. Chehbouni, "Comparing pan-sharpened landsat-9 and Sentinel-2 for land-use classification using machine learning classifiers," *J. Geovisualization Spatial Anal.*, vol. 6, no. 2, pp. 1–17, Dec. 2022, doi: 10.1007/S41651-022-00130-0/METRICS.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.

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

[4] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[5] H. Ghanbari, M. Mahdianpari, S. Homayouni, and F. Mohammadimanesh, "A meta-analysis of convolutional neural networks for remote sensing applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3602–3613, 2021.

[6] S. Jozdani, D. Chen, D. Pouliot, and B. Alan Johnson, "A review and meta-analysis of generative adversarial networks and their applications in remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, Apr. 2022, Art. no. 102734, doi: 10.1016/J.JAG.2022.102734.

[7] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review," *Remote Sens. (Basel)*, vol. 13, no. 15, 2021, Art. no. 2965.

[8] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, 2019.

[9] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[10] T. Hoeser and C. Kuenzer, "Object detection and image segmentation with deep learning on earth observation data: A review-part I: Evolution and recent trends," *Remote Sens. (Basel)*, vol. 12, no. 10, 2020, Art. no. 1667.

[11] T. Hoeser, F. Bachofer, and C. Kuenzer, "Object detection and image segmentation with deep learning on Earth observation data: A review—Part II: Applications," *Remote Sens. (Basel)*, vol. 12, no. 18, 2020, Art. no. 3053.

[12] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385–126400, 2020.

[13] G. Dong, G. Liao, H. Liu, and G. Kuang, "A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 44–68, Sep. 2018.

[14] S. S. Sawant and M. Prabukumar, "A review on graph-based semi-supervised learning methods for hyperspectral image classification," *Egyptian J. Remote Sens. Space Sci.*, vol. 23, no. 2, pp. 243–248, 2020.

[15] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, 2018.

[16] Y. Chong, Y. Ding, Q. Yan, and S. Pan, "Graph-based semi-supervised learning: A review," *Neurocomputing*, vol. 408, pp. 216–230, 2020.

[17] K. Ohri and M. Kumar, "Review on self-supervised image recognition using deep neural networks," *Knowl. Based Syst.*, vol. 224, 2021, Art. no. 107090.

[18] L. Chum, A. Subramanian, V. N. Balasubramanian, and C. V. Jawahar, "Beyond supervised learning: A computer vision perspective," *J. Indian Inst. Sci.*, vol. 99, no. 2, pp. 177–199, 2019.

[19] J. Peng and Y. Wang, "Medical image segmentation with limited supervision: A review of deep network models," *IEEE Access*, vol. 9, pp. 36827–36851, 2021.

[20] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, no. 7, 2009, Art. no. e1000097.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[22] M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, pp. 40, 2023.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[24] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[25] D. Ma, P. Tang, and L. Zhao, "SiftingGAN: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1046–1050, Jul. 2019.

[26] K. Qin et al., "Hapke data augmentation for deep learning-based hyperspectral data analysis with limited samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 886–890, May 2021.

[27] B. Hosseiny, H. Rastiveis, and S. Homayouni, "An automated framework for plant detection based on deep simulated learning from drone imagery," *Remote Sens. (Basel)*, vol. 12, no. 21, 2020, Art. no. 3521.

[28] A. Singh and L. Bruzzone, "SIGAN: Spectral index generative adversarial network for data augmentation in multispectral remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6003305.

[29] Y. Yuan, Y. Wu, Y. Fu, Y. Wu, L. Zhang, and Y. Jiang, "An advanced SAR image despeckling method by Bernoulli-sampling-based self-supervised deep learning," *Remote Sens.*, vol. 13, no. 18, Sep. 2021, Art. no. 3636, doi: 10.3390/RS13183636.

[30] G. Csurka, R. Volpi, and B. Chidlovskii, "Semantic image segmentation: Two decades of research," *Found. Trends Comput. Graph. Vis.*, vol. 14, no. 1-2, pp. 1–162, 2022.

[31] J. Zheng et al., "A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S^2M^2T DA) scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609213.

[32] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[33] C. Ma, X. Mu, P. Zhao, and X. Yan, "Meta-learning based on parameter transfer for few-shot classification of remote sensing scenes," *Remote Sens. Lett.*, vol. 12, no. 6, pp. 531–541, 2021.

[34] X. Li, Z. Sun, J.-H. Xue, and Z. Ma, "A concise review of recent few-shot meta-learning methods," *Neurocomputing*, vol. 456, pp. 463–468, 2021.

[35] M. Huisman, J. N. van Rijn, and A. Plaat, "A survey of deep meta-learning," *Artif. Intell. Rev.*, vol. 54, pp. 4483–4541, 2021.

[36] Y. Xiong, K. Xu, Y. Dou, Y. Zhao, and Z. Gao, "WRMatch: Improving FixMatch with weighted nuclear-norm regularization for few-shot remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5612214.

[37] Z. Zhao, P. Tang, L. Zhao, and Z. Zhang, "Few-shot object detection of remote sensing images via two-stage fine-tuning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8021805.

[38] K. Fu, T. Zhang, Y. Zhang, Z. Wang, and X. Sun, "Few-shot SAR target classification via metalearning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 2000314.

[39] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2387–2402, 2021, doi: 10.1109/JSTARS.2021.3052869.

[40] Y. Wang, C. Albrecht, N. A. A. Braham, L. Mou, and X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, 2022.

[41] K. Fu et al., "WSF-NET: Weakly supervised feature-fusion network for binary segmentation in remote sensing image," *Remote Sens. (Basel)*, vol. 10, no. 12, 2018, p. 1970.

[42] R. Qiao, A. Ghodsi, H. Wu, Y. Chang, and C. Wang, "Simple weakly supervised deep learning pipeline for detecting individual red-attacked trees in VHR remote sensing images," *Remote Sens. Lett.*, vol. 11, no. 7, pp. 650–658, 2020.

[43] Y. Wei and S. Ji, "Scribble-based weakly supervised deep learning for road surface extraction from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602312.

[44] R. Lian and L. Huang, "Weakly supervised road segmentation in high-resolution remote sensing images using point annotations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4501013.

[45] B. Aygunes, R. G. Cinbis, and S. Aksoy, "Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 176, pp. 262–274, 2021.

[46] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112045.

[47] E. A. Kringle, E. C. Knutson, C. Engstrom, and L. Terhorst, "Iterative processes: A review of semi-supervised machine learning in rehabilitation science," *Disabil. Rehabil. Assistive Technol.*, vol. 15, no. 5, pp. 515–520, 2020.

[48] S.-J. Liu, H. Luo, and Q. Shi, "Active ensemble deep learning for polarimetric synthetic aperture radar image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1580–1584, Sep. 2021.

[49] J. M. Haut, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.

[50] B. Liu et al., "Active deep densely connected convolutional network for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 42, no. 15, pp. 5905–5924, 2021.

[51] P. Liu, H. Zhang, and K. B. Eom, "Active deep learning for classification of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 712–724, Feb. 2017.

[52] S. Yang, L. Gu, X. Li, F. Gao, and T. Jiang, "Fully automated classification method for crops based on spatiotemporal deep-learning fusion technology," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5405016.

[53] H. Dong, W. Ma, L. Jiao, F. Liu, and L. Li, "A multiscale self-attention deep clustering for change detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5207016.

[54] S. Saha and X. X. Zhu, "Patch-level unsupervised planetary change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6504405.

[55] T. Zhan, M. Gong, X. Jiang, and M. Zhang, "Unsupervised scale-driven change detection with deep spatial–spectral features for VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5653–5665, Aug. 2020.

[56] S. Saha, F. Bovolo, and L. Bruzzone, "Unsupervised deep change vector analysis for multiple-change detection in VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3677–3693, Jun. 2019.

[57] T. Kurihana, E. Moyer, R. Willett, D. Gilton, and I. Foster, "Data-driven cloud clustering via a rotationally invariant autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4103325.

[58] B. Hosseiny and R. Shah-Hosseini, "A hyperspectral anomaly detection framework based on segmentation and convolutional neural network algorithms," *Int. J. Remote Sens.*, vol. 41, no. 18, pp. 6946–6975, 2020.

[59] N. Merrill and A. Eskandarian, "Modified autoencoder training and scoring for robust unsupervised anomaly detection in deep learning," *IEEE Access*, vol. 8, pp. 101824–101833, 2020.

[60] P. Zhang, M. Gong, H. Zhang, J. Liu, and Y. Ban, "Unsupervised difference representation learning for detecting multiple types of changes in multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2277–2289, Apr. 2019.

[61] W. Li, H. Chen, and Z. Shi, "Semantic segmentation of remote sensing images with self-supervised multi-task representation learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6438–6450, 2021.

[62] K. Heidler et al., "Self-supervised audiovisual representation learning for remote sensing data," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 116, pp. 103–130, 2023.

[63] J. Xie, L. Fang, B. Zhang, J. Chanussot, and S. Li, "Super resolution guided deep network for land cover classification from remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5611812.

[64] C. Ren, X. Wang, J. Gao, X. Zhou, and H. Chen, "Unsupervised change detection in satellite images with generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10047–10061, Dec. 2021.

[65] N. Zhang, Y. Wang, X. Zhang, D. Xu, and X. Wang, "An unsupervised remote sensing single-image super-resolution method based on generative adversarial network," *IEEE Access*, vol. 8, pp. 29027–29039, 2020.

[66] M. Wang, X. Zhang, X. Niu, F. Wang, and X. Zhang, "Scene classification of high-resolution remotely sensed image based on ResNet," *J. Geovisualization Spatial Anal.*, vol. 3, no. 2, pp. 1–9, Dec. 2019, doi: 10.1007/S41651-019-0039-9/METRICS.

[67] F. Hu, G. S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, Nov. 2015, doi: 10.3390/RS71114680.

[68] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.

[69] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5901–5904, doi: 10.1109/IGARSS.2019.8900532.

[70] G. Sumbul et al., "BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 174–180, Sep. 2021, doi: 10.1109/MGRS.2021.3089174.

[71] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.

[72] A. Jamali, M. Mahdianpari, F. Mohammadimanesh, B. Brisco, and B. Salehi, "A synergic use of Sentinel-1 and Sentinel-2 imagery for complex wetland classification using generative adversarial network (GAN) scheme," *Water (Basel)*, vol. 13, no. 24, 2021, Art. no. 3601.

[73] Z. Yuan, W. Huang, C. Tang, A. Yang, and X. Luo, "Graph-based embedding smoothing network for few-shot scene classification of remote sensing images," *Remote Sens. (Basel)*, vol. 14, no. 5, 2022, Art. no. 1161.

[74] Z. Cao, X. Li, J. Jianfeng, and L. Zhao, "3D convolutional siamese network for few-shot hyperspectral classification," *J. Appl. Remote Sens.*, vol. 14, no. 4, 2020, Art. no. 48504.

[75] Y. Li, Z. Shao, X. Huang, B. Cai, and S. Peng, "Meta-FSEO: A meta-learning fast adaptation with self-supervised embedding optimization for few-shot remote sensing scene classification," *Remote Sens. (Basel)*, vol. 13, no. 14, 2021, Art. no. 2776.

[76] G. Lenczner, A. Chan-Hon-Tong, B. Le Saux, N. Luminari, and G. Le Besnerais, "DIAL: Deep interactive and active learning for semantic segmentation in remote sensing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3376–3389, 2022.

[77] J. Chen, W. Xu, Y. Yu, C. Peng, and W. Gong, "Reliable label-supervised pixel attention mechanism for weakly supervised building segmentation in UAV imagery," *Remote Sens. (Basel)*, vol. 14, no. 13, 2022, Art. no. 3196.

[78] M. Lu, L. Fang, M. Li, B. Zhang, Y. Zhang, and P. Ghamisi, "NFANet: A novel method for weakly supervised water extraction from high-resolution remote-sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5617114.

[79] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sens. (Basel)*, vol. 12, no. 2, 2020, Art. no. 207.

[80] D. Guo, Y. Xia, and X. Luo, "Self-supervised GANs with similarity loss for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2508–2521, 2021.

[81] Y. Yuan and L. Lin, "Self-supervised pretraining of transformers for satellite image time series classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 474–487, 2021.

[82] Y. Yuan, L. Lin, Q. Liu, R. Hang, and Z.-G. Zhou, "SITS-former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 106, 2022, Art. no. 102651.

[83] K. Li, Y. Qin, Q. Ling, Y. Wang, Z. Lin, and W. An, "Self-supervised deep subspace clustering for hyperspectral images with adaptive self-expressive coefficient matrix initialization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3215–3227, 2021.

[84] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, "Self-supervised learning with adaptive distillation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501813.

[85] B. Ren, Y. Zhao, B. Hou, J. Chanussot, and L. Jiao, "A mutual information-based self-supervised learning model for PolSAR land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9224–9237, Nov. 2021.

[86] Z. Xue, X. Yu, A. Yu, B. Liu, P. Zhang, and S. Wu, "Self-supervised feature learning for multimodal remote sensing image land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533815.

[87] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2438–2442, Dec. 2015.

[88] K. R. Shahi, P. Ghamisi, B. Rasti, P. Scheunders, and R. Gloaguen, "Unsupervised data fusion with deeper perspective: A novel multisensor deep clustering algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 284–296, 2022.

[89] A. Radman, R. Shah-Hosseini, and S. Homayouni, "An unsupervised saliency-guided deep convolutional neural network for accurate burn mapping from sentinel-1 SAR data," *Remote Sens. (Basel)*, vol. 15, no. 5, 2023, Art. no. 1184.

[90] Z. Chen, T. Zhang, and C. Ouyang, "End-to-end airplane detection using transfer learning in remote sensing images," *Remote Sens. (Basel)*, vol. 10, no. 1, 2018, Art. no. 139.

[91] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, "Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8534–8545, Nov. 2019.

[92] H. Chen, L. Zhang, J. Ma, and J. Zhang, "Target heat-map network: An end-to-end deep network for target detection in remote sensing images," *Neurocomputing*, vol. 331, pp. 375–387, 2019.

[93] Q. Li, Y. Chen, and Y. Zeng, "Transformer with transfer CNN for remote-sensing-image object detection," *Remote Sens. (Basel)*, vol. 14, no. 4, 2022, Art. no. 984.

[94] G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova, "Transferring deep learning models for cloud detection between landsat-8 and proba-V," *ISPRS J. Photogrammetry Remote Sens.*, vol. 160, pp. 1–17, 2020.

[95] D. Zhu et al., "Diverse sample generation with multi-branch conditional generative adversarial network for remote sensing objects detection," *Neurocomputing*, vol. 381, pp. 40–51, 2020.

[96] H. Liu, Z. Hu, J. Yu, and S. Gao, "Conditional GAN-based remote sensing target image generation method," *Int. J. Adv. Netw., Monit. Controls*, vol. 5, no. 4, pp. 66–74, 2020.

[97] J. Oh and M. Kim, "PeaceGAN: A GAN-based multi-task learning method for SAR target image generation with a pose estimator and an auxiliary classifier," *Remote Sens. (Basel)*, vol. 13, no. 19, 2021, Art. no. 3939.

[98] Y. Zhang, H. Sun, J. Zuo, H. Wang, G. Xu, and X. Sun, "Aircraft type recognition in remote sensing images based on feature learning with conditional generative adversarial networks," *Remote Sens. (Basel)*, vol. 10, no. 7, 2018, Art. no. 1123.

[99] K. Zheng, M. Wei, G. Sun, B. Anas, and Y. Li, "Using vehicle synthesis generative adversarial networks to improve vehicle detection in remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 9, 2019, Art. no. 390.

[100] G. Xu, X. Zhu, and N. Tapper, "Using convolutional neural networks incorporating hierarchical active learning for target-searching in large-scale remote sensing images," *Int. J. Remote Sens.*, vol. 41, no. 11, pp. 4057–4079, 2020.

[101] L. Xia, X. Zhang, J. Zhang, H. Yang, and T. Chen, "Building extraction from very-high-resolution remote sensing images using semi-supervised semantic edge detection," *Remote Sens. (Basel)*, vol. 13, no. 11, 2021, Art. no. 2187.

[102] L. Liao, L. Du, and Y. Guo, "Semi-supervised SAR target detection based on an improved faster R-CNN," *Remote Sens. (Basel)*, vol. 14, no. 1, 2021, Art. no. 143.

[103] X. Yan, L. Shen, J. Wang, X. Deng, and Z. Li, "MSG-SR-Net: A weakly supervised network integrating multiscale generation and superpixel refinement for building extraction from high-resolution remotely sensed imageries," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1012–1023, 2022.

[104] W. Diao, X. Sun, F. Dou, M. Yan, H. Wang, and K. Fu, "Object recognition in remote sensing images using sparse deep belief networks," *Remote Sens. Lett.*, vol. 6, no. 10, pp. 745–754, 2015.

[105] Z. Shao, J. Deng, L. Wang, Y. Fan, N. S. Sumari, and Q. Cheng, "Fuzzy autoencode based cloud detection for remote sensing imagery," *Remote Sens. (Basel)*, vol. 9, no. 4, 2017, Art. no. 311.

[106] C. Xiao, R. Qin, and X. Huang, "Treetop detection using convolutional neural networks trained through automatically generated pseudo labels," *Int. J. Remote Sens.*, vol. 41, no. 8, pp. 3010–3030, 2020.

[107] Z. Wen, Z. Liu, S. Zhang, and Q. Pan, "Rotation awareness based self-supervised learning for SAR target recognition with limited training samples," *IEEE Trans. Image Process.*, vol. 30, pp. 7266–7279, 2021.

[108] Y. Xu, H. Sun, J. Chen, L. Lei, K. Ji, and G. Kuang, "Adversarial self-supervised learning for robust SAR target recognition," *Remote Sens. (Basel)*, vol. 13, no. 20, 2021, Art. no. 4158.

[109] C. Yao, Y. Yuan, and Z. Jiang, "Self-supervised spectral matching network for hyperspectral target detection," *IEEE Int. Geosci. Remote Sens. Sympos.*, 2021.

[110] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, "Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022.

[111] W. Rao, Y. Qu, L. Gao, X. Sun, Y. Wu, and B. Zhang, "Transferable network with Siamese architecture for anomaly detection in hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 106, 2022, Art. no. 102669.
[112] L. T. Luppino et al., "Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2022.3172183.
[113] H. Lyu et al., "Long-term annual mapping of four cities on different continents by applying a deep information learning method to landsat data," *Remote Sens. (Basel)*, vol. 10, no. 3, 2018, Art. no. 471.
[114] P. J. Soto Vega et al., "An unsupervised domain adaptation approach for change detection and its application to deforestation mapping in tropical biomes," *ISPRS J. Photogrammetry Remote Sens.*, vol. 181, pp. 113–128, Nov. 2021, doi: 10.1016/J.ISPRSJPRS.2021.08.026.
[115] J. Yan, L. Wang, H. He, D. Liang, W. Song, and W. Han, "Large-area land-cover changes monitoring with time-series remote sensing images using transferable deep models," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
[116] T. Liu et al., "Building change detection for vhr remote sensing images via local-global pyramid network and cross-task transfer learning strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4704817.
[117] H. Fang, P. Du, and X. Wang, "A novel unsupervised binary change detection method for VHR optical remote sensing imagery over urban areas," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 108, 2022, Art. no. 102749.
[118] P. J. Soto, G. A. Costa, R. Q. Feitosa, M. X. Ortega, J. D. Bermudez, and J. N. Turnes, "Domain-adversarial neural networks for deforestation detection in tropical forests," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2504505.
[119] F. Hao, Z.-F. Ma, H.-P. Tian, H. Wang, and D. Wu, "Semi-supervised label propagation for multi-source remote sensing image change detection," *Comput. Geosci.*, vol. 170, 2023, Art. no. 105249.
[120] C. Sun, J. Wu, H. Chen, and C. Du, "SemiSANet: A semi-supervised high-resolution remote sensing image change detection model using siamese networks with graph attention," *Remote Sens. (Basel)*, vol. 14, no. 12, 2022, Art. no. 2801.
[121] H. Dong, W. Ma, Y. Wu, J. Zhang, and L. Jiao, "Self-supervised representation learning for remote sensing image change detection based on temporal prediction," *Remote Sens. (Basel)*, vol. 12, no. 11, 2020, Art. no. 1868.
[122] Y. Chen and L. Bruzzone, "A self-supervised approach to pixel-level change detection in bi-temporal RS images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.
[123] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6792–6810, Nov. 2018.
[124] H. V. Nguyen, M. O. Ulfarsson, J. R. Sveinsson, and M. Dalla Mura, "Sentinel-2 sharpening using a single unsupervised convolutional neural network with MTF-based degradation model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6882–6896, 2021.
[125] J. Ma, L. Zhang, and J. Zhang, "SD-GAN: Saliency-discriminated GAN for remote sensing image superresolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1973–1977, Nov. 2020.
[126] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, 2023.
[127] O. Sommervold, M. Gazzea, and R. Arghandeh, "A survey on SAR and optical satellite image registration," *Remote Sens. (Basel)*, vol. 15, no. 3, 2023, Art. no. 850.
[128] M. Papadomanolaki, S. Christodoulidis, K. Karantzalos, and M. Vakalopoulou, "Unsupervised multistep deformable registration of remote sensing imagery based on deep learning," *Remote Sens. (Basel)*, vol. 13, no. 7, 2021, Art. no. 1294.
[129] L. H. Hughes, M. Schmitt, and X. X. Zhu, "Mining hard negative samples for SAR-optical image matching using generative adversarial networks," *Remote Sens. (Basel)*, vol. 10, no. 10, 2018, Art. no. 1552.
[130] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: A semi-supervised despeckling algorithm for SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4321–4329, 2021.
[131] G. Zhang, Z. Li, X. Li, and S. Liu, "Self-supervised despeckling algorithm with an enhanced U-net for synthetic aperture radar images," *Remote Sens. (Basel)*, vol. 13, no. 21, 2021, Art. no. 4383.
[132] R. Imamura, T. Itasaka, and M. Okuda, "Zero-shot hyperspectral image denoising with separable image prior," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, *arXiv:1907.00651*.
[133] K. F. Niresi and C.-Y. Chi, "Unsupervised hyperspectral denoising based on deep image prior and least favorable distribution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5967–5983, 2022.
[134] T. Burgert, M. Ravanbakhsh, and B. Demir, "On the effects of different types of label noise in multi-label remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5413713.
[135] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 103, 2021, Art. no. 102520.
[136] B. Hosseiny, A. M. Abdi, and S. Jamali, "Urban land use and land cover classification with interpretable machine learning–A case study using Sentinel-2 and auxiliary data," *Remote Sens. Appl.*, vol. 28, 2022, Art. no. 100843.
[137] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
[138] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1135–1144.
[139] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. ACM Int. Symp. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279, doi: 10.1145/1869790.1869829.
[140] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[141] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
[142] N. Weir et al., "Spacenet mvoi: A multi-view overhead imagery dataset," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 992–1001.
[143] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2115–2118.
[144] J. López-Fandiño, A. S. Garea, D. B. Heras, and F. Argüello, "Stacked autoencoders for multiclass change detection in hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1906–1909.
[145] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
[146] J. R. Diemunsch and J. Wissinger, "Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: Search technology for a robust ATR," *Proc. SPIE*, vol. 3370, pp. 481–492, 1998.
[147] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS–A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Annu. Photogrammetry, Remote Sens. Spatial Inform. Sci.*, vol. 4, pp. 153–160, 2019.
[148] P. Ebel, Y. Xu, M. Schmitt, and X. X. Zhu, "SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5222414.

Benyamin Hosseiny received the B.Sc. degree in geomatics engineering and the master's degree in remote sensing, in 2016 and 2018, respectively, from the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, where he is currently working toward the Ph.D. degree in remote sensing.

His Ph.D. studies primarily focus on the development of a multiple-input multiple output ground-based synthetic aperture radar for structural monitoring. Throughout his research, he has developed interferometric and beamforming algorithms to measure displacement from radar signals and enable multidimensional imaging.

His research interests include radar and synthetic aperture radar signal processing, geosensors, pattern recognition, machine (deep) learning applications for remote sensing, and earth observation. Over the past three years, he has authored and coauthored more than ten peer-reviewed papers and actively served as a Reviewer in the well-known journals in these fields. In 2021, he served as a Visiting Researcher with the Department of Technology and Society, Lund University, Lund, Sweden, where he studied on the explainable machine learning for remote sensing image classification.

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

In 2019, he was an Ocean Frontier Institute (OFI) Postdoctoral Fellow with Memorial University and C-CORE. He is currently a Remote Sensing Technical Lead with C-CORE and a Cross Appointed Professor with the Faculty of Engineering and Applied Science, Memorial University. He has authored or coauthored more than 100 publications, including peer-reviewed journal articles, conference papers, books, and book chapters. His research interests include remote sensing and image analysis, with a special focus on PolSAR image processing, multisensor data classification, machine learning, geo Big Data, and deep learning.

Dr. Mahdianpari is also an Editorial Team Member for the IEEE Geoscience and Remote Sensing Letters, *Remote Sensing Journal*, and the *Canadian Journal of Remote Sensing*. He was a recipient of the Research and Development Corporation Ocean Industries Student Research Award, organized by Newfoundland Industry and Innovation Center, amongst more than 400 submissions, in 2016, the T. David Collett Best Industry Paper Award organized by IEEE in 2016, the Com-Adv Devices, Inc., Scholarship for Innovation, Creativity, and Entrepreneurship from the Memorial University in 2017, the Artificial Intelligence for Earth Grant organized by Microsoft in 2018, and the Graduate Academic Excellence Award organized by Memorial University in 2019, also the Best Industry Paper Award organized by IEEE Newfoundland Electrical and Computer Engineering Conference, in 2020.

Mohammadali Hemati (Student Member, IEEE) received the B.Sc. degree in surveying and geomatics engineering from Shahid Rajaee Teacher Training University, Tehran, Iran, in 2019, and the M.Sc. degree in remote sensing engineering from the School of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, in 2021. He is currently working toward the Ph.D. degree in electrical engineering in the Department of Engineering and Applied Science, Memorial University, St. John's, NL, Canada.

His research interests include remote sensing and Earth observation, with a specific focus on environmental applications and climate change. His expertise extends to machine learning and deep learning, image processing and classification, analysis of geo Big Data for large-scale mapping, synthetic aperture radar, hyperspectral and multispectral data analysis, as well as time-series analysis and change detection.

Ali Radman received the B.Sc. degree in geomatics engineering and the master's degree in remote sensing from the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in electrical engineering in Memorial University of Newfoundland, St. John's, NL, Canada.

His research interests include remote sensing, environmental monitoring, multispectral and SAR remote sensing, deep learning applications for remote sensing, and Earth observation.

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

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

Dr. Mohammadimanesh was the recipient of several awards, including the EmeraGraduate Scholarship (2016–2019), CIG-NLAward (2018), and the IEEE NECEC Best Industry Paper Award (2016 and 2020).

Jocelyn Chanussot (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory, Delegation Generale de l'Armement (DGA - French National Defense Department). Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. He is conducting his research with Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, data fusion, and machine learning in remote sensing. He has been a Visiting Scholar with Stanford University, Stanford, CA, USA, KTH Royal Institute of Technology, Stockholm, Sweden, and National University of Singapore, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. In 2015–2017, he was a Visiting Professor with the University of California, Los Angeles. He holds the AXA Chair in remote sensing and is an Adjunct Professor with the Chinese Academy of Sciences, Aerospace Information Research Institute, Beijing, China.

Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French Chapter in 2007–2010, which received the 2010 IEEE GRSS Chapter Excellence Award. He was the recipient of multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia in 2017–2019. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing. He was the Chair in 2009–2011 and the Cochair of the GRSS Data Fusion Technical Committee in 2005–2008. He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society in 2006–2008 and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He is an Associate Editor for IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Image Processing, and the Proceedings of the IEEE. He was the Editor-in-Chief for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing in 2011–2015. In 2014, he served as a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE. He is a member of the Institut Universitaire de France in 2012–2017 and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, since 2018).