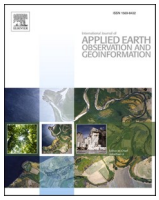
International Journal of Applied Earth Observation and Geoinformation 125 (2023) 103569

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Ten deep learning techniques to address small data problems with remote sensing

Anastasiia Safonova [a,*], Gohar Ghazaryan [a,d], Stefan Stiller [a,b], Magdalena Main-Knorn [a], Claas Nendel [a,c], Masahiro Ryo [a,b]

[a] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany
[b] Environment and Natural Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany
[c] Institute of Biochemistry and Biology, University of Potsdam, Germany
[d] Department of Geography, Humboldt University of Berlin, Berlin, Germany

ARTICLE INFO

Keywords:
Small data problems
Remote sensing
Deep learning
Transfer learning
Few-shot learning
Zero-shot learning
Self-supervised learning

ABSTRACT

Researchers and engineers have increasingly used Deep Learning (DL) for a variety of Remote Sensing (RS) tasks. However, data from local observations or via ground truth is often quite limited for training DL models, especially when these models represent key socio-environmental problems, such as the monitoring of extreme, destructive climate events, biodiversity, and sudden changes in ecosystem states. Such cases, also known as *small data problems*, pose significant methodological challenges. This review summarises these challenges in the RS domain and the possibility of using emerging DL techniques to overcome them. We show that the small data problem is a common challenge across disciplines and scales that results in poor model generalisability and transferability. We then introduce an overview of ten promising DL techniques: transfer learning, self-supervised learning, semi-supervised learning, few-shot learning, zero-shot learning, active learning, weakly supervised learning, multitask learning, process-aware learning, and ensemble learning; we also include a validation technique known as spatial k-fold cross validation. Our particular contribution was to develop a flowchart that helps DL users select which technique to use given by answering a few questions. We hope that our review article facilitate DL applications to tackle societally important environmental problems with limited reference data.

1. Introduction

Over the last decade, Artificial Intelligence (AI) technologies, especially Machine Learning (ML) and Deep Learning (DL), have been increasingly used for understanding and predicting human-environment interactions (LeCun et al., 2015). ML is a subset of AI that implements algorithms which use data to learn how to perform a specific task without being explicitly programmed. DL is a subset of ML that focuses on training deep neural networks capable of implicit feature extraction from unstructured data, such as images, text, and sound (Chai et al., 2021; Lauriola et al., 2022; Sztahó et al., 2021). Scientists have actively employed DL for image processing and data analysis, recently providing innovative solutions in the field of Remote Sensing (RS) to detect and classify objects on Earth. This study defines RS as the use of satellite and aircraft-based sensors.

The expanding field of RS provides an abundance of data streams from numerous sources. This, combined with the growing array of available data products, delivers a wide range of data that is useful for addressing various problems. Among them, Landsat, has been operational since the early 1970 s and provides a unique long-term record of satellite imagery with a 30-metre spatial resolution. The Copernicus programme's Sentinel-2 system generates data with a 10-metre spatial resolution, offering a balance between spatial detail and data continuity as well as radar imagery based on the Sentinel-1 mission. One recently launched hyperspectral mission, the Environmental Mapping and Analysis Program (EnMap), stands out with over 200 spectral bands and a 30-metre spatial resolution; this offers unique opportunities for researchers to map ecosystems and their changes in detail. In addition, commercial platforms, such as SkySat, provide extremely high-resolution data with a spatial resolution of less than one metre (Fruth

Abbreviations: RS, Remote Sensing; AI, Artificial Intelligence; ML, Machine Learning; DL, Deep Learning; CNN, Convolutional Neural Network; TL, Transfer Learning; NLP, Natural Language Processing; HIS, Hyperspectral Imaging; UAV, Unmanned Aerial Vehicle.

* Corresponding author.
E-mail address: anastasiia.safonova@zalf.de (A. Safonova).

https://doi.org/10.1016/j.jag.2023.103569
Received 15 September 2023; Received in revised form 1 November 2023; Accepted 12 November 2023
Available online 18 November 2023

et al., 2018; Murthy et al., 2014). Together, these diverse RS platforms contribute to a more comprehensive understanding of the Earth's surface across different scales and domains (Spoto et al., 2012). For this reason, these RS products are widely used to study local and regional environmental problems, including agricultural productivity (Sawada et al., 2020; Taiwo et al., 2023), the water quality of lakes and ponds (Bhateria and Jain, 2016), the ecological patterns of forests and grasslands (Zhu et al., 2023), and damage to natural, cultivated, and inhabited land through extreme weather events. Since the spatial and temporal resolution of RS products is likely to continue increasing, DL applications are expected to become even more popular for solving fine-scale local issues where each local site has its own unique conditions and context (Bai et al., 2022b; Kattenborn et al., 2021; Ma et al., 2019a).

Since DL algorithms have fewer inductive biases but larger parameter counts than conventional ML algorithms, DL models normally require a large amount of data for training (Adugna et al., 2022; Akar and Güngör, 2012; Fang et al., 2021; Sharma et al., 2013; Thanh Noi and Kappas, 2018). DL methods usually learn from raw data and skip manual feature engineering steps; this means that human efforts are not needed to quantitatively measure some attributes from the data. For example, DL algorithms can learn from image data directly, instead of using the extracted shape and size of an object in an image. When sufficient data is available, DL methods can automatically extract the meaningful features from low to high levels for prediction (Zhang et al., 2019a). However, although raw data of common events is generally abundant, the lack of sufficient labelling information makes the collection and preparation of a large reference dataset (Russakovsky et al., 2015) a persistent challenge for many RS applications. Moreover, certain scenarios also lack available reference data. For instance, biodiversity monitoring needs a large number of human observers well trained in taxonomic classification, which often prevents observation campaigns from generating datasets large enough for sound DL applications. Furthermore, anomaly events such as climate extremes and disease outbreaks are too rare for researchers to acquire sufficient data coverage. Their sample size is often as small as n = 1–300, which is usually insufficient for DL application (Kokol et al., 2022).

The gap between the large data availability of RS imagery and the small data (Brigato and Iocchi, 2020) availability of several important real-world environmental problems (referred to as the "small data problem") is a very common challenge. It is hard to acquire the ground-truth response labels associated with the input features. This makes sense, because the goal of most of these studies is to develop a model designed to predict a specific response variable from the various observed input features. However, traditional DL training methodologies require a large initial set of labelled data to train predictive models. It is increasingly clear that this is an emerging problem for AI, and researchers have proposed several novel DL techniques that require less labelled data (e.g., transfer learning and self-supervised learning). However, to the best of our knowledge, there is no review article that offers an overview of these techniques and their applications in the RS domain.

In this review, we show that the small data problem in the RS domain is a common technical challenge (particularly as it relates to DL) and then offer an overview of ten promising DL techniques to address this problem with different conditions. First, we explore how the small data problem can be defined. Second, we describe a few common elements of the previous studies. Third, we present the advantages and disadvantages of using a small dataset. Last but not least, we provide an overview of ten DL techniques that can address the small data problem. Our particular contribution was the development of a flowchart that guide users to effectively identify which technique to use in their use cases just by answering a few questions.

We believe that the small data problem is a common – but still understudied – issue for recent RS applications, and therefore, this review should serve as a valuable resource for supporting DL applications in the RS domains while addressing a broader range of environmental problems where reference data is often hardly available.

2. What is the small data problem?

We argue that a dataset can be considered large (not small) when the dataset consists of $>$ 100,000 annotated samples, or when it covers the entire probability distribution in a high-dimensional space. For example, there are several free large datasets that can be used for DL: the ImageNet dataset, containing over 14 million annotated images (Russakovsky et al., 2015), the Common Objects in Context (COCO) dataset, containing 330 K images, 1.5 million object instances, and 80 object categories (Lin et al., 2015), and the OpenImages dataset, containing over 9 million images (Kuznetsova et al., 2020). These datasets can be used for training a large DL model with thousands to millions of parameters. In the RS domain, land use / land cover classification would be a typical example. In these cases, model generalisability and transferability are expected to be high. Generalisability refers to how well a DL model can make accurate predictions on new, unseen data (Habib et al., 2023; Krois et al., 2021; Shah et al., 2022), and transferability is the ability of a trained model to perform well on a task or dataset different from the one on which it was originally trained (Romão et al., 2020; Wang et al., 2019b; Zhang et al., 2020).

In contrast, data is more likely to be regarded as small (or not large enough) when the dataset consists of $<$ 1,000 annotated samples, the dataset covers the distribution poorly, or the number of samples is expected to be insufficient when using DL to find meaningful features. It is noteworthy that not only the total data volume but also class imbalance and skewed data distribution can be regarded as a part of the small data problem. This is a frequently occurring situation, but it can be a significant challenge for training deep neural networks (Adadi, 2021; Du et al., 2019). A relatively small dataset can negatively affect the performance of a DL model due to overfitting, which is when a model performs well with the training data but poorly on new, independent testing data. This therefore results in low levels of model generalisability and transferability. A common case within the RS domain (but particularly relevant) is that the data can be "extra-small", meaning that the dataset consists of just 1–10 annotated samples (e.g., historical natural disasters and disease outbreaks). The size would be sufficient for human beings to start guessing what features can uniquely describe the target, but it would not be sufficient for automated, implicit DL feature extraction.

In the DL domain, the Tiny ImageNet Dataset (also known as MicroImageNet) contains 500 images for each class (of 200 classes), indicating that DL scientists regard this level of data size as small. According to the articles we reviewed in the following section, the majority of the studies targeted classification of a few types, and many of them collected less than 500 images for each class (Blekos et al., 2020; Dyrmann et al., 2016; Freeman et al., 2019; Guirado et al., 2017; Hong-Yu et al., 2023; Li et al., 2022b; Liu et al., 2022d; Liu and Zheng, 2017; Malambo et al., 2019; Pang et al., 2020; Putra and Wijayanto, 2023; Safonova et al., 2019, 2021, 2022; Sapkota et al., 2022; Windrim et al., 2019; Zenkl et al., 2022).

There should not be a strict divide between "small" and "large" when training DL models, because the size of the dataset required may depend on various factors such as the complexity of the task and the number of features in the data. Typically, the challenges stemming from a limited amount of labelled data increase with system complexity, the rarity of observations (e.g., endangered biological species), and the coverage of geographic area. Also, classification and detection problems may need less data than regression problems. Nevertheless, a convincing theoretical argument for separating the two could be made based on whether a trained DL model exhibits a "double descent", which means that a model's performance initially improves with increased complexity, worsens, and then improves again, contradicting the traditional expectations of a bias-variance trade off (Nakkiran et al., 2019). When the dataset is not large enough, the model tends to remember all possible

case-by-case instances without generalisation (and thus overfit). However, once the dataset is large enough, the DL model starts learning a handful of general features (Elhage et al., 2022).

In RS, the "small data" problem is often associated with the challenges in collecting reference data, which is labour-intensive and expensive. Data collection is often limited to small areas and short temporal coverage. Collected data may cover the spatial variability and heterogeneity sparsely when targeting diverse environmental conditions. Moreover, when one attempts to integrate multiple datasets, data integration often faces differences in data collection protocols, completeness, consistency, and compatibility (Wu et al., 2019b). This challenge can further prevent from collecting a large amount of labelled reference data from the viewpoint of data harmonisation. There are standardised initiatives like GeoWiki (Fritz et al., 2012), Joint Experiment for Crop Assessment and Monitoring (JECAM) (Borg et al., 2018), and European-wide land survey LUCAS (Land Use/Cover Area frame Survey) that gather substantial reference data (Martino and Fritz, 2008). However, their primary concentration is on land cover and land use. Specifically, LUCAS also provides soil information (Panagos et al., 2013), but primarily over Europe. Apart from land use/cover, there is a notable data deficiency for other applications, such as crop yield and vegetation biomass estimation especially with high spatial granularity. Additionally, achieving a balanced geographic distribution of this data remains challenging as hydrometeorological, agricultural, and various other monitoring networks continue to be sparsely distributed across many regions (Sheffield et al., 2018).

The small data problem may be relevant to the "small n, large p ($n < p$)" problem in statistics, where the sample size n is much smaller than the number of parameters p (also known as the "short, fat data" problem). As a rule of thumb, each parameter can be reasonably estimated with $n = 5$–10. According to this logic, it is then possible to estimate how many samples might be needed for a given DL model. Even one of the simplest convolutional neural network (CNN) architectures, LeNet-5 (two convolutional and three fully connected layers), still has about 60,000 parameters (Lecun et al., 1998). The most popular CNN architectures have 10 million to 100 million parameters (e.g., AlexNet, VGG, Inception, and ResNet), (Khan et al., 2020), and there is a trend towards increasing the number of parameters for achieving better performance (e.g., large language models). For instance, vision transformers (Dosovitskiy et al., 2021) have recently gained more and more traction, with some models consisting of more than 20 *billion* parameters (Dehghani et al., 2023).

3. Deep learning applications in RS with the small data problem

This section is a literature survey describing how common the small data problem is in the RS domain. The aim of this section is not to have a comprehensive overview of all relevant literature but to offer some common issues with the small data problems. Our approach may have omitted several articles that also addressed the small data problem, but the goal here was not to cover every single previous study. Instead, we have attempted to describe how broad the issue is, and such an effort may not be worthwhile, given the speed of scientific progress, where today's comprehensiveness may be far less important in the next couple of years.

Initially, we made a Web of Science Core Collection search withapp the following keyword combination in the "all fields" category: {"*remote sensing*" AND ("*deep learning*" OR "*convolutional neural*" OR "*recurrent neural*") AND ("*small data*" OR "*small sample*" OR "*limited sample*" OR "*limited data*")}.

We found 161 articles as of 18 January 2023. We first examined all titles and abstracts and discarded irrelevant articles. Moreover, we relied on snowball sampling of relevant papers from the reference lists of the literature that had not emerged via the systematic search but were relevant to the main goals of this review. This resulted in 80 additional articles. The list of articles with detailed information is available at DOI 10.13140/RG.2.2.33529.24161.

As a general publication trend over time, we found that the number of papers, as well as the spatial extent of interest, have increased over the years; most papers were related to vegetation monitoring (Fig. 1). In addition, we summarise and describe the reviewed publications in the following subsections based on five key findings:

(1) Various DL algorithms with various RS data sources have been used for a few common problems.

(2) The small data problem is a scale-dependent issue.

(3) Data augmentation and transfer learning are popular, but other techniques are rarely used.

(4) Reported model performances are suspiciously high, indicating a lack of appropriate evaluation schemes.

(5) Using a small dataset has several attractive benefits.

3.1. Various DL algorithms using different RS data sources are used for a few common problems

RS and DL have had a major impact in many areas, particularly in vegetation-related applications (49 of the 80 articles), followed by land use / land cover classification (16 articles), and vehicle detection or classification (5 articles). The majority of the studies conducted classification (44 articles), followed by segmentation (19 articles), and object detection (17 articles). The majority (86 %) of the studies had 1,000 or fewer annotated samples per class (median: 242 samples per class; mean 616). Few studies addressed a regression problem, but this does not necessarily indicate a lack of research on RS and DL for regression. Rather, it suggests that such tasks are more complex and require larger labelled datasets, as evidenced by the reliance on larger amounts of labelled data in several published studies (Osco et al., 2021; Yuan et al., 2020).

Vegetation-related applications included mapping crop type (Lange et al., 2022; Li and Stein, 2020; Odebiri et al., 2022), as well as monitoring plant health (Feng et al., 2022a; Ho et al., 2022; Safonova et al., 2019; Xue et al., 2022b; Astolfi et al., 2021) and predicting crop yields (Kim et al., 2021; Li et al., 2022a; Pang et al., 2020; Sagan et al., 2021). For instance, around 800 labelled data points were used for classification of crops using Sentinel-1 data (Zhao et al., 2019), and around 300 field data points were used for yield estimation with Planet and WorldView data from 2D and 3D CNN (Sagan et al., 2021). In addition, RS has contributed to biodiversity conservation by its use in analysing complex ecosystems, tracking habitat changes, and identifying plant species. For example, (Muro et al., 2022) used Sentinel-1 and −2 data in a DL model to predict plant biomass and species richness, using around 500 observations. Another example is a study by (Lange et al., 2022), where CNNs were used to map grassland use intensity.

Besides vegetation studies employing global satellite data, various other RS sources have been used. For example, one study focused on mapping urban areas used in high-resolution satellite imagery to create detailed maps of buildings and infrastructure (Li and Stein, 2020), while another study focused on monitoring changes in vegetation cover used in lower-resolution imagery from a different satellite sensor (Lioutas, 2020). In addition, a study of land cover mapping used a combination of different data sources, including aerial imagery, LiDAR, and field data (Sanlang et al., 2021; Uhl et al., 2021). Hyperspectral RS, a technology that acquires high-dimensional spectral information across hundreds of contiguous spectral bands, is another popular data source. However, obtaining manual annotations for hyperspectral data is challenging, leading to an insufficient number of labelled pixels (Chen et al., 2015; Pan et al., 2016). While DL techniques may hold potential for hyperspectral image classification (e.g., Spectral MugNet), further research is required to explore their effectiveness in scenarios with limited data (Jia et al., 2021; Pan et al., 2018).

We also found that several studies used the same dataset repeatedly by using different DL algorithms. One series of studies (Ding et al., 2022; Feng et al., 2022a; Gao et al., 2021a; Kang et al., 2019; Wang et al.,

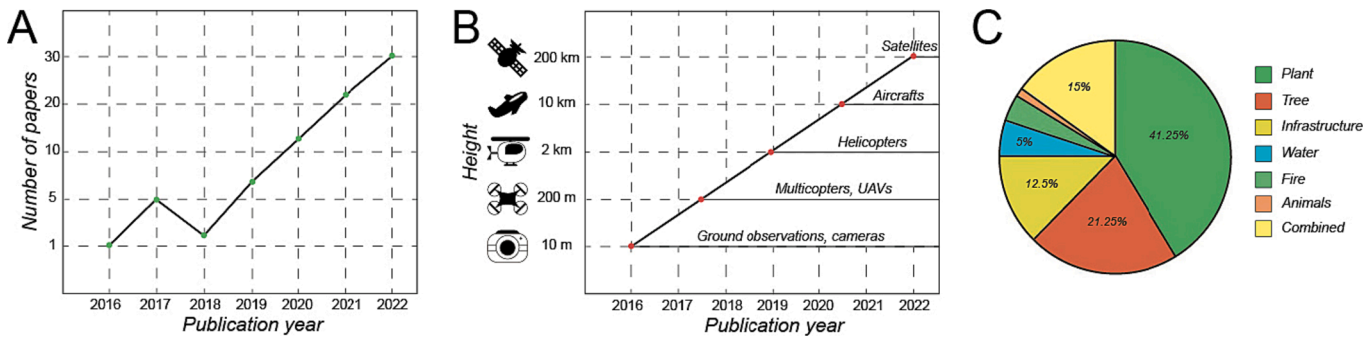


Fig. 1. Overview of publication trends with studies using remote sensing with limited annotated data and deep learning since 2016. (A) is the number of peer-reviewed articles per year, (B) is the maximum spatial coverage increasing over time, and (C) is the relative proportion of focal topics.

2020b; Wu et al., 2019a; Xu et al., 2022; Xue et al., 2022b; Zuo et al., 2020; Blekos et al., 2020) used the same hyperspectral images (HSIs) from a dataset held by the Italian University of Pavia. Some of these authors compared these HSIs with other available HSI datasets: Salinas – six papers (Ding et al., 2022; Gao et al., 2021b; Wang et al., 2020b; Wu et al., 2019a; Xue et al., 2022b; Zuo et al., 2020), Indian Pines – six papers (Ding et al., 2022; Feng et al., 2022a; Kang et al., 2019; Wang et al., 2020b; Wu et al., 2019a; Xue et al., 2022b; Zuo et al., 2020), Kennedy Space Center – two papers (Gao et al., 2021a; Xu et al., 2022), and Houston – two papers (Feng et al., 2022b; Xue et al., 2022b).

Analysing the same dataset using different approaches is a reasonable way to confirm how well a new technique may perform in comparison to previous ones, but this repetition also indicates that similar investigations of various issues is still challenging, probably due to the small data problem. Nevertheless, we found that a few studies have investigated the monitoring of extreme events, including natural fire occurrences (Kato et al., 2021; Xue et al., 2022a) and algal bloom events (Shin et al., 2022). We expect future studies to address other equally important global change events, such as conflicts, energy issues, and biodiversity problems.

3.2. The small data problem is a scale-dependent issue

The importance of spatial resolution and spatial extent in RS data is another factor that directly influences the level of detail captured and the subsequent insights that can be derived from the imagery (Kattenborn et al., 2019; Leitão et al., 2018). The small data problem becomes more pronounced when analysing high-resolution data, such as data obtained from Unmanned Aerial Vehicles (UAVs), which can offer centimetre-level granularity. This is because the fine-scale details captured in high-resolution imagery increase the variability and heterogeneity of the landscape, making it more difficult to generalise from a limited set of labelled examples. In contrast, satellite data with spatial resolution generally in the range of a few metres have different challenges related to labelled data. This relative lack of resolution can reduce the variability of the terrain and oversimplify the representation of features in the imagery.

Different sources of data also exhibit different advantages and disadvantages related to their spatial extent. Satellite data typically covers a much larger area compared to UAV data. The spatial predictions generated from satellite datacan be reliable across vast regions if there is an adequate number of labelled samples and if their spatial distribution is representative of the entire area of interest (e.g., geographically isolated areas). The distribution of labelled data plays a critical role in the performance and generalisation capabilities of DL models trained on RS data. In contrast, when working with high-resolution UAV data, the focus is more on capturing the fine-scale details and variations within a smaller area of interest. In this context, the challenge lies in accurate local measurements that reveal subtle differences.

3.3. Data augmentation and transfer learning are popular, while other techniques are rarely used

Data augmentation and transfer learning (TL) have become very common ways to improve models when applying DL techniques to RS data under conditions of small data. Our review found that 71 % of all studies employed some sort of data augmentation technique (Shorten and Khoshgoftaar, 2019) while TL was present in 14 works (Guirado et al., 2017; Reedha et al., 2022; Yu et al., 2022). Data augmentation is a technique to artificially expand a dataset by creating new samples through various transformations, such as rotation, scaling, and flipping, to improve a model's generalisability and robustness. Transfer learning is a technique where a pre-trained model, often on a large-scale dataset, is fine-tuned for a different but related task or dataset, leveraging the previously learned features to improve generalisability.

Various data augmentation techniques were used in most of the papers included in this review. The choice of data augmentation technique depends on the quantity, quality, and type of RS data. Most commonly, this method was applied to limited data from satellite imagery obtained from Landsat, WorldView, extremely high-resolution imagery, images from UAVs, and others. The most common methods of increasing RS data were: manual or automatic cropping of a large image or orthophoto image into small patches ranging from 15 × 15 pixels to 250 × 250 pixels or more, geometric image transformations (resizing, cropping, rotation, horizontal reflection, etc.), and colour transformations (changing contrast, brightness, colour, applying various noise filters, etc.). Nevertheless, as we show in the practical recommendation section, other DL techniques exist, but they are still rarely used.

3.4. Reported model performances are suspiciously high

While it was challenging to compare the studies under review in terms of model performance because they reported different metrics, we found an interesting but potentially problematic trend throughout the previous works. In essence, many studies tended to report an over-optimistic, overfit result, without testing model generalisability and transferability.

One of the most popular metrics was *Accuracy* (59 papers). The following evaluation metrics were also frequently used: *Precision* (P) in 39 papers, *Recall* (R) in 30 papers, *F1 score* (F1) in 28 papers, *Kappa coefficient* (k) in 16 papers, *Intersection over Union* (IoU) in 13 papers, *mean Average Precision* (mAP) in 10 papers, *Sensitivity* (S) in 3 papers, *p-value* by Freeman et al. (Freeman et al., 2019), *Root Mean Squared Error* by Wang et al. (Wang et al., 2022a) and Hong-Yu et al. (Hong-Yu et al., 2023), *Mean Absolute Percentage Error* by (Barbosa et al., 2021), and *Dice Similarity Coefficient* by (Khan et al., 2021). Most papers used metric combinations (55 articles, 68 %).

We found that the reported performance was extremely high. In 32 of the 59 articles reporting "Accuracy", the score was 95 % or more. Accuracy of 99–100 % was achieved in 11 studies. Measurements of

"Precision" were higher than 0.98 in some studies. The F1 metric was 90 or higher in at least 14 of the 28 papers. Some studies reported even 100 % accuracy or an F1 metric of 100, which is a clear sign of overfitting. The tendency of these outstandingly high scores might result from the model evaluation scheme. Typically, the test dataset should be collected independently from the data collection used for model training and validation, and the model performance should be evaluated using the test dataset. Otherwise, the model test was done for the same, biased dataset, e.g., the test was not done in another region, spatial autocorrelation was ignored, or the test was not done with data from other years. As a result, the models are neither generalisable nor scalable, but highly specialised for the particular data acquisition pipeline. Also, it is possible that augmented data was used in both training and test datasets. In the practical recommendation section, we suggest some promising solutions – cross validation in particular – to these problems.

3.5. Using a small dataset has several attractive benefits

It is often the case that it is impossible to obtain additional annotated data regardless of whether the researcher wants it or not: these include, for instance, studies that investigate rarely observed phenomena (e.g., climate extremes, rare species, and disease outbreaks), cover a specific narrow geographic space (e.g., a single agricultural field), are limited by time and resources (e.g., studies in low-income countries), or those that employ data recorded before digitalisation. However, this does not mean there are no advantages of using a small dataset to solve various RS problems by using DL.

One of the biggest advantages of using a small RS dataset is faster training times. This can be particularly useful for prototyping and experimentation with different models and hyperparameters. With smaller datasets, multiple models can be trained in a relatively short time, making it easier to compare and select the best model (Althnian et al., 2021; Prusa et al., 2015). Another benefit is reduced memory requirements and less storage. The use of a smaller dataset reduces the memory requirements, making it possible to train on resource-constrained devices, such as laptops or embedded systems (Katsaragakis et al., 2020). A small dataset may also make it difficult to train (overly) complex models with many parameters. In these cases, simpler models may indeed be more suitable; they also have the advantage of being easier to train and validate the performance of the model (D'souza et al., 2020; Elsken et al., 2017; Keshari et al., 2018; Liu and Deng, 2015). The use of a small dataset combined with a low-complexity model can still result in sufficient performance (Brigato and Iocchi, 2020). Collecting and labelling an RS dataset is in some cases a complex, time-consuming, and costly task for research centres and organisations, which often leads to the use and preparation of a small set of available data. It could be drone data, digital camera imagery, or a few plots of satellite data. Thus, the use of a small dataset can reduce operating costs (Wang et al., 2023; Zhao, 2017). Finally, small datasets are easier to annotate, which is useful in cases where manual annotation is required.

Of course, small datasets have downsides as well. The main disadvantage is the lack of generalisability and transferability due to overfitting (Liu et al., 2017a), resulting in poor performance when using unforeseen datasets (Bailly et al., 2022; Wu et al., 2021a; Power et al., 2022). Small datasets also may be biassed (Althnian et al., 2021; Lones, 2023; Schat et al., 2020).

4. Practical recommendations for DL implementation strategies

In the previous section, we noted that data augmentation and transfer learning are popular, but other techniques are also promising. To address this issue, this section offers practical recommendations on strategies for the implementation of DL. We introduce the following techniques: TL, self-supervised learning, semi-supervised learning, few-shot learning, zero-shot learning, weakly supervised learning, process-aware learning, multitask learning, and ensemble learning. The literature search in this section was conducted in the same way as the previous search for major issues. However, we added a query that was able to return studies applying one of the above-mentioned techniques (e.g., "transfer learning", "semi-supervised learning", "few-shot learning") to the main search keywords. In addition, Table 1 shows a

Table 1
Summary of ten useful deep learning techniques + one validation technique for effectively analysing small datasets. No. corresponds to the subsection number in Section 4 in this article.

No.	DL technique	Short description with potential advantages (+) and disadvantages (–)
1	Transfer	To use a model that is pre-trained using a large, relevant dataset after fine-tuning using the target dataset +) improved performance, reduced data requirements, enhanced generalisability –) risk of performance reduction if transferred to a different domain, unnecessarily large model size
2	Self-supervised	To build a model that is pre-trained using the unlabeled target dataset with self-created labels, followed by supervision with provided labels +) no label required for learning features, unlabelled data usability, enhanced generalisability –) computationally expensive, chance that a model stops learning with some methods
3	Semi-supervised	To use a mix of supervised and unsupervised learning for training a model with labelled and unlabelled datasets +) mixed use of labelled and unlabelled data, enhanced generalisability –) computationally expensive, over-fitting risk, sensitive to data quality
4	Few-shot	To teach a model to generalise for new tasks or problems with only a few labelled examples per class +) directly targeting small data problems, rapid model adaptation, enhanced generalisability –) limited task complexity, over-fitting risk, sensitive to data quality
5	Zero-shot	To use a few-shot learning model that is trained to recognise objects/classes it has never seen before +) adaptable to an unknown class or entity, enhanced transferability –) extremely sensitive to data quality of new instances
6	Active	To train a model while selecting informative examples, labelling them, and adding them to the training dataset +) reduced labelling cost, efficient training for a specific task –) over-fitting risk, lower generalisability
7	Weakly supervised	To train a model by using data that are labelled partially, noisily, or imprecisely +) reduced labelling cost, allowing inaccurate label (uncertainty of ground truth) for scalability –) computationally expensive, less accurate than (fully) supervised learning
8	Multi-task	To develop a model that learns general features useful for solving multiple tasks +) efficient training for multiple tasks, enhanced generalisation, reduced data requirement –) modeling complexity, task interference, limited scalability
9	Ensemble	To combine many individual models that learned differently from each other for prediction +) enhanced generalisability, robustness for data perturbation, accounting for uncertainty –) computationally expensive, lower interpretability than a single model
10	Process-aware	To incorporate process-based regulation into learning +) relying on mechanistic understanding, enhanced transferability –) risk of performance reduction if relying on a wrong assumption
11	Cross validation	To train and validate a model for several times using alternating partitions for training and validation to avoid gaining an over-optimistic model performance score +) less biassed model evaluation, enhanced generalisability –) computationally expensive

succinct summary about the advantages and disadvantages of each DL technique for use with small datasets. We found 37 articles as of 16 February 2023 (Fig. 2). We also present a practical flowchart for deciding which algorithm to use in each specific use case: We believe that this flowchart is a unique contribution to DL users as it can help identify which techniques to use for different use cases in a simple way (Fig. 3). We do not explicitly cover other methods such as data augmentation and regularisation, since they have been widely covered in various literature, such as by the work by (Shorten and Khoshgoftaar, 2019).

4.1. Transfer learning

As described before, TL is a popular technique that derives learning from one task and reuses it to solve another (similar) task. As (Iman et al., 2022) have explained TL is widely used in labelled dataset fields such as radar images, medical images, malware classification, facial emotion recognition, mechanics, vision, human activity recognition, civil engineering, Natural Language Processing (NLP), military, human sciences/psychology, chemistry, security, physics/astrophysics, and telecommunications.

Typically, TL takes place when a neural network is pre-trained on a large dataset, such as ImageNet (Russakovsky et al., 2015), and then its weights are used to fine-tune it on a smaller dataset for a specific task (Fig. 4). TL can also be used as a feature extraction method to develop a second model that can be trained on the target data. The idea of TL is to apply knowledge from the source task to the target task, potentially improving performance, reducing the need for large training data, preventing overfitting, reducing the otherwise huge computation cost, and saving time (Rawat and Wang, 2017). Pre-training on a general dataset is particularly effective when the task-specific dataset is small or when there is limited labelled data available.

We found three ways that TL was used in the literature: fine-tuning a pre-trained model, using pre-trained features as input for a new model, and combining pre-trained models. Fine-tuning involves taking a pre-trained model and training it further on the target task (Dong et al., 2021; Ziegler et al., 2020). Using pre-trained features means that the output of one or more layers of a previously trained model is used to develop a new model trained on the target task. Combining pre-trained models involves training multiple pre-trained models on related tasks and then subsequently combining them to make predictions about a different, previously unexplored phenomenon.

TL is already actively used to solve a variety of problems in RS when the dataset is small. We found a total of 14 papers regarding the use of TL in RS on a small sample. For example, a paper by (Wang et al., 2018b) proposed a DL framework for RS image registration based on TL that would reduce the huge computational cost in the training stage, speed up the framework, and achieve additional performance gains. The experiments conducted on seven sets of RS images acquired by RADARSAT, SPOT, and Landsat showed that the proposal improved registration accuracy by between 2.4 % and 53.7 %.

(Zhang et al., 2019b) used TL to classify HSI due to very limited training data and the massive parameters of end-to-end 3-D lightweight models. Moving to the problem of radar-jamming detection (Hou et al., 2022) and (Lv et al., 2022) separately proposed methods based on TL. In (Character et al., 2021), researchers used TL not only to compensate for a small dataset (Lidar and Sonar), but also to address false positives by training the YOLOv3 model on both shipwrecks and background topography. Another example was forest-fire detection using YOLOv5 by (Xue et al., 2022a), improving the performance of mAP@0.5 by up to 10.1 %. (Wang et al., 2022b) applied TL to weed density extraction based on few-shot learning through UAV and multispectral images in an ecological irrigation area using a pre-trained AlexNet algorithm.

A kind of TL known as domain adaptation was applied to synthesise training data under diverse environmental conditions with automatic labels using YOLOv3 (Zhao et al., 2021b). The results from that paper showed that their proposed method improved bale detection. Moreover, this approach could be easily scaled to many other crop field objects. (Chen et al., 2022) used the Faster R-CNN domain adaptation for aircraft detection on the DOTA dataset. In (Yu et al., 2022), the authors showed that their method based on TL could accurately extract terraced field surfaces and segment terraced field boundaries with an overall accuracy above 93.12 %.

In another experiment, TL solved the problem of poor adaptability of the DenseNet-121 network to RS images acquired from different platforms, and was able to properly identify disaster-damaged buildings (Yang et al., 2021). Other examples included scattering shrub detection (Guirado et al., 2017), fir tree detection (Safonova et al., 2019), HSI classification (Feng et al., 2022a), land cover classification (Naushad et al., 2021; Qiu et al., 2022), and seismic data analysis (El Zini et al., 2020). The average accuracy in these works after applying TL to the new small datasets was over 93 %.

Key recommendations for using TL include selecting the right pre-trained model, determining the level of TL (feature extraction, fine tuning, or both), determining which layer(s) to transfer, generously employing data augmentation, regularising the network, and evaluating

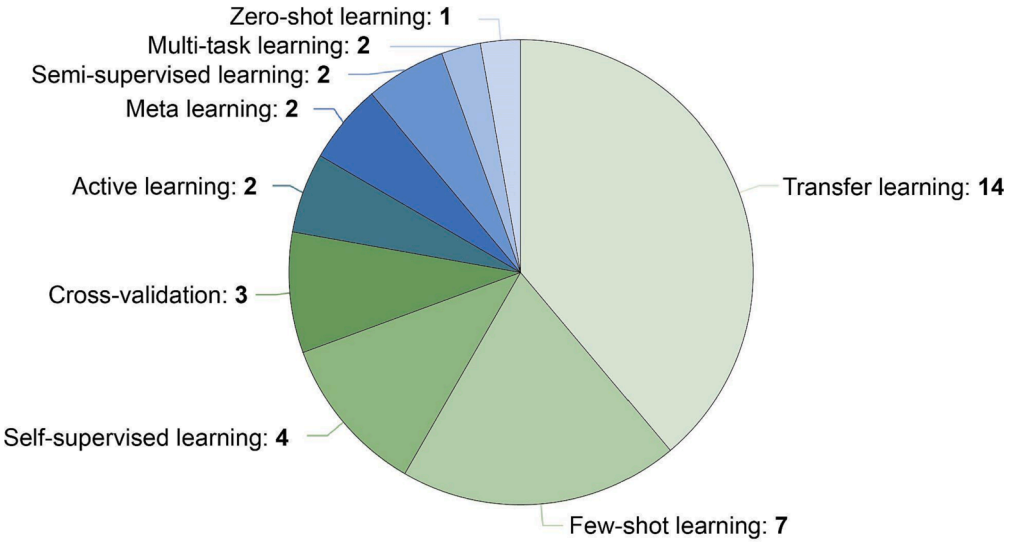


Fig. 2. The number of articles that use particular deep learning techniques addressing the small data problem in remote sensing applications as of February 2023.

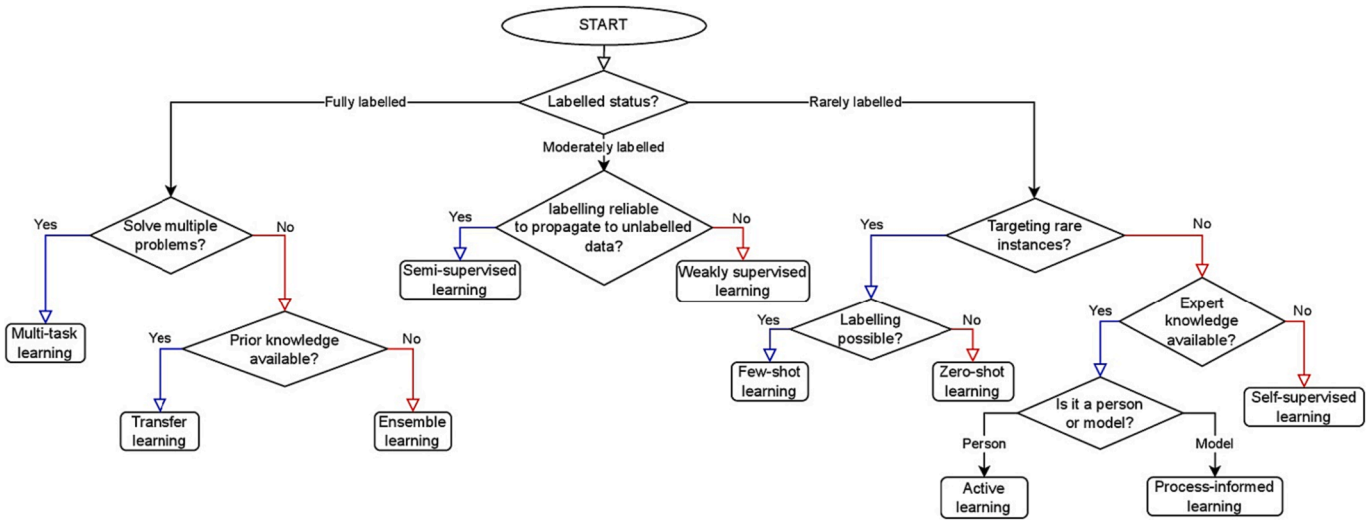


Fig. 3. Practical flowchart for selecting an appropriate deep learning technique (as of May 2023) to address the small data problem in remote sensing applications.

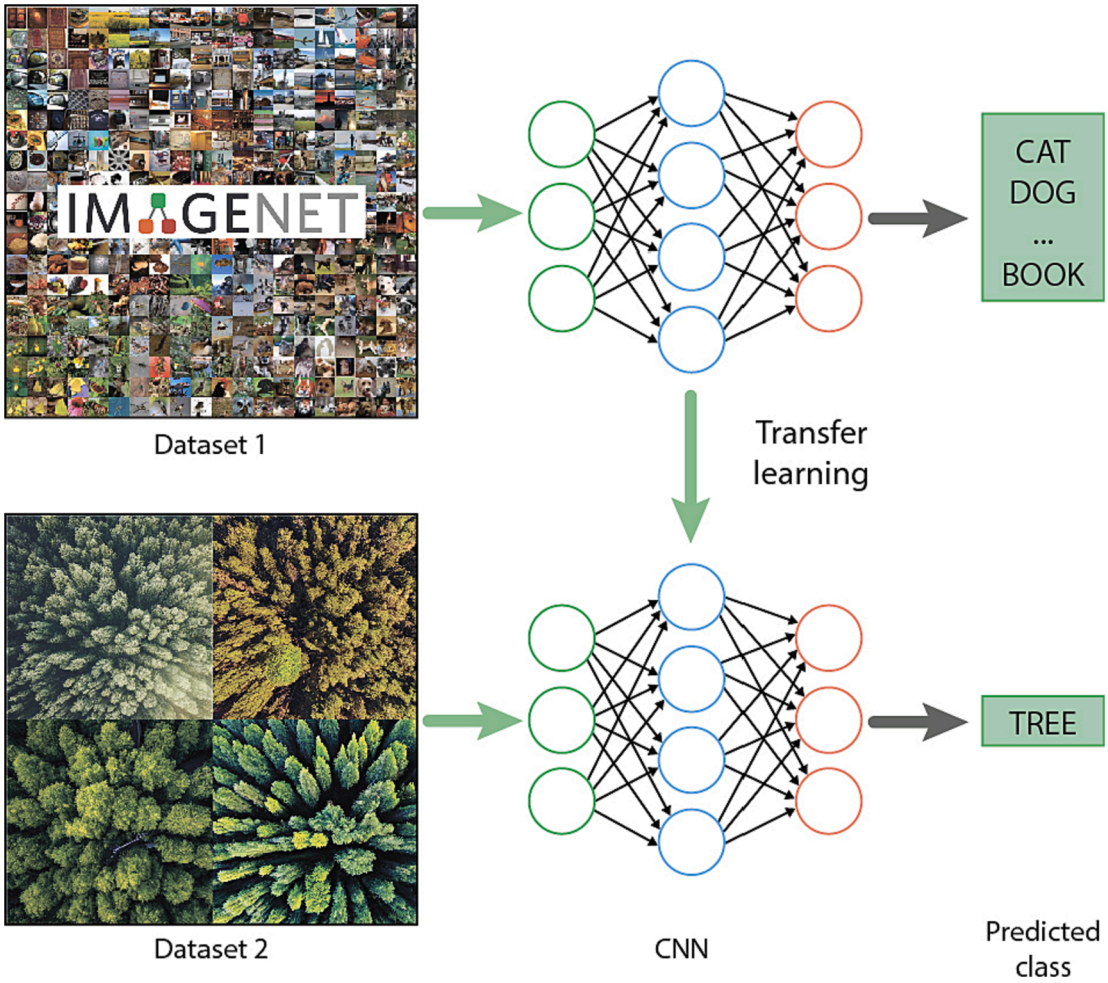


Fig. 4. A schematic diagram of transfer learning. The upper part depicts model pre-training with Dataset 1 (e.g., ImageNet). The weights are transferred and fine-tuned for a specific task with Dataset 2. Here, Dataset 2 contains only images of land cover taken with UAVs, where the convolutional neural network (CNN) model aims to predict the class (tree or not).

performance. One common practice is to import and use ready-made models from DL libraries such as TensorFlow, Keras, Theano, and PyTorch. The most popular models were AlexNet, VGG, Xception, Inception, MobileNet, DenseNet, ResNet, GoogleLeNet, and YOLOs. In (Abu et al., 2022; Sharma et al., 2021; Zhao, 2017), all suggested considering fine-tuning several hyperparameters (feature map, filter size, activation function, pool size, optimiser, learning rate, batch size, epoch, dropout rate, loss function, and evaluation metric) of the pre-

trained model.

4.2. Self-supervised learning

Self-supervised learning is a technique related to transfer learning. However, in contrast to traditional transfer learning, self-supervised learning does not require labelled data for pre-training – it can leverage the structure of unlabelled data to generate labels for the pre-training task. However, like transfer learning, a model pre-trained with self-supervised learning is further fine-tuned on the labelled downstream task (Rani et al., 2023).

This technique has been employed in medicine and healthcare (Chen et al., 2019; Krishnan et al., 2022), physics (Ma and Liu, 2020), speech representation (Mohamed et al., 2022), RS (Wang et al., 2022b), time-series analysis (Pöppelbaum et al., 2022), vide o processing (Wang et al., 2022d; Yan et al., 2020; Jing and Tian, 2019), speech processing (Hsu et al., 2021), and target tracking (Yuan et al., 2021) for example. The most-cited papers involve medical research, solving problems such as accurate detection of tissue in monocular endoscopy (Liu et al., 2019b), retinal disease diagnosis (Li et al., 2020), MRI parameter mapping or reconstruction (Liu et al., 2021; Yaman et al., 2020), 3D medical-image analysis (Zhu et al., 2020), and homography estimation (Wang et al., 2019a).

However, we only found four papers dealing with self-supervised learning in RS using a small sample (Liu et al., 2022a; Rangnekar et al., 2020; Song et al., 2022; Xue et al., 2022b). These were related to solving the HSI classification problem. The main reason for using self-supervised learning was the scarcity and high cost of labelled HSI samples. In (Song et al., 2022), the authors proposed a dual-branch residual neural network (ResNet) to fuse spectral and spatial information. (Liu et al., 2022a) presented a novel ensemble self-supervised feature learning method using multiple HSI datasets. As the papers presented performed the same task on the same datasets, some of their comparative results can be presented in Fig. 5. (Xue et al., 2022b) proposed a generative self-supervised feature learning architecture for multimodal RS-imaged land cover classification. In this case, the self-supervised feature learning architecture was able to extract highly sophisticated, robust feature representations from multi-view data; this process did not require any labelled information, thus alleviating the otherwise critical need for annotated samples. To solve the same problem, (Rangnekar et al., 2020) compared the performance of SegNet, U-Net, and Res-U-Net for scene understanding and object identification by using dense semantic segmentation to establish a benchmark for a given scene.

Despite the fact that the use of self-supervised learning technology on

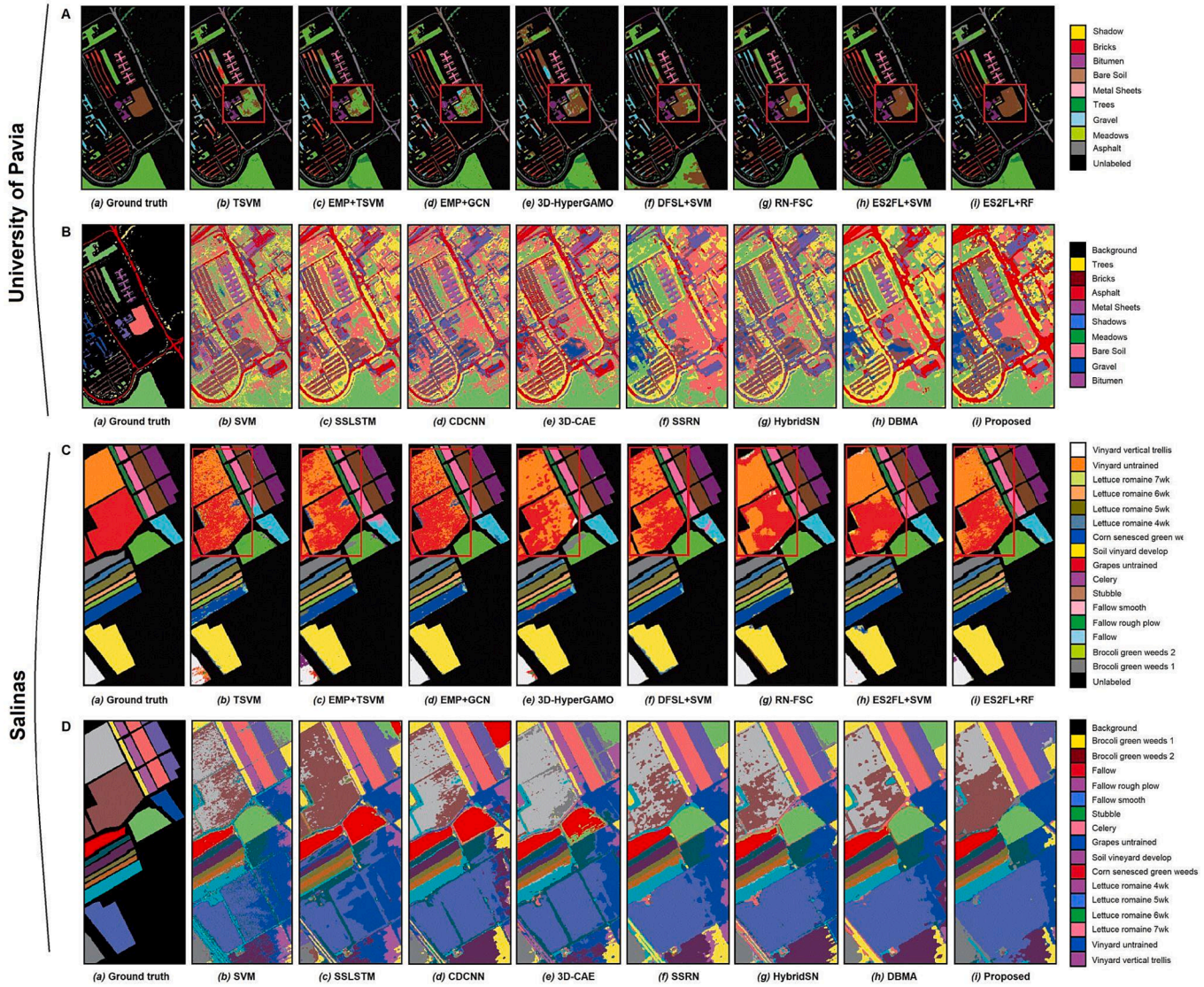


Fig. 5. Comparative results from the application of self-supervised learning technology in remote sensing using hyperspectral images (HSIs) from Italian University of Pavia (UP) and Salinas (SA) datasets, where A and C are the results presented by (Song et al., 2022), and B and D are the results presented by (Liu et al., 2022a).

small datasets is relatively unexplored, we have come across some works on similar topics (Cao and Wu, 2021; Su et al., 2020). (Su et al., 2020) presented a systematic study by varying the degree of domain shift and analysing the performance of multiple *meta*-learners on a variety of domains. The authors found that the improvements were greater when the training set was smaller or the task was more challenging. They also noted that self-supervised learning can degrade performance if the distributions of the images used for *meta*-learning and self-supervised learning are different. (Cao and Wu, 2021) proposed a system of scaled-down self-supervised learning, which included three parts: small resolution, small architecture, and small data. The authors showed that this approach could achieve impressive results on small data alone, even without a large pre-training dataset.

4.3. Semi-supervised learning

Semi-supervised learning is a technique that has been actively implemented in a number of different areas over the past few years. Self-supervised learning is a mix of supervised learning and unsupervised learning to leverage unlabelled data, in which a model is repeatedly trained and updated using both the labelled and the generated pseudo-labels (that is, predicted labels) for the unlabelled data. This can result in better performance than supervised learning alone (Han et al., 2022; Wang et al., 2021).

Across all scientific disciplines, one of the most-cited papers was presented by (Ma et al., 2019b) dealing with probabilistic representation and the inverse design of metamaterials. Another work was dedicated to detecting fake users on Twitter (BalaAnand et al., 2019). Later, (Xu et al., 2021) presented a paper on image recognition and facial attribute recognition using a semi-supervised, self-growing generative adversarial network (SGGAN). The authors claimed that when they used training data with only 4 % labelled facial features, their approach was nevertheless able to achieve accuracy comparable to that of leading supervised DL methods with all labelled facial features. (Rostami et al., 2020) used semi-supervised learning to choose a subset of available features that had the lowest redundancy with each other but also the highest relevance to the target class with limited training data in a Synthetic Aperture Radar (SAR) classification. (Tseng et al., 2021) proposed DNetUnet for medical image segmentation. As of 2022, some of the most cited papers investigated road damage detection (Shim et al., 2022), drift compensation for olfactory sensors (Lu et al., 2022), and mechanical fault diagnosis (Feng et al., 2022b).

As for the implementation of the semi-supervised learning strategy in the field of RS under conditions of small sample sizes, we came across only two articles. (Jozdani et al., 2021) deployed a teacher-student semi-supervised learning approach (based on the U-Net and U-Net++ networks) involving unlabelled UAV and WorldView-2 data to assist with improving model performance to map caribou lichen. This approach produced a reasonably accurate (overall accuracy of 85 % and F1 score of 84 %) lichen map at the WorldView scale.

Although semi-supervised learning is suitable for unlabelled data, it is recommended to use labelled data because the quality of the labelled dataset will directly affect model performance. Another important point is to experiment with different ratios of labelled and unlabelled data to find the optimal balance for a particular task (Chapelle et al., 2009; He et al., 2021).

4.4. Few-shot learning

The goal of few-shot learning is to teach models to generalise for new tasks or problems with only a few labelled examples per class (Fig. 6).

Few-shot learning is therefore a type of *meta*-learning, which involves training a model on a set of related tasks so that the model can then learn to quickly adapt to new, similar tasks with only a few examples (Hospedales et al., 2020). This method has gained popularity in RS for its ability to solve the problems of agriculture and areal scene classification (Li and Yang, 2021; Zhang et al., 2021). Recently, (Gao et al., 2021b), (Zuo et al., 2022), and (Li et al., 2022d) all applied *meta*-learning to HSI classification. Another approach is to use a widely cited generative adversarial network to teach the model a high-level representation of the data (Goodfellow et al., 2014). Despite the widespread popularity of such methods in the field of medicine (Yi et al., 2019), they are only recently beginning to be used in RS (Liu et al., 2018; Zhong et al., 2020; Zhu et al., 2018).

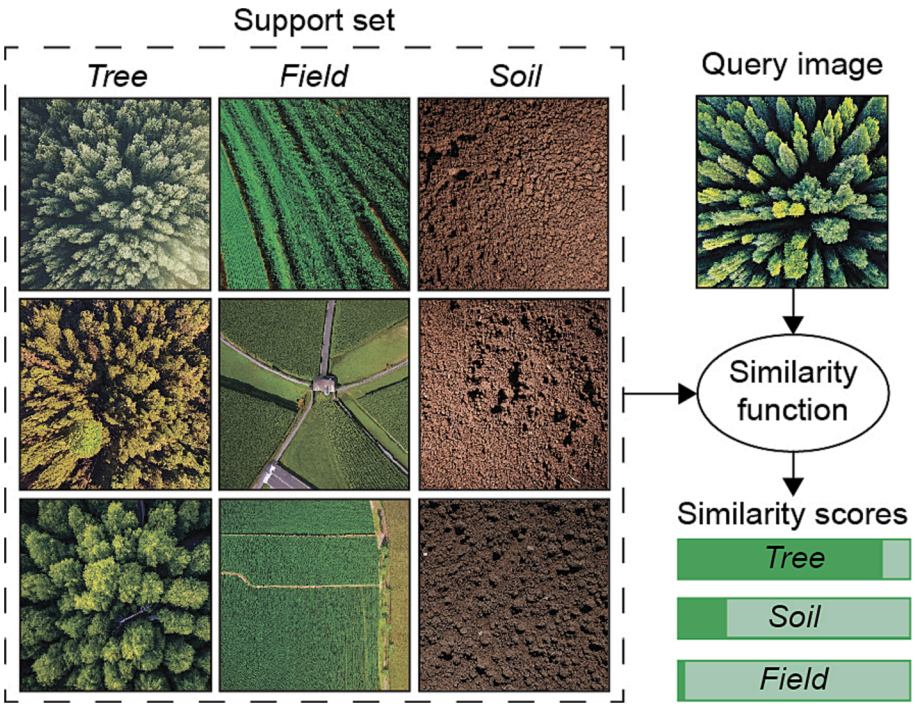


Fig. 6. An example of few-shot learning. The "tree" class label prediction by the pre-trained model to determine similarity probabilities on the query image by using the support set information.

The application of few-shot learning in RS was found in seven articles in our search: (Bai et al., 2022a), (Li et al., 2022d), (Liu et al., 2022b), (Rao et al., 2019), (Zuo et al., 2022), (Wang et al., 2022b), and (Wang et al., 2022c), and other. Most of the papers dealt with HSI classification problems. (Wang et al., 2022b) proposed weed and crop density extraction using RGB and multispectral images in an ecological irrigation area. (Liu et al., 2019a) proposed an algorithm based on few-shot learning in three steps. First, spectral-spatial features are extracted to reduce the labelling uncertainty via a deep residual 3-D CNN. Second, the network is trained in episodes to learn about a metric space where samples from the same class are close and those from different classes are far. Finally, the testing samples are classified by a nearest neighbour classifier in the learned metric space. A similar algorithm was also proposed by (Bai et al., 2022a). A small-scale high-precision network called "3-D convolution random Fourier features (3-DCRFF)" was presented by (Wang et al., 2022b). Yet another method was based on an edge-labelling graph neural network (FSL-EGNN) created by (Zuo et al., 2022).

4.5. Zero-shot learning

Zero-shot learning is a special type of few-shot learning method, which is trained to recognise objects or classes it has never seen before. The model is trained on a set of known classes or objects, but is also given additional information about the relationships between these classes, such as semantic or visual similarities (Wang et al., 2020a). This additional information is used to help the model recognise new, unseen classes or objects that are related to the known classes. This method is useful in situations where it is difficult or expensive to obtain labelled data for new classes or objects. The method has been applied to medical image segmentation (Wang et al., 2018a), attribute-based classification (Lampert et al., 2014), industrial fault diagnosis (Feng et al., 2022b), label-embedding for image classification (Akata et al., 2016), and instance segmentation RS images based on the segment anything foundation model, incorporating semantic category information (Chen et al., 2023). We found only one article in RS applications with small dataset, by (Sumbul et al., 2019). Their paper presented object recognition for 40 different types of street trees using areal data. Experiments showed that their proposed model achieved a 14.3 % normalised recognition accuracy for the classes with no training examples, which was significantly better than a random guess accuracy of 6.3 % for 16 test classes, as well as the accuracy levels of three other zero-shot learning algorithms.

Both few-shot and zero-shot learning methods can be a valuable technique for extracting the best performance from limited data (Villon et al., 2021; Zhang et al., 2019b). However, when using them with small samples, special attention should be paid to data augmentation, the selection of an appropriate evaluation metric, and the ensemble of multiple few-shot learning models. All of these can significantly improve the performance of the model.

4.6. Active learning

Active learning is a powerful technique that can help ML models achieve greater accuracy, while reducing the amount of labelled data required (Koller et al., 2022; Pardakhti et al., 2021; Ren et al., 2021; Settles, 2009). This technique involves selecting the most informative examples, labelling them, and adding them to the training dataset. Active learning has been successfully applied in many disciplines, including medicine (Littlewood et al., 2013; Nakarmi and Santosh, 2023), NLP (Arora and Agarwal, 2007; Zhang et al., 2022) and computer vision (Takezoe et al., 2022; Wu et al., 2022a). The idea of active learning use in RS in particular has been presented by (Liu et al., 2017b) and (Cao et al., 2020) for HSI image classification. However, we found only two papers using active learning for RS on a small dataset. For example, (Han et al., 2020) select information-rich and representative samples by combining the idea of active learning and input into a support vector machine (SVM) classifier, and achieve superior classification accuracy of remote sensing sea ice images with small samples. In another work (Zhou et al., 2020), SAR target classification was achieved with limited data using data-driven active learning. Since active learning can be guided by the user's intention, it can reduce labelling costs and training time, and improve accuracy compared to randomly selecting examples for a specific task. Active learning can be especially useful for small datasets, as it allows the model to learn more efficiently from a limited number of labelled examples (Pardakhti et al., 2021). However, due to this data limitation, the model may be more prone to overfitting (Farquhar et al., 2020). To avoid this, we recommend applying regularisation methods and monitoring the performance of the model during training.

4.7. Weakly supervised learning

In cases when collecting full ground-truth labels is time-consuming, expensive, or otherwise practically impossible, then it is useful to use weakly supervised learning (Zhou, 2018). This is a type of ML in which the training data has been labelled partially, noisily, or imprecisely. One popular application of this is label propagation: the use of a small set of labelled data to generate labels for a larger set of unlabelled data. Another application is multi-instance learning, where each point of the training data exists in multiple instances, but a subset of these instances are labelled.

Weakly supervised learning has been successfully applied in various applications, including image classification, object detection, semantic segmentation, and NLP. However, the use of weakly supervised learning under the conditions of a limited dataset has only been presented in a small number of papers. In one striking case, (Liu et al., 2022c) presented an application that could identify acute lymphoblastic leukaemia with outstanding accuracy, approximately 91.9 %. Another example was the classification of tiny spike-like projections on the basement membrane of the glomerulus by (Wu et al., 2022b). According to the results of the trial, the accuracy was 94.05 %. (Ruan et al., 2022) conducted two fault diagnosis experiments on ball bearings and bevel gears with 97.23 % and 99.76 % accuracy. Another example is a work presented by (Kim et al., 2021) with segmentation for an autonomous combine harvester. Their results showed that their proposed weakly supervised crop area segmentation (WSCAS) method could be performed with the lowest inference time, and crop area could be localised with an IoU of about 94 %. In all experiments, the authors of these papers claim that their proposed algorithms are superior to other existing methods, even under the conditions of a small sample size. It should be noted that we did not come across works where the methodology was used for RS on a small dataset. However, this technology has been actively used for optical images (Cheng and Han, 2016; Han et al., 2015), areal and satellite images (Wei and Ji, 2020), and HIS (Yao et al., 2016).

Weakly supervised learning has the potential to reduce the cost and effort of collecting accurately labelled data, and can be used in various applications where obtaining reliably labelled data is difficult or impractical, even in small datasets. One common approach is to use transfer learning and active learning, which can help improve model performance. In addition, using the weakly supervised learning method with limited data can help reduce the need for large amounts of labelled data, while still achieving high accuracy.

4.8. Multi-task learning

Multi-task learning is another powerful technique that can be implemented in RS when only a small dataset is available. It is designed to improve the performance of multiple related learning tasks by leveraging useful information among them (Zhang and Yang, 2018). The goal is to develop a model that can learn to generalise well about new instances of each task, while also benefiting from the shared knowledge learned across all tasks. This means that the model learns to solve

multiple related problems using the same or shared representations, instead of developing independent models for each task. The model typically shares lower-level layers across all tasks, while having task-specific layers at higher levels. This way, the model can extract general key features shared across multiple tasks.

Multi-task learning's impressive track record has helped it gain popularity in recent years. The frequency of publications describing the use of this technology is growing by 25–30 % every year. It has led to success in many ML applications, from NLP and speech recognition to computer vision and drug discovery (Ruder, 2017; Sosnin et al., 2019; Zhao et al., 2023). Multi-task learning is also actively used in RS, particularly for classification (Qi et al., 2017), target detection (Wu et al., 2019c; Zhang et al., 2017), semantic segmentation (Li et al., 2022c; Volpi and Tuia, 2018), and feature representation tasks (Xiong et al., 2019). However, its application when using a small number of training samples is still rare, as only two papers have discussed this (Quan et al., 2023; Zhao et al., 2021a). The first paper proposed a multi-aspect SAR target recognition method based on a prototypical network. The second paper deals with the detection of building changes using pseudo-labels generated by high-availability semantic segmentation on three available RS datasets. This method can significantly improve the recognition performance of the DL model under a small number of samples, and thus the recognition accuracy can approach that of a model with a complete training set.

4.9. Ensemble deep learning

Ensemble learning is a method that combines many individual models to obtain better generalisation performance (e.g., random forests and boosting). While this approach is commonly used for tabular data analysis, its application in DL models is far less popular because it requires huge computational resources and time. Nevertheless, ensemble DL models have the potential to harness the benefits of DL architecture as well as ensemble learning (e.g., to avoid overfitting). Previously, this method has been applied to predicting short-term traffic flow (Zhang and Xin, 2022), predicting plant miRNA–lncRNA (Hamdy et al., 2023), and identifying the drivers of vehicles by using Controller Area Network (CAN) bus data (Hu et al., n.d.). Ganaie et al. also reviewed a variety of techniques that have been applied in different domains (Ganaie et al., 2022). (Liu et al., 2022a) presented a novel ensemble self-supervised feature learning method on multiple HSI datasets. Since any of the above-mentioned learning techniques can be combined, ensemble learning has promising potential for further applications.

4.10. Process-aware learning

Process-aware learning refers to the process of incorporating knowledge into ML models about the underlying processes or mechanisms that generate data. It is particularly helpful for understanding the underlying causal relationships between variables, thus leading to better predictions and decisions. One popular application is known as "physics-informed learning" (Karniadakis et al., 2021; Raissi et al., 2019, 2017) in the domain of physics. Also, this technology was previously used in mechanics (Cai et al., 2021), medicine (Fossan et al., 2021; Meier and Heijman, 2022), and computer science (Doan et al., 2019; Kashefi and Mukerji, 2022). Although we did not find the process-aware learning approach in any RS domain, we can imagine several use cases. For instance, researchers could use a vegetation growth model to simulate parameters that are difficult to measure in the field, use these simulated parameters as labelling for images, and then train a DL model with the labelled data.

4.11. Cross validation

As we mentioned above in subsection 3.4, it seems that previously reported model performances have been suspiciously high, indicating a lack of appropriate evaluation systems. Applying an appropriate validation strategy is important for model generalisability and transferability (Vabalas et al., 2019), while preventing overfitting (Ying, 2019). Note that cross validation does not belong to learning technique, but the method can be jointly used with any of the above-introduced learning methods.

The most commonly used type of validation strategy in ML is cross-validation (Little et al., 2017). This procedure is quite common in ML via tabular dataset analysis, but it is rarely applied to large datasets for DL. We believe that k-fold cross-validation would be useful for evaluating DL model performance with a small dataset – and this capacity for validation is another benefit of using a small dataset. Moreover, in the area of RS, random sampling for validation may not be the best idea, because spatial and temporal data typically reveal high autocorrelation levels. Several recent studies have pointed out that autocorrelation leads to a violation of the assumption of data independence between training data and the validation set (Kattenborn et al., 2022; Le Rest et al., 2014; Ploton et al., 2020). Spatial rather than random cross-validation can be used for less biased model assessments (Roberts et al., 2017) (Fig. 7).

The cross-validation technique was used by (Xue et al., 2019) and by (Chen et al., 2018) to solve problems with estimating PM2.5 concentrations across China, and (Yang et al., 2018) to estimate grassland biomass. Other case studies have included mapping soil properties from high-resolution RS data (Forkuor et al., 2017), mapping fire intensity (Gibson et al., 2020), and quantifying rangelands (Rigge et al., 2020). We found four papers that directly investigated cross-validation for RS imagery under small sample conditions (Freeman et al., 2019; Lange et al., 2022; Odebiri et al., 2022; Wu et al., 2021b). All the publications applied random cross-validation with different fold sizes (4 and 10 folds), each at different scales (local, regional, and national). The tasks performed were the mapping of regressed soil organic carbon content, HSI-based land cover classification, and plant water stress detection.

5. Conclusions

In this review paper, we performed a survey of the small data problem in RS data in DL implementation and suggested promising DL techniques to address the problem. First, we summarised 80 studies from 2016 to 2023, and presented the possibilities to address the small data problem with advanced DL techniques beyond conventional learning methods. For this, we first had to define what "small data" means. Then, we described the few previous studies that had analysed RS processes employing DL techniques under conditions of small data, and we looked at the advantages and disadvantages of using small datasets. Finally, we offered a set of practical recommendations about how RS scientists can better implement DL techniques to fully take advantage of a small dataset. As one previous paper noted (Keshari et al., 2020) a variety of approaches can be used to solve the small data problem, such as data augmentation, data fine-tuning, the adaptation of pre-trained models, and reducing the dependence on large-sample learning. However, in our review, we also presented even more techniques that are worth considering when working with a small dataset. We identified a total of ten learning techniques for addressing the small data problem: Transfer learning, self-supervised learning, semi-supervised learning, few-shot/zero-shot learning, active learning, weakly supervised learning, multi-task learning, ensemble learning, and process-aware learning. Cross-validation is also a valuable tool for improving the use of DL in RS (Fig. 3). Our goal has been to show ways to implement DL applications for research where ground-truth (annotation) data is difficult to obtain, while making it possible to solve various problems involving classification, detection, or segmentation. These problems could include biodiversity loss, climate extremes, and sudden changes in socio-environmental systems. For future work, we plan to apply different DL strategies to solve various practical problems in RS under small dataset conditions, where this modern technique has not been adequately used. In addition, the practical flowchart shown in Fig. 3 and the summary

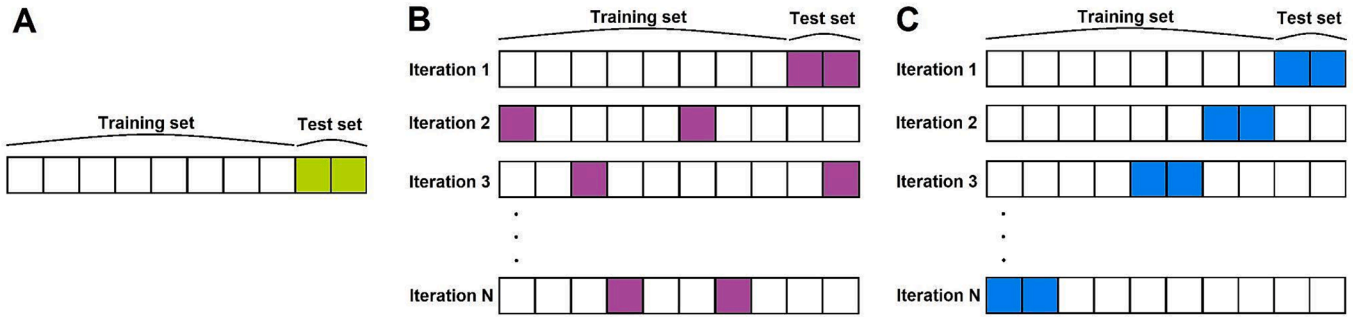


Fig. 7. Model validation schemes. (A) Simple data split without iteration, which is often done practically when a deep learning model is used in the remote sensing domain; (B) repeated random subsampling validation, where one repeats model evaluation while conducting subsampling to estimate model performance instability; (C) k-fold cross validation, where one firstly divides the data into groups and then recursively evaluate model performance using one of the groups. If grouping is done based on the spatial coordinate, it is so-called spatial cross validation.

table (Table 1) will be further improved based on the results of our future experiments and new studies presented by the RS scientific community using other AI techniques in small data problems. Finally, we hope that by justifying the use of small datasets, this review will motivate more researchers to experiment with other techniques and apply them to different RS problems.

Declaration of Competing Interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

This study was supported by the Federal Ministry of Education and Research (BMBF – Bundesministerium für Bildung und Forschung) project “Multi-modale Datenintegration, domänenspezifische Methoden und KI zur Stärkung der Datenkompetenz in der Agrarforschung” (16DKWN089). We thank Sebastian Raschka for offering valuable feedback.

References

Abu, M., Zahri, N.A.H., Amir, A., Ismail, M.I., Yaakub, A., Anwar, S.A., Ahmad, M.I., 2022. A Comprehensive Performance Analysis of Transfer Learning Optimization in Visual Field Defect Classification. Diagnostics 12, 1258. https://doi.org/10.3390/diagnostics12051258.

Adadi, A., 2021. A survey on data-efficient algorithms in big data era. J. Big Data 8, 24. https://doi.org/10.1186/s40537-021-00419-9.

Adugna, T., Xu, W., Fan, J., 2022. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sens. 14, 574. https://doi.org/10.3390/rs14030574.

Akar, Ö., Güngör, O., 2012. Classification of multispectral images using Random Forest algorithm. J. Geod. Geoinformation 1, 105–112. https://doi.org/10.9733/jgg.241212.1.

Akata, Z., Perronnin, F., Harchaoui, Z., Schmid, C., 2016. Label-Embedding for Image Classification. IEEE Trans. Pattern Anal. Mach. Intell. 38, 1425–1438. https://doi.org/10.1109/TPAMI.2015.2487986.

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A.B., Alzakari, N., Abou Elwafa, A., Kurdi, H., 2021. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. Appl. Sci. 11, 796. https://doi.org/10.3390/app11020796.

Arora, S., Agarwal, S., 2007. Active Learning for Natural Language Processing. Literature Review. Language Technology Institute, Scool of Computer Scince, Carnegie Mellon University, Pittsburgh, United States.

G. Astolfi M.C.B. Pache G.V. Menezes A. da Oliveira Junior S., Menezes, G.K., Weber, V. A. de M., Castelão Tetila, E., Belete, N.A. de S., Matsubara, E.T., Pistori, H., Combining Syntactic Methods With LSTM to Classify Soybean Aerial Images IEEE Geosci. Remote Sens. Lett. 18 2021 2182 2186 10.1109/LGRS.2020.3014938.

Bai, J., Huang, S., Xiao, Z., Li, X., Zhu, Y., Regan, A.C., Jiao, L., 2022a. Few-Shot Hyperspectral Image Classification Based on Adaptive Subspaces and Feature Transformation. IEEE Trans. Geosci. Remote Sens. 60, 1–17. https://doi.org/10.1109/TGRS.2022.3149947.

Bai, T., Wang, L., Yin, D., Sun, K., Chen, Y., Li, W., Li, D., 2022b. Deep learning for change detection in remote sensing: a review. Geo-Spat. Inf. Sci. 1–27. https://doi.org/10.1080/10095020.2022.2085633.

Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., Roy, P., 2022. Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Comput. Methods Programs Biomed. 213, 106504 https://doi.org/10.1016/j.cmpb.2021.106504.

BalaAnand, M., Karthikeyan, N., Karthik, S., Varatharajan, R., Manogaran, G., Sivaparthipan, C.B., 2019. An enhanced graph-based semi-supervised learning algorithm to detect fake users on Twitter. J. Supercomput. 75, 6085–6105. https://doi.org/10.1007/s11227-019-02948-w.

B.D.S. Barbosa G.A. Ferraz e S., Costa, L., Ampatzidis, Y., Vijayakumar, V., dos Santos, L. M., UAV-based coffee yield prediction utilizing feature selection and deep learning Smart Agric. Technol. 1 2021 100010 10.1016/j.atech.2021.100010.

Bhateria, R., Jain, D., 2016. Water quality assessment of lake water: a review. Sustain. Water Resour. Manag. 2, 161–173. https://doi.org/10.1007/s40899-015-0014-7.

Blekos, K., Nousias, S., Lalos, A.S., 2020. Efficient automated U - Net based tree crown delineation using UAV multi-spectral imagery on embedded devices, in: 2020 IEEE 18th International Conference on Industrial Informatics (INDIN). Presented at the 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), pp. 541–546. https://doi.org/10.1109/INDIN45582.2020.9442183.

Borg, E., Conrad, C., Truckenbrodt, S., Hüttich, C., Ahmadian, N., Dahms, T., Heupel, K., Spengler, D., Missling, K.-D., 2018. Joint Experiment for Crop Assessment and Monitoring (JECAM) - Test Site DEMMIN, 2018. Presented at the International Geoscience and Remote Sensing Symposium, IGARSS 2018, the 38th annual symposium. Valencia, Spain, pp. 1–4.

L. Brigato L. Iocchi A Close Look at Deep Learning with Small Data 2020 https://doi.org/10.48550/arXiv.2003.12843.

Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E., 2021. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mech. Sin. 37, 1727–1738. https://doi.org/10.1007/s10409-021-01148-1.

Y.-H. Cao J. Wu Rethinking Self-Supervised Learning: Small is Beautiful 2021 https://doi.org/10.48550/arXiv.2103.13559.

Cao, X., Yao, J., Xu, Z., Meng, D., 2020. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. IEEE Trans. Geosci. Remote Sens. 58, 4604–4616. https://doi.org/10.1109/TGRS.2020.2964627.

Chai, J., Zeng, H., Li, A., Ngai, E.W.T., 2021. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Mach. Learn. Appl. 6, 100134 https://doi.org/10.1016/j.mlwa.2021.100134.

Chapelle, O., Scholkopf, B., Zien, A., Eds., 2009. Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. IEEE Trans. Neural Netw. 20, 542–542. https://doi.org/10.1109/TNN.2009.2015974.

L. Character J.R. Ortiz A., Beach, T., Luzzadder-Beach, S., Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar Remote Sens. 13 2021 1759 10.3390/rs13091759.

Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., Rueckert, D., 2019. Self-supervised learning for medical image analysis using image context restoration. Med. Image Anal. 58, 101539 https://doi.org/10.1016/j.media.2019.101539.

K. Chen C. Liu H. Chen H. Zhang W. Li Z. Zou Z. Shi RSPrompter: Learning to Prompt for Remote Sensing Instance Segmentation based on Visual Foundation Model 2023 https://doi.org/10.48550/arXiv.2306.16269.

Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018. A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ. 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251.

Chen, J., Sun, J., Li, Y., Hou, C., 2022. Object detection in remote sensing images based on deep transfer learning. Multimed. Tools Appl. 81, 12093–12109. https://doi.org/10.1007/s11042-021-10833-z.

Chen, Y., Zhao, X., Jia, X., 2015. Spectral-Spatial Classification of Hyperspectral Data Based on Deep Belief Network. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 2381–2392. https://doi.org/10.1109/JSTARS.2015.2388577.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 117, 11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014.

D'souza, R.N., Huang, P.-Y., Yeh, F.-C., 2020. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size. Sci. Rep. 10, 834. https://doi.org/10.1038/s41598-020-57866-2.

M. Dehghani J. Djolonga B. Mustafa P. Padlewski J. Heek J. Gilmer A. Steiner M. Caron R. Geirhos I. Alabdulmohsin R. Jenatton L. Beyer M. Tschannen A. Arnab X. Wang C. Riquelme M. Minderer J. Puigcerver U. Evci M. Kumar S. van Steenkiste G.F. Elsayed A. Mahendran F. Yu A. Oliver F. Huot J. Bastings M.P. Collier A. Gritsenko V. Birodkar C. Vasconcelos Y. Tay T. Mensink A. Kolesnikov F. Pavetić D. Tran T. Kipf M. Lučić X. Zhai D. Keysers J. Harmsen N. Houlsby Scaling Vision Transformers to 22 Billion Parameters 2023 https://doi.org/10.48550/arXiv.2302.05442.

Ding, C., Chen, Y., Li, R., Wen, D., Xie, X., Zhang, L., Wei, W., Zhang, Y., 2022. Integrating Hybrid Pyramid Feature Fusion and Coordinate Attention for Effective Small Sample Hyperspectral Image Classification. Remote Sens. 14, 2355. https://doi.org/10.3390/rs14102355.

Doan, N.A.K., Polifke, W., Magri, L., 2019. Physics-Informed Echo State Networks for Chaotic Systems Forecasting, in: Rodrigues, J.M.F., Cardoso, P.J.S., Monteiro, J., Lam, R., Krzhizhanovskaya, V.V., Lees, M.H., Dongarra, J.J., Sloot, P.M.A. (Eds.), Computational Science – ICCS 2019, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 192–198. https://doi.org/10.1007/978-3-030-22747-0_15.

X. Dong L.A. Tuan M. Lin S. Yan H. Zhang How Should Pre-Trained Language Models Be Fine-Tuned Towards Adversarial Robustness? 2021 https://doi.org/10.48550/arXiv.2112.11668.

A. Dosovitskiy L. Beyer A. Kolesnikov D. Weissenborn X. Zhai T. Unterthiner M. Dehghani M. Minderer G. Heigold S. Gelly J. Uszkoreit N. Houlsby An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021 https://doi.org/10.48550/arXiv.2010.11929.

S.S. Du Y. Wang X. Zhai S. Balakrishnan R. Salakhutdinov A. Singh How Many Samples are Needed to Estimate a Convolutional or Recurrent Neural Network? 2019 https://doi.org/10.48550/arXiv.1805.07883.

Dyrmann, M., Karstoft, H., Midtiby, H.S., 2016. Plant species classification using deep convolutional neural network. Biosyst. Eng. 151, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024.

El Zini, J., Rizk, Y., Awad, M., 2020. A Deep Transfer Learning Framework for Seismic Data Analysis: A Case Study on Bright Spot Detection. IEEE Trans. Geosci. Remote Sens. 58, 3202–3212. https://doi.org/10.1109/TGRS.2019.2950888.

N. Elhage T. Hume C. Olsson N. Schiefer T. Henighan S. Kravec Z. Hatfield-Dodds R. Lasenby D. Drain C. Chen R. Grosse S. McCandlish J. Kaplan D. Amodei M. Wattenberg C. Olah Toy Models of Superposition. https:// 2022 doi.org/10.48550/arXiv.2209.10652.

T. Elsken J.-H. Metzen F. Hutter Simple And Efficient Architecture Search for Convolutional Neural Networks 2017 https://doi.org/10.48550/arXiv.1711.04528.

Fang, Y., Wang, J., Ou, X., Ying, H., Hu, C., Zhang, Z., Hu, W., 2021. The impact of training sample size on deep learning-based organ auto-segmentation for head-and-neck patients. Phys. Med. Biol. 66, 185012 https://doi.org/10.1088/1361-6560/ac2206.

S. Farquhar Y. Gal T. Rainforth On Statistical Bias In Active Learning: How and When to Fix It. Presented at the International Conference on Learning Representations 2020.

Feng, Y., Chen, J., Zhang, T., He, S., Xu, E., Zhou, Z., 2022b. Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis. ISA Trans. 120, 383–401. https://doi.org/10.1016/j.isatra.2021.03.013.

Feng, F., Zhang, Y., Zhang, J., Liu, B., 2022a. Small Sample Hyperspectral Image Classification Based on Cascade Fusion of Mixed Spatial-Spectral Features and Second-Order Pooling. Remote Sens. 14, 505. https://doi.org/10.3390/rs14030505.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. PLOS ONE 12, e0170478.

Fossan, F.E., Müller, L.O., Sturdy, J., Bråten, A.T., Jørgensen, A., Wiseth, R., Hellevik, L. R., 2021. Machine learning augmented reduced-order models for FFR-prediction. Comput. Methods Appl. Mech. Eng. 384, 113892 https://doi.org/10.1016/j.cma.2021.113892.

Freeman, D., Gupta, S., Smith, D.H., Maja, J.M., Robbins, J., Owen, J.S., Peña, J.M., de Castro, A.I., 2019. Watson on the Farm: Using Cloud-Based Artificial Intelligence to Identify Early Indicators of Water Stress. Remote Sens. 11, 2645. https://doi.org/10.3390/rs11222645.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., Obersteiner, M., 2012. Geo-Wiki: An online platform for improving global land cover. Environ. Model. Softw. 31, 110–123. https://doi.org/10.1016/j.envsoft.2011.11.015.

Fruth, T., Lenzen, C., Gross, E., Mrowka, F., 2018. The EnMAP Mission Planning System, in: 2018 SpaceOps Conference, SpaceOps Conferences. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2018-2525.

Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N., 2022. Ensemble deep learning: A review. Eng. Appl. Artif. Intell. 115, 105151 https://doi.org/10.1016/j.engappai.2022.105151.

Gao, K., Liu, B., Yu, X., Zhang, P., Tan, X., Sun, Y., 2021b. Small sample classification of hyperspectral image using model-agnostic meta-learning algorithm and convolutional neural network. Int. J. Remote Sens. 42, 3090–3122. https://doi.org/10.1080/01431161.2020.1864060.

Gao, H., Xiong, X., Cao, L., Yu, D., Yang, G., Yang, L., 2021a. Pixel-Level Prediction for Ocean Remote Sensing Image Features Fusion Based on Global and Local Semantic Relations. IEEE Access 9, 11644–11654. https://doi.org/10.1109/ACCESS.2021.3052021.

Gibson, R., Danaher, T., Hehir, W., Collins, L., 2020. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. Remote Sens. Environ. 240, 111702 https://doi.org/10.1016/j.rse.2020.111702.

I.J. Goodfellow J. Pouget-Abadie M. Mirza B. Xu D. Warde-Farley S. Ozair A. Courville Y. Bengio Generative Adversarial Networks. https:// 2014 doi.org/10.48550/arXiv.1406.2661.

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., Herrera, F., 2017. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus lotus as Case Study. Remote Sens. 9, 1220. https://doi.org/10.3390/rs9121220.

Habib, A.-R., Xu, Y., Bock, K., Mohanty, S., Sederholm, T., Weeks, W.B., Dodhia, R., Ferres, J.L., Perry, C., Sacks, R., Singh, N., 2023. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Sci. Rep. 13, 5368. https://doi.org/10.1038/s41598-023-31921-0.

Hamdy, W., Ismail, A., Awad, W.A., Ibrahim, A.H., Hassanien, A.E., 2023. An Optimized Ensemble Deep Learning Model for Predicting Plant miRNA–lncRNA Based on Artificial Gorilla Troops Algorithm. Sensors 23, 2219. https://doi.org/10.3390/s23042219.

M. Han H. Wu Z. Chen M. Li X. Zhang A survey of multi-label classification based on supervised and semi-supervised learning 2022 J. Mach. Learn. Cybern Int 10.1007/s13042-022-01658-9.

Han, Y., Wei, C., Zhou, R., Hong, Z., Zhang, Y., Yang, S., 2020. Combining 3D-CNN and Squeeze-and-Excitation Networks for Remote Sensing Sea Ice Image Classification. Math. Probl. Eng. 2020, e8065396.

Han, J., Zhang, D., Cheng, G., Guo, L., Ren, J., 2015. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning. IEEE Trans. Geosci. Remote Sens. 53, 3325–3337. https://doi.org/10.1109/TGRS.2014.2374218.

J. He A. Kortylewski S. Yang S. Liu C. Yang C. Wang A. Yuille Rethinking Re-Sampling in Imbalanced Semi-Supervised Learning 2021 https://doi.org/10.48550/arXiv.2106.00209.

Ho, B., Kocer, B.B., Kovac, M., 2022. Vision based crown loss estimation for individual trees with remote aerial robots. ISPRS J. Photogramm. Remote Sens. 188, 75–88. https://doi.org/10.1016/j.isprsjprs.2022.04.002.

Hong-Yu, F., Yun-Kai, Y., Wei, W., Ao, L., Ming-Zhi, X., Xihong, G., Wei, S., Guo-Xian, C., 2023. Ramie Plant Counting Based on UAV Remote Sensing Technology and Deep Learning. J. Nat. Fibers 20, 2159610. https://doi.org/10.1080/15440478.2022.2159610.

T. Hospedales A. Antoniou P. Micaelli A. Storkey Meta-Learning in Neural Networks: A Survey 2020 https://doi.org/10.48550/arXiv.2004.05439.

Hou, Y., Ren, H., Lv, Q., Wu, L., Yang, X., Quan, Y., 2022. Radar-Jamming Classification in the Event of Insufficient Samples Using Transfer Learning. Symmetry 14, 2318. https://doi.org/10.3390/sym14112318.

W.-N. Hsu B. Bolte Y.-H.-H. Tsai K. Lakhotia R. Salakhutdinov A. Mohamed HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units 2021 https://doi.org/10.48550/arXiv.2106.07447.

H. Hu J. Liu G. Chen Y. Zhao Y. Men P. Wang Driver identification through vehicular CAN bus data: An ensemble deep learning approach IET Intell. Transp. Syst. n/a. n.d. 10.1049/itr2.12311.

M. Iman K. Rasheed H.R. Arabnia A Review of Deep Transfer Learning and Recent Advancements 2022 https://doi.org/10.48550/arXiv.2201.09679.

Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., Yu, S., 2021. A survey: Deep learning for hyperspectral image classification with few labeled samples. Neurocomputing 448, 179–204. https://doi.org/10.1016/j.neucom.2021.03.035.

L. Jing Y. Tian Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey 2019 https://doi.org/10.48550/arXiv.1902.06162.

Jozdani, S., Chen, D., Chen, W., Leblanc, S.G., Prévost, C., Lovitt, J., He, L., Johnson, B. A., 2021. Leveraging Deep Neural Networks to Map Caribou Lichen in High-Resolution Satellite Images Based on a Small-Scale. Noisy UAV-Derived Map. Remote Sens. 13, 2658. https://doi.org/10.3390/rs13142658.

Kang, X., Zhuo, B., Duan, P., 2019. Semi-supervised deep learning for hyperspectral image classification. Remote Sens. Lett. 10, 353–362. https://doi.org/10.1080/2150704X.2018.1557787.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-Informed Machine Learning. Nat. Rev. Phys. 3, 422–440. https://doi.org/10.1038/s42254-021-00314-5.

Kashefi, A., Mukerji, T., 2022. Physics-informed PointNet: A deep learning solver for steady-state incompressible flows and thermal fields on multiple sets of irregular geometries. J. Comput. Phys. 468, 111510 https://doi.org/10.1016/j.jcp.2022.111510.

Kato, S., Miyamoto, H., Amici, S., Oda, A., Matsushita, H., Nakamura, R., 2021. Automated classification of heat sources detected using SWIR remote sensing. Int. J. Appl. Earth Obs. Geoinformation 103, 102491. https://doi.org/10.1016/j.jag.2021.102491.

Katsaragakis, M., Papadopoulos, L., Konijnenburg, M., Catthoor, F., Soudris, D., 2020. Memory Footprint Optimization Techniques for Machine Learning Applications in Embedded Systems, in: 2020 IEEE International Symposium on Circuits and Systems (ISCAS). Presented at the 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4. https://doi.org/10.1109/ISCAS45731.2020.9181038.

Kattenborn, T., Lopatin, J., Förster, M., Braun, A.C., Fassnacht, F.E., 2019. UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. Remote Sens. Environ. 227, 61–73. https://doi.org/10.1016/j.rse.2019.03.025.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D., Dormann, C.F., 2022. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open J. Photogramm. Remote Sens. 5, 100018 https://doi.org/10.1016/j.ophoto.2022.100018.

R. Keshari M. Vatsa R. Singh A. Noore Learning Structure and Strength of CNN Filters for Small Sample Size Training 2018 https://doi.org/10.48550/arXiv.1803.11405.

R. Keshari S. Ghosh S. Chhabra M. Vatsa R. Singh Unravelling Small Sample Size Problems in the Deep Learning World 2020 https://doi.org/10.48550/arXiv.2008.03522.

Khan, A.H., Fraz, M.M., Shahzad, M., 2021. Deep Learning Based Land Cover and Crop Type Classification: A Comparative Study, in: 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). Presented at the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), pp. 1–6. https://doi.org/10.1109/ICoDT252288.2021.9441483.

Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53, 5455–5516. https://doi.org/10.1007/s10462-020-09825-6.

Kim, W.-S., Lee, D.-H., Kim, T., Kim, H., Sim, T., Kim, Y.-J., 2021. Weakly Supervised Crop Area Segmentation for an Autonomous Combine Harvester. Sensors 21, 4801. https://doi.org/10.3390/s21144801.

P. Kokol M. Kokol S. Zagoranski Machine learning on small size samples: A synthetic knowledge synthesis Sci. Prog. 105 2022 00368504211029777 10.1177/00368504211029777.

Koller, C., Shahrad, M., Zhu, X.X., 2022. Uncertainty-Guided Representation Learning in Local Climate Zone Classification, in: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 183–186. https://doi.org/10.1109/IGARSS46834.2022.9883897.

Krishnan, R., Rajpurkar, P., Topol, E.J., 2022. Self-supervised learning in medicine and healthcare. Nat. Biomed. Eng. 6, 1346–1352. https://doi.org/10.1038/s41551-022-00914-1.

Krois, J., Garcia Cantu, A., Chaurasia, A., Patil, R., Chaudhari, P.K., Gaudin, R., Gehrung, S., Schwendicke, F., 2021. Generalizability of deep learning models for dental image analysis. Sci. Rep. 11, 6102. https://doi.org/10.1038/s41598-021-85454-5.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., Ferrari, V., 2020. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. Int. J. Comput. vis. 128, 1956–1981. https://doi.org/10.1007/s11263-020-01316-z.

Lampert, C.H., Nickisch, H., Harmeling, S., 2014. Attribute-Based Classification for Zero-Shot Visual Object Categorization. IEEE Trans. Pattern Anal. Mach. Intell. 36, 453–465. https://doi.org/10.1109/TPAMI.2013.140.

Lange, M., Feilhauer, H., Kühn, I., Doktor, D., 2022. Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. Remote Sens. Environ. 277, 112888 https://doi.org/10.1016/j.rse.2022.112888.

Lauriola, I., Lavelli, A., Aiolli, F., 2022. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. Neurocomputing 470, 443–456. https://doi.org/10.1016/j.neucom.2021.05.103.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., Bretagnolle, V., 2014. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Glob. Ecol. Biogeogr. 23, 811–820. https://doi.org/10.1111/geb.12161.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Leitão, P.J., Schwieder, M., Pötzschner, F., Pinto, J.R.R., Teixeira, A.M.C., Pedroni, F., Sanchez, M., Rogass, C., van der Linden, S., Bustamante, M.M.C., Hostert, P., 2018. From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. Ecosphere 9, e02298.

Li, L., Cao, G., Liu, J., Tong, Y., 2022b. Efficient Detection in Aerial Images for Resource-Limited Satellites. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2020.3046739.

X. Li M. Jia M.T. Islam L. Yu L. Xing Self-supervised Feature Learning via Exploiting Multi-modal Data for Retinal Disease Diagnosis 2020 https://doi.org/10.48550/arXiv.2007.11067.

Li, Z., Liu, M., Chen, Y., Xu, Y., Li, W., Du, Q., 2022d. Deep Cross-Domain Few-Shot Learning for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–18. https://doi.org/10.1109/TGRS.2021.3057066.

Li, H., Lu, J., Tian, G., Yang, H., Zhao, J., Li, N., 2022a. Crop Classification Based on GDSSM-CNN Using Multi-Temporal RADARSAT-2 SAR with Limited Labeled Data. Remote Sens. 14, 3889. https://doi.org/10.3390/rs14163889.

Li, Y., Si, Y., Tong, Z., He, L., Zhang, J., Luo, S., Gong, Y., 2022c. MQANet: Multi-Task Quadruple Attention Network of Multi-Object Semantic Segmentation from Remote Sensing Images. Remote Sens. 14, 6256. https://doi.org/10.3390/rs14246256.

Li, M., Stein, A., 2020. Mapping Land Use from High Resolution Satellite Images by Exploiting the Spatial Arrangement of Land Cover Objects. Remote Sens. 12, 4158. https://doi.org/10.3390/rs12244158.

Li, Y., Yang, J., 2021. Meta-learning baselines and database for few-shot classification in agriculture. Comput. Electron. Agric. 182, 106055 https://doi.org/10.1016/j.compag.2021.106055.

T.-Y. Lin M. Maire S. Belongie L. Bourdev R. Girshick J. Hays P. Perona D. Ramanan C.L. Zitnick P. Dollár Microsoft COCO: Common Objects in Context 2015 https://doi.org/10.48550/arXiv.1405.0312.

V. Lioutas Mapping Low-Resolution Images To Multiple High-Resolution Images Using Non-Adversarial Mapping 2020 https://doi.org/10.48550/arXiv.2006.11708.

Little, M.A., Varoquaux, G., Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., Kording, K.P., 2017. Using and understanding cross-validation strategies. Perspectives on Saeb et al. GigaScience 6, gix020. https://doi.org/10.1093/gigascience/gix020.

Littlewood, K.E., Shilling, A.M., Stemland, C.J., Wright, E.B., Kirk, M.A., 2013. High-fidelity simulation is superior to case-based discussion in teaching the management of shock. Med. Teach. 35, e1003–e1010. https://doi.org/10.3109/0142159X.2012.733043.

Liu, Y., Chen, P., Zhang, J., Liu, N., Liu, Y., 2022c. Weakly Supervised Ternary Stream Data Augmentation Fine-Grained Classification Network for Identifying Acute Lymphoblastic Leukemia. Diagnostics 12, 16. https://doi.org/10.3390/diagnostics12010016.

Liu, S., Deng, W., 2015. Very deep convolutional neural network based image classification using small training sample size. In: In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). Presented at the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 730–734. https://doi.org/10.1109/ACPR.2015.7486599.

B. Liu Y. Wei Y. Zhang Q. Yang Deep Neural Networks for High Dimension, Low Sample Size Data 2287–2293 2017 https://doi.org/10.24963/ijcai.2017/318.

X. Liu A. Sinha M. Ishii G.D. Hager A. Reiter R.H. Taylor M. Unberath Dense Depth Estimation in Monocular Endoscopy with Self-supervised Learning Methods 2019 https://doi.org/10.48550/arXiv.1902.07766.

Liu, Z., Guan, R., Hu, J., Chen, W., Li, X., 2022d. Remote Sensing Scene Data Generation Using Element Geometric Transformation and GAN-Based Texture Synthesis. Appl. Sci. 12, 3972. https://doi.org/10.3390/app12083972.

Liu, F., Kijowski, R., El Fakhri, G., Feng, L., 2021. Magnetic resonance parameter mapping using model-guided self-supervised deep learning. Magn. Reson. Med. 85, 3211–3226. https://doi.org/10.1002/mrm.28659.

Liu, X., Wang, Y., Liu, Q., 2018. Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening. In: IEEE International Conference on Image Processing (ICIP). Presented at the 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 873–877. https://doi.org/10.1109/ICIP.2018.8451049.

Liu, B., Yu, X., Yu, A., Zhang, P., Wan, G., Wang, R., 2019a. Deep Few-Shot Learning for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 57, 2290–2304. https://doi.org/10.1109/TGRS.2018.2872830.

Liu, B., Gao, K., Yu, A., Ding, L., Qiu, C., Li, J., 2022a. ES2FL: Ensemble Self-Supervised Feature Learning for Small Sample Classification of Hyperspectral Images. Remote Sens. 14, 4236. https://doi.org/10.3390/rs14174236.

Liu, P., Zhang, H., Eom, K.B., 2017b. Active Deep Learning for Classification of Hyperspectral Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 712–724. https://doi.org/10.1109/JSTARS.2016.2598859.

Liu, J., Zhang, K., Wu, S., Shi, H., Zhao, Y., Sun, Y., Zhuang, H., Fu, E., 2022b. An Investigation of a Multidimensional CNN Combined with an Attention Mechanism Model to Resolve Small-Sample Problems in Hyperspectral Image Classification. Remote Sens. 14, 785. https://doi.org/10.3390/rs14030785.

Liu, Y., Zheng, F., 2017. Object-oriented and multi-scale target classification and recognition based on hierarchical ensemble learning. Comput. Electr. Eng. 62, 538–554. https://doi.org/10.1016/j.compeleceng.2016.12.026.

M.A. Lones How to avoid machine learning pitfalls: a guide for academic researchers 2023 https://doi.org/10.48550/arXiv.2108.02497.

Lu, S., Guo, J., Liu, S., Yang, B., Liu, M., Yin, L., Zheng, W., 2022. An Improved Algorithm of Drift Compensation for Olfactory Sensors. Appl. Sci. 12, 9529. https://doi.org/10.3390/app12199529.

Lv, Q., Quan, Y., Feng, W., Sha, M., Dong, S., Xing, M., 2022. Radar Deception Jamming Recognition Based on Weighted Ensemble CNN With Transfer Learning. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2021.3129645.

W. Ma F. Cheng Y. Xu Q. Wen Y. Liu Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy 2019 https://doi.org/10.48550/arXiv.1901.10819.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019a. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Ma, W., Liu, Y., 2020. A data-efficient self-supervised deep learning model for design and characterization of nanophotonic structures. Sci. China Phys. Mech. Astron. 63, 284212 https://doi.org/10.1007/s11433-020-1575-2.

Malambo, L., Popescu, S., Ku, N.-W., Rooney, W., Zhou, T., Moore, S., 2019. A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. Remote Sens. 11, 2939. https://doi.org/10.3390/rs11242939.

Martino, L., Fritz, M., 2008. New Insight into Land Cover and Land Use in Europe, No. 33. European Communities, Luxembourg.

Meier, S., Heijman, J., 2022. Commentary: EP-PINNs: Cardiac electrophysiology characterisation using physics-informed neural networks. Front. Cardiovasc, Med, p. 9.

Mohamed, A., Lee, H., Borgholt, L., Havtorn, J.D., Edin, J., Igel, C., Kirchhoff, K., Li, S.-W., Livescu, K., Maaløe, L., Sainath, T.N., Watanabe, S., 2022. Self-Supervised Speech Representation Learning: A Review. IEEE J. Sel. Top. Signal Process. 16, 1179–1210. https://doi.org/10.1109/JSTSP.2022.3207050.

Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F.A., Schwarz, L.-M., Ghazaryan, G., Schultz, J., Malenovský, Z., Dubovyk, O., 2022. Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. Remote Sens. Environ. 282, 113262 https://doi.org/10.1016/j.rse.2022.113262.

Murthy, K., Shearn, M., Smiley, B.D., Chau, A.H., Levine, J., Robinson, M.D., 2014. SkySat-1: very high-resolution imagery from a small satellite, in: Sensors, Systems, and Next-Generation Satellites XVIII. Presented at the Sensors, Systems, and Next-Generation Satellites XVIII, SPIE, pp. 367–378. https://doi.org/10.1117/12.2074163.

Nakarmi, S., Santosh, K., 2023. Active Learning to Minimize the Risk from Future Epidemics. In: In: 2023 IEEE Conference on Artificial Intelligence (CAI). Presented at the 2023 IEEE Conference on Artificial Intelligence (CAI), pp. 329–330. https://doi.org/10.1109/CAI54212.2023.00145.

P. Nakkiran G. Kaplun Y. Bansal T. Yang B. Barak I. Sutskever Deep Double Descent: Where Bigger Models and More Data Hurt 2019 https://doi.org/10.48550/arXiv.1912.02292.

Naushad, R., Kaur, T., Ghaderpour, E., 2021. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. Sensors 21, 8083. https://doi.org/10.3390/s21238083.

Odebiri, O., Mutanga, O., Odindi, J., 2022. Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data. Geoderma 411, 115695. https://doi.org/10.1016/j.geoderma.2022.115695.

Osco, L.P., Marcato Junior, J., Marques Ramos, A.P., de Castro Jorge, L.A., Fatholahi, S. N., de Andrade Silva, J., Matsubara, E.T., Pistori, H., Gonçalves, W.N., Li, J., 2021. A review on deep learning in UAV remote sensing. Int. J. Appl. Earth Obs. Geoinformation 102, 102456. https://doi.org/10.1016/j.jag.2021.102456.

Pan, B., Shi, Z., Zhang, N., Xie, S., 2016. Hyperspectral Image Classification Based on Nonlinear Spectral-Spatial Network. IEEE Geosci. Remote Sens. Lett. 13, 1782–1786. https://doi.org/10.1109/LGRS.2016.2608963.

Pan, B., Shi, Z., Xu, X., 2018. MugNet: Deep learning for hyperspectral image classification using limited samples. ISPRS J. Photogramm. Remote Sens. Deep Learning RS Data 145, 108–119. https://doi.org/10.1016/j.isprsjprs.2017.11.003.

Panagos, P., Ballabio, C., Yigini, Y., Dunbar, M.B., 2013. Estimating the soil organic carbon content for European NUTS2 regions based on LUCAS data collection. Sci. Total Environ. 442, 235–246. https://doi.org/10.1016/j.scitotenv.2012.10.017.

Pang, Y., Shi, Y., Gao, S., Jiang, F., Veeranampalayam-Sivakumar, A.-N., Thompson, L., Luck, J., Liu, C., 2020. Improved crop row detection with deep neural network for early-season maize stand count in UAV imagery. Comput. Electron. Agric. 178, 105766 https://doi.org/10.1016/j.compag.2020.105766.

Pardakhti, M., Mandal, N., Ma, A.W.K., Yang, Q., 2021. Practical Active Learning with Model Selection for Small Data. In: IEEE International Conference on Machine Learning and Applications (ICMLA). Presented at the 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1647–1653. https://doi.org/10.1109/ICMLA52953.2021.00263.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540. https://doi.org/10.1038/s41467-020-18321-y.

Pöppelbaum, J., Chadha, G.S., Schwung, A., 2022. Contrastive learning based self-supervised time-series analysis. Appl. Soft Comput. 117, 108397 https://doi.org/10.1016/j.asoc.2021.108397.

A. Power Y. Burda H. Edwards I. Babuschkin V. Misra Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets 2022 https://doi.org/10.48550/arXiv.2201.02177.

Prusa, J., Khoshgoftaar, T.M., Seliya, N., 2015. The Effect of Dataset Size on Training Tweet Sentiment Classifiers. In: In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). Presented at the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 96–102. https://doi.org/10.1109/ICMLA.2015.22.

Putra, Y.C., Wijayanto, A.W., 2023. Automatic detection and counting of oil palm trees using remote sensing and object-based deep learning. Remote Sens. Appl. Soc. Environ. 29, 100914 https://doi.org/10.1016/j.rsase.2022.100914.

Qi, K., Liu, W., Yang, C., Guan, Q., Wu, H., 2017. Multi-Task Joint Sparse and Low-Rank Representation for the Scene Classification of High-Resolution Remote Sensing Image. Remote Sens. 9, 10. https://doi.org/10.3390/rs9010010.

Qiu, T., He, H., Liang, X., Chen, F., Chen, Z., Liu, Y., 2022. Using different training strategies for urban land-use classification based on convolutional neural networks. Front. Environ, Sci, p. 10.

Quan, Y., Yu, A., Guo, W., Lu, X., Jiang, B., Zheng, S., He, P., 2023. Unified building change detection pre-training method with masked semantic annotations. Int. J. Appl. Earth Obs. Geoinformation 120, 103346. https://doi.org/10.1016/j.jag.2023.103346.

M. Raissi P. Perdikaris G.E. Karniadakis Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations 2017 https://doi.org/10.48550/arXiv.1711.10561.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

Rangnekar, A., Mokashi, N., Ientilucci, E.J., Kanan, C., Hoffman, M.J., 2020. AeroRIT: A New Scene for Hyperspectral Image Analysis. Ieee Trans. Geosci. Remote Sens. 58, 8116–8124. https://doi.org/10.1109/TGRS.2020.2987199.

Rani, V., Nabi, S.T., Kumar, M., Mittal, A., Kumar, K., 2023. Self-supervised Learning: A Succinct Review. Arch. Comput. Methods Eng. https://doi.org/10.1007/s11831-023-09884-2.

Rao, M., Tang, P., Zhang, Z., 2019. Spatial-Spectral Relation Network for Hyperspectral Image Classification With Limited Training Samples. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 5086–5100. https://doi.org/10.1109/JSTARS.2019.2957047.

Rawat, W., Wang, Z., 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Comput. 29, 2352–2449. https://doi.org/10.1162/neco_a_00990.

Reedha, R., Dericquebourg, E., Canals, R., Hafiane, A., 2022. Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. Remote Sens. 14, 592. https://doi.org/10.3390/rs14030592.

P. Ren Y. Xiao X. Chang P.-Y. Huang Z. Li B.B. Gupta X. Chen X. Wang A Survey of Deep Active Learning 2021 https://doi.org/10.48550/arXiv.2009.00236.

Rigge, M., Homer, C., Cleeves, L., Meyer, D.K., Bunde, B., Shi, H., Xian, G., Schell, S., Bobo, M., 2020. Quantifying Western U.S. Rangelands as Fractional Components with Multi-Resolution Remote Sensing and In Situ Data. Remote Sens. 12, 412. https://doi.org/10.3390/rs12030412.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929. https://doi.org/10.1111/ecog.02881.

Romão, M.C., Castro, N.F., Pedro, R., Vale, T., 2020. Transferability of deep learning models in searches for new physics at colliders. Phys. Rev. D 101, 035042. https://doi.org/10.1103/PhysRevD.101.035042.

Rostami, M., Berahmand, K., Forouzandeh, S., 2020. A novel method of constrained feature selection by the measurement of pairwise constraints uncertainty. J. Big Data 7, 83. https://doi.org/10.1186/s40537-020-00352-3.

Ruan, H., Wang, Y., Li, X., Qin, Y., Tang, B., 2022. An enhanced non-local weakly supervised fault diagnosis method for rotating machinery. Measurement 189, 110433. https://doi.org/10.1016/j.measurement.2021.110433.

S. Ruder An Overview of Multi-Task Learning in Deep Neural Networks 2017 https://doi.org/10.48550/arXiv.1706.05098.

O. Russakovsky J. Deng H. Su J. Krause S. Satheesh S. Ma Z. Huang A. Karpathy A. Khosla M. Bernstein A.C. Berg L. Fei-Fei ImageNet Large Scale Visual Recognition Challenge 2015 https://doi.org/10.48550/arXiv.1409.0575.

Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., Herrera, F., 2019. Detection of Fir Trees (Abies sibirica) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. Remote Sens. 11, 643. https://doi.org/10.3390/rs11060643.

Safonova, A., Guirado, E., Maglinets, Y., Alcaraz-Segura, D., Tabik, S., 2021. Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN. Sensors 21, 1617. https://doi.org/10.3390/s21051617.

Safonova, A., Hamad, Y., Alekhina, A., Kaplun, D., 2022. Detection of Norway Spruce Trees (Picea Abies) Infested by Bark Beetle in UAV Images Using YOLOs Architectures. IEEE Access 10, 10384–10392. https://doi.org/10.1109/ACCESS.2022.3144433.

Sagan, V., Maimaitijiang, M., Bhadra, S., Maimaitiyiming, M., Brown, D.R., Sidike, P., Fritschi, F.B., 2021. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. ISPRS J. Photogramm. Remote Sens. 174, 265–281. https://doi.org/10.1016/j.isprsjprs.2021.02.008.

Sanlang, S., Cao, S., Du, M., Mo, Y., Chen, Q., He, W., 2021. Integrating Aerial LiDAR and Very-High-Resolution Images for Urban Functional Zone Mapping. Remote Sens. 13, 2573. https://doi.org/10.3390/rs13132573.

Sapkota, B.B., Popescu, S., Rajan, N., Leon, R.G., Reberg-Horton, C., Mirsky, S., Bagavathiannan, M.V., 2022. Use of synthetic images for training a deep learning model for weed detection and biomass estimation in cotton. Sci. Rep. 12, 19580. https://doi.org/10.1038/s41598-022-23399-z.

Sawada, Y., Koike, T., Ikoma, E., Kitsuregawa, M., 2020. Monitoring and Predicting Agricultural Droughts for a Water-Limited Subcontinental Region by Integrating a Land Surface Model and Microwave Remote Sensing. IEEE Trans. Geosci. Remote Sens. 58, 14–33. https://doi.org/10.1109/TGRS.2019.2927342.

Schat, E., van de Schoot, R., Kouw, W.M., Veen, D., Mendrik, A.M., 2020. The data representativeness criterion: Predicting the performance of supervised classification based on data set similarity. PLOS ONE 15, e0237009.

Settles, B., 2009. Active Learning Literature Survey. University of Wisconsin-Madison Department of Computer Sciences. Technical Report).

Shah, V., Flood, M.W., Grimm, B., Dixon, P.C., 2022. Generalizability of deep learning models for predicting outdoor irregular walking surfaces. J. Biomech. 139, 111159 https://doi.org/10.1016/j.jbiomech.2022.111159.

Sharma, P., Amhia, H., Sharma, S.D., 2021. Performance analysis of pre-trained transfer learning models for the classification of the rolling bearing faults. J. Phys. Conf. Ser. 2070, 012141 https://doi.org/10.1088/1742-6596/2070/1/012141.

Sharma, A., Ghosh, A., Joshi, P.K., 2013. Decision tree approach for classification of remotely sensed satellite data using open source support. J. Earth Syst. Sci. 122, 1237–1247. https://doi.org/10.1007/s12040-013-0339-2.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K., 2018. Satellite Remote Sensing for Water Resources Management: Potential for Supporting Sustainable Development in Data-Poor Regions. Water Resour. Res. 54, 9724–9758. https://doi.org/10.1029/2017WR022437.

Shim, S., Kim, J., Lee, S.-W., Cho, G.-C., 2022. Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. Autom. Constr. 135, 104139 https://doi.org/10.1016/j.autcon.2022.104139.

Shin, J., Khim, B.-K., Jang, L.-H., Lim, J., Jo, Y.-H., 2022. Convolutional neural network model for discrimination of harmful algal bloom (HAB) from non-HABs using Sentinel-3 OLCI imagery. ISPRS J. Photogramm. Remote Sens. 191, 250–262. https://doi.org/10.1016/j.isprsjprs.2022.07.012.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.

Song, L., Feng, Z., Yang, S., Zhang, X., Jiao, L., 2022. Self-Supervised Assisted Semi-Supervised Residual Network for Hyperspectral Image Classification. Remote Sens. 14, 2997. https://doi.org/10.3390/rs14132997.
Sosnin, S., Vashurina, M., Withnall, M., Karpov, P., Fedorov, M., Tetko, I.V., 2019. A Survey of Multi-task Learning Methods in Chemoinformatics. Mol. Inform. 38, 1800108. https://doi.org/10.1002/minf.201800108.
Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B., Meygret, A., 2012. Overview Of Sentinel-2, in: 2012 IEEE International Geoscience and Remote Sensing Symposium. Presented at the 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 1707–1710. https://doi.org/10.1109/IGARSS.2012.6351195.
J.-C. Su S. Maji B. Hariharan When Does Self-supervision Improve Few-shot Learning? 2020 https://doi.org/10.48550/arXiv.1910.03560.
Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2019. BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 5901–5904. https://doi.org/10.1109/IGARSS.2019.8900532.
Sztahó, D., Szaszák, G., Beke, A., 2021. Deep Learning Methods in Speaker Recognition: A Review. Period. Polytech. Electr. Eng. Comput. Sci. 65, 310–328. https://doi.org/10.3311/PPee.17024.
Taiwo, B.E., Kafy, A.-A., Samuel, A.A., Rahaman, Z.A., Ayowole, O.E., Shahrier, M., Duti, B.M., Rahman, M.T., Peter, O.T., Abosede, O.O., 2023. Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques. Environ. Sustain. Indic. 18, 100248 https://doi.org/10.1016/j.indic.2023.100248.
Takezoe, R., Liu, X., Mao, S., Chen, M.T., Feng, Z., Zhang, S., Wang, X., 2022. Deep Active Learning for Computer Vision: past and Future. https://doi.org/10.1561/116.00000057.
Thanh Noi, P., Kappas, M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors 18, 18. https://doi.org/10.3390/s18010018.
Tseng, K.-K., Zhang, R., Chen, C.-M., Hassan, M.M., 2021. DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service. J. Supercomput. 77, 3594–3615. https://doi.org/10.1007/s11227-020-03407-7.
Uhl, J.H., Leyk, S., Li, Z., Duan, W., Shbita, B., Chiang, Y.-Y., Knoblock, C.A., 2021. Combining Remote-Sensing-Derived Data and Historical Maps for Long-Term Back-Casting of Urban Extents. Remote Sens. 13, 3672. https://doi.org/10.3390/rs13183672.
Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PLOS ONE 14, e0224365.
Villon, S., Iovan, C., Mangeas, M., Claverie, T., Mouillot, D., Villéger, S., Vigliola, L., 2021. Automatic underwater fish species classification with limited data using few-shot learning. Ecol. Inform. 63, 101320 https://doi.org/10.1016/j.ecoinf.2021.101320.
Volpi, M., Tuia, D., 2018. Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. ISPRS J. Photogramm. Remote Sens. 144, 48–60. https://doi.org/10.1016/j.isprsjprs.2018.06.007.
Wang, Q., Du, P., Yang, J., Wang, G., Lei, J., Hou, C., 2019b. Transferred deep learning based waveform recognition for cognitive passive radar. Signal Process. 155, 259–267. https://doi.org/10.1016/j.sigpro.2018.09.038.
Wang, H., Duentsch, I., Guo, G., Khan, S.A., 2023. Special issue on small data analytics. Int. J. Mach. Learn. Cybern. 14, 1–2. https://doi.org/10.1007/s13042-022-01699-0.
Y. Wang Q. Yao J. Kwok L.M. Ni Generalizing from a Few Examples: A Survey on Few-Shot Learning 2020 https://doi.org/10.48550/arXiv.1904.05046.
S. Wang Y. Han J. Chen X. He Z. Zhang X. Liu K. Zhang Weed Density Extraction Based on Few-Shot Learning Through UAV Remote Sensing RGB and Multispectral Images in Ecological Irrigation Area. Front. Plant Sci 2022 12.
Wang, Z., Hou, C., Yue, G., Yang, Q., 2022d. Dynamic-boosting attention for self-supervised video representation learning. Appl. Intell. 52, 3143–3155. https://doi.org/10.1007/s10489-021-02440-0.
Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T., 2018a. Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. IEEE Trans. Med. Imaging 37, 1562–1573. https://doi.org/10.1109/TMI.2018.2791721.
Wang, L., Liu, Y., Di, H., Qin, C., Sun, G., Fu, Y., 2021. Semi-Supervised Dual Relation Learning for Multi-Label Classification. IEEE Trans. Image Process. 30, 9125–9135. https://doi.org/10.1109/TIP.2021.3122003.
Wang, T., Liu, H., Li, J., 2022c. Spectral-Spatial Classification of Few Shot Hyperspectral Image With Deep 3-D Convolutional Random Fourier Features Network. IEEE Trans. Geosci. Remote Sens. 60, 1–18. https://doi.org/10.1109/TGRS.2022.3185612.
Wang, S., Quan, D., Liang, X., Ning, M., Guo, Y., Jiao, L., 2018b. A deep learning framework for remote sensing image registration. ISPRS J. Photogramm. Remote Sens. Deep Learning RS Data 145, 148–164. https://doi.org/10.1016/j.isprsjprs.2017.12.012.
Wang, C., Wang, X., Bai, X., Liu, Y., Zhou, J., 2019a. Self-Supervised deep homography estimation with invertibility constraints. Pattern Recognit. Lett. 128, 355–360. https://doi.org/10.1016/j.patrec.2019.09.021.
Wang, D., Zhang, D., Yang, G., Xu, B., Luo, Y., Yang, X., 2022a. SSRNet: In-Field Counting Wheat Ears Using Multi-Stage Convolutional Neural Network. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2021.3093041.
Wang, Z., Zou, C., Cai, W., 2020b. Small Sample Classification of Hyperspectral Remote Sensing Images Based on Sequential Joint Deeping Learning Model. IEEE Access 8, 71353–71363. https://doi.org/10.1109/ACCESS.2020.2986267.
Y. Wei S. Ji Scribble-based Weakly Supervised Deep Learning for Road Surface Extraction from Remote Sensing Images 2020 https://doi.org/10.48550/arXiv.2010.13106.
Windrim, L., Bryson, M., McLean, M., Randle, J., Stone, C., 2019. Automated Mapping of Woody Debris over Harvested Forest Plantations Using UAVs, High-Resolution Imagery, and Machine Learning. Remote Sens. 11, 733. https://doi.org/10.3390/rs11060733.
Wu, X., Chen, C., Zhong, M., Wang, J., Shi, J., 2021a. COVID-AL: The diagnosis of COVID-19 with deep active learning. Med. Image Anal. 68, 101913 https://doi.org/10.1016/j.media.2020.101913.
Wu, X., Chen, Y., Li, X., Liu, X., Liu, Y., Wu, Y., Li, M., Zhou, X., Wang, C., 2022b. IDA-MIL: Classification of Glomerular with Spike-like Projections via Multiple Instance Learning with Instance-level Data Augmentation. Comput. Methods Programs Biomed. 225, 107106 https://doi.org/10.1016/j.cmpb.2022.107106.
Wu, M., Li, C., Yao, Z., 2022a. Deep Active Learning for Computer Vision Tasks: Methodologies, Applications, and Challenges. Appl. Sci. 12, 8103. https://doi.org/10.3390/app12168103.
Wu, Z., Liu, J., Yang, J., Xiao, Z., Xiao, L., 2021b. Composite kernel learning network for hyperspectral image classification. Int. J. Remote Sens. 42, 6066–6089. https://doi.org/10.1080/01431161.2021.1934599.
Wu, C., Wang, M., Gao, L., Song, W., Tian, T., Choo, K.-K.-R., 2019a. Convolutional Neural Network with Expert Knowledge for Hyperspectral Remote Sensing Imagery Classification. KSII Trans. Internet Inf. Syst. 13, 3917–3941.
Wu, X., Xiao, Q., Wen, J., You, D., Hueni, A., 2019b. Advances in quantitative remote sensing product validation: Overview and current status. Earth-Sci. Rev. 196, 102875 https://doi.org/10.1016/j.earscirev.2019.102875.
Wu, X., Zhang, X., Wang, N., Cen, Y., 2019c. Joint Sparse and Low-Rank Multi-Task Learning with Extended Multi-Attribute Profile for Hyperspectral Target Detection. Remote Sens. 11, 150. https://doi.org/10.3390/rs11020150.
Xiong, W., Lv, Y., Cui, Y., Zhang, X., Gu, X., 2019. A Discriminative Feature Learning Approach for Remote Sensing Image Retrieval. Remote Sens. 11, 281. https://doi.org/10.3390/rs11030281.
Xu, Z., Wang, H., Yang, Y., 2021. Semi-supervised self-growing generative adversarial networks for image recognition. Multimed. Tools Appl. 80, 17461–17486. https://doi.org/10.1007/s11042-020-09602-1.
Xu, Y., Zhang, Y., Yu, C., Ji, C., Yue, T., Li, H., 2022. Residual Spatial Attention Kernel Generation Network for Hyperspectral Image Classification With Small Sample Size. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2022.3175494.
Xue, Z., Lin, H., Wang, F., 2022a. A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement. Forests 13, 1332. https://doi.org/10.3390/f13081332.
Xue, Z., Yu, X., Yu, A., Liu, B., Zhang, P., Wu, S., 2022b. Self-Supervised Feature Learning for Multimodal Remote Sensing Image Land Cover Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3190466.
Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., Zhang, Q., 2019. Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations. Environ. Int. 123, 345–357. https://doi.org/10.1016/j.envint.2018.11.075.
Yaman, B., Hosseini, S.A.H., Moeller, S., Ellermann, J., Uğurbil, K., Akçakaya, M., 2020. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. Magn. Reson. Med. 84, 3172–3191. https://doi.org/10.1002/mrm.28378.
Yan, X., Gilani, S.Z., Feng, M., Zhang, L., Qin, H., Mian, A., 2020. Self-Supervised Learning to Detect Key Frames in Videos. Sensors 20, 6941. https://doi.org/10.3390/s20236941.
Yang, S., Feng, Q., Liang, T., Liu, B., Zhang, W., Xie, H., 2018. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. Remote Sens. Environ. 204, 448–455. https://doi.org/10.1016/j.rse.2017.10.011.
Yang, W., Zhang, X., Luo, P., 2021. Transferability of Convolutional Neural Network Models for Identifying Damaged Buildings Due to Earthquake. Remote Sens. 13, 504. https://doi.org/10.3390/rs13030504.
Yao, X., Han, J., Cheng, G., Qian, X., Guo, L., 2016. Semantic Annotation of High-Resolution Satellite Images via Weakly Supervised Learning. IEEE Trans. Geosci. Remote Sens. 54, 3660–3671. https://doi.org/10.1109/TGRS.2016.2523563.
Yi, X., Walia, E., Babyn, P., 2019. Generative Adversarial Network in Medical Imaging: A Review. Med. Image Anal. 58, 101552 https://doi.org/10.1016/j.media.2019.101552.
Ying, X., 2019. An Overview of Overfitting and its Solutions. J. Phys. Conf. Ser. 1168, 022022 https://doi.org/10.1088/1742-6596/1168/2/022022.
Yu, M., Rui, X., Xie, W., Xu, X., Wei, W., 2022. Research on Automatic Identification Method of Terraces on the Loess Plateau Based on Deep Transfer Learning. Remote Sens. 14, 2446. https://doi.org/10.3390/rs14102446.
Yuan, D., Chang, X., Huang, P.-Y., Liu, Q., He, Z., 2021. Self-Supervised Deep Correlation Tracking. IEEE Trans. Image Process. 30, 976–985. https://doi.org/10.1109/TIP.2020.3037518.
Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716 https://doi.org/10.1016/j.rse.2020.111716.
Zenkl, R., Timofte, R., Kirchgessner, N., Roth, L., Hund, A., Van Gool, L., Walter, A., Aasen, H., 2022. Outdoor Plant Segmentation With Deep Learning for High-Throughput Field Phenotyping on a Diverse Wheat Dataset. Front. Plant Sci. 12.
Zhang, P., Bai, Y., Wang, D., Bai, B., Li, Y., 2021. Few-Shot Classification of Aerial Scene Images via Meta-Learning. Remote Sens. 13, 108. https://doi.org/10.3390/rs13010108.

Zhang, C., Li, G., Du, S., Tan, W., Gao, F., 2019a. Three-dimensional densely connected convolutional network for hyperspectral remote sensing image classification. J. Appl. Remote Sens. 13, 016519 https://doi.org/10.1117/1.JRS.13.016519.

Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q., Shen, C., 2019b. Hyperspectral Classification Based on Lightweight 3-D-CNN With Transfer Learning. IEEE Trans. Geosci. Remote Sens. 57, 5813–5828. https://doi.org/10.1109/TGRS.2019.2902568.

Zhang, W., Liljedahl, A.K., Kanevskiy, M., Epstein, H.E., Jones, B.M., Jorgenson, M.T., Kent, K., 2020. Transferability of the Deep Learning Mask R-CNN Model for Automated Mapping of Ice-Wedge Polygons in High-Resolution Satellite and UAV Images. Remote Sens. 12, 1085. https://doi.org/10.3390/rs12071085.

Zhang, Y., Wu, K., Du, B., Zhang, L., Hu, X., 2017. Hyperspectral Target Detection via Adaptive Joint Sparse Representation and Multi-Task Learning with Locality Information. Remote Sens. 9, 482. https://doi.org/10.3390/rs9050482.

Zhang, Y., Xin, D., 2022. A Diverse Ensemble Deep Learning Method for Short-Term Traffic Flow Prediction Based on Spatiotemporal Correlations. IEEE Trans. Intell. Transp. Syst. 23, 16715–16727. https://doi.org/10.1109/TITS.2021.3131248.

Zhang, Y., Yang, Q., 2018. An overview of multi-task learning. Natl. Sci. Rev. 5, 30–43. https://doi.org/10.1093/nsr/nwx105.

Zhang, C., Zhou, J., Wang, H., Tan, T., Cui, M., Huang, Z., Wang, P., Zhang, L., 2022. Multi-Species Individual Tree Segmentation and Identification Based on Improved Mask R-CNN and UAV Imagery in Mixed Forests. Remote Sens. 14, 874. https://doi.org/10.3390/rs14040874.

Zhao, W., 2017. Research on the deep learning of the small sample data based on transfer learning. AIP Conf. Proc. 1864, 020018 https://doi.org/10.1063/1.4992835.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L., Feng, M., 2019. Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang. China. Remote Sens. 11, 2673. https://doi.org/10.3390/rs11222673.

Zhao, P., Huang, L., Xin, Y., Guo, J., Pan, Z., 2021a. Multi-Aspect SAR Target Recognition Based on Prototypical Network with a Small Number of Training Samples. Sensors 21, 4333. https://doi.org/10.3390/s21134333.

Zhao, Y., Wang, X., Che, T., Bao, G., Li, S., 2023. Multi-task deep learning for medical image computing and analysis: A review. Comput. Biol. Med. 153, 106496 https://doi.org/10.1016/j.compbiomed.2022.106496.

Zhao, W., Yamada, W., Li, T., Digman, M., Runge, T., 2021b. Augmenting Crop Detection for Precision Agriculture with Deep Visual Transfer Learning—A Case Study of Bale Detection. Remote Sens. 13, 23. https://doi.org/10.3390/rs13010023.

Zhong, Z., Li, J., Clausi, D.A., Wong, A., 2020. Generative Adversarial Networks and Conditional Random Fields for Hyperspectral Image Classification. IEEE Trans. Cybern. 50, 3318–3329. https://doi.org/10.1109/TCYB.2019.2915094.

Zhou, Z.-H., 2018. A brief introduction to weakly supervised learning. Natl. Sci. Rev. 5, 44–53. https://doi.org/10.1093/nsr/nwx106.

Zhou, Y., Jiang, X., Li, Z., Liu, X., 2020. SAR Target Classification with Limited Data via Data Driven Active Learning, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 2475–2478. https://doi.org/10.1109/IGARSS39084.2020.9324364.

Zhu, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2018. Generative Adversarial Networks for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 56, 5046–5063. https://doi.org/10.1109/TGRS.2018.2805286.

Zhu, J., Li, Y., Hu, Y., Ma, K., Zhou, S.K., Zheng, Y., 2020. Rubik's Cube+: A self-supervised feature learning framework for 3D medical image analysis. Med. Image Anal. 64, 101746 https://doi.org/10.1016/j.media.2020.101746.

Zhu, X., Zheng, J., An, Y., Xin, X., Xu, D., Yan, R., Xu, L., Shen, B., Hou, L., 2023. Grassland Ecosystem Progress: A Review and Bibliometric Analysis Based on Research Publication over the Last Three Decades. Agronomy 13, 614. https://doi.org/10.3390/agronomy13030614.

D.M. Ziegler N. Stiennon J. Wu T.B. Brown A. Radford D. Amodei P. Christiano G. Irving Fine-Tuning Language Models from Human Preferences 2020 https://doi.org/10.48550/arXiv.1909.08593.

Zuo, X., Yu, X., Liu, B., Zhang, P., Tan, X., Wei, X., 2020. Graph inductive learning method for small sample classification of hyperspectral remote sensing images. Eur. J. Remote Sens. 53, 349–357. https://doi.org/10.1080/22797254.2021.1901064.

Zuo, X., Yu, X., Liu, B., Zhang, P., Tan, X., 2022. FSL-EGNN: Edge-Labeling Graph Neural Network for Hyperspectral Image Few-Shot Classification. IEEE Trans. Geosci. Remote Sens. 60, 1–18. https://doi.org/10.1109/TGRS.2022.3165025.