

Exploring Applications of Convolutional Neural Networks in Analyzing Multispectral Satellite Imagery: A Systematic Review

Antonia Ivanda*, Ljiljana Šerić, and Maja Braović

Abstract: Remote sensing is of great importance for analyzing and studying various phenomena occurrence and development on Earth. Today is possible to extract features specific to various fields of application with the application of modern machine learning techniques, such as Convolutional Neural Networks (CNN) on MultiSpectral Images (MSI). This systematic review examines the application of 1D-, 2D-, 3D-, and 4D-CNNs to MSI, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. This review addresses three Research Questions (RQ): RQ1: “In which application domains different CNN models have been successfully applied for processing MSI data?”, RQ2: “What are the commonly utilized MSI datasets for training CNN models in the context of processing multispectral satellite imagery?”, and RQ3: “How does the degree of CNN complexity impact the performance of classification, regression or segmentation tasks for multispectral satellite imagery?”. Publications are selected from three databases, Web of Science, IEEE Xplore, and Scopus. Based on the obtained results, the main conclusions are: (1) The majority of studies are applied in the field of agriculture and are using Sentinel-2 satellite data; (2) Publications implementing 1D-, 2D-, and 3D-CNNs mostly utilize classification. For 4D-CNN, there are limited number of studies, and all of them use segmentation; (3) This study shows that 2D-CNNs prevail in all application domains, but 3D-CNNs prove to be better for spatio-temporal pattern recognition, more specifically in agricultural and environmental monitoring applications. 1D-CNNs are less common compared to 2D-CNNs and 3D-CNNs, but they show good performance in spectral analysis tasks. 4D-CNNs are more complex and still underutilized, but they have potential for complex data analysis. More details about metrics according to each CNN are provided in the text and supplementary files, offering a comprehensive overview of the evaluation metrics for each type of machine learning technique applied.

Key words: remote sensing; MultiSpectral Images (MSI); satellite images; Convolutional Neural Networks (CNN); machine learning; classification; regression; segmentation

* Antonia Ivanda, Ljiljana Šerić, and Maja Braović are with Department of Electronics and Computer Science, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split 21000, Croatia. E-mail: asenta00@fesb.hr; ljiljana@fesb.hr; mbraovic@fesb.hr.

* To whom correspondence should be addressed.

Manuscript received: 2024-08-13; revised: 2024-10-12; accepted: 2024-10-28

1 Introduction

Remote sensing is a valuable source of information for better understanding the Earth's system, as it enables the continuous and systematic collection of data about the Earth's surface over time. These data are typically acquired through sensors installed on aircrafts, Unmanned Aerial Vehicles (UAVs), or satellites,

which receive and measure electromagnetic energy emitted or reflected from the observed surface^[1]. Multispectral and hyperspectral imagery are common types of remote sensing data that provide information about the Earth's surface in multiple wavelengths of light in the range from visible to infrared regions of the electromagnetic spectrum. MultiSpectral Image (MSI) consists of multiple bands, where each band captures information in a specific wavelength range, usually ranging from three to ten bands. MSI also has a relatively low cost compared to hyperspectral imagery, making it a more accessible option for many researchers and organizations^[2]. For example, NASA's Landsat program has been collecting multispectral images of the Earth's surface since the 1970s. The program provides free data downloads through the USGS EarthExplorer website (<https://earthexplorer.usgs.gov/>). On the other hand, the European Space Agency's (ESA) Sentinel program has been providing multispectral and radar images since 2014, and also provides free downloads through the Copernicus Open Access Hub. Multispectral imagery has found widespread application across various fields, owing to its free availability, and the valuable spatial and spectral information that it provides. These applications range from environmental monitoring^[3], forestry^[4], geology^[5], agriculture^[6, 7], urban planning^[8], and more.

Pixel classification is an important task in remote sensing, where in the context of multispectral imagery allows for the identification and mapping of land cover features with high spatial resolution. Conventional machine learning techniques are restricted in their capability to handle raw data directly. Developing a machine learning system requires significant domain knowledge to design a feature extractor that would convert raw data, such as image pixel values, into an appropriate feature vector. The developed algorithm, often a classifier, would then use this feature vector to identify or classify patterns in the input data^[9]. Machine learning algorithms are widely used in multispectral imagery for tasks, such as regression, classification, and feature extraction. Prominent examples of these algorithms include the random forest^[10, 11], support vector machine^[12], k-nearest neighbors^[13], and ridge regression^[14, 15] approaches that have been successfully applied to perform pixel-based analysis and classification tasks. The disadvantage of these algorithms is using only spectral

information from image, which affects the capabilities of proposed algorithms. However, many machine learning algorithms used in multispectral imagery only consider spectral information and neglect important spatial information. This limitation can cause decreased accuracy and even lead to misclassification in certain cases. To address this, deep learning has shown promise in incorporating not only spectral information but also spatial and temporal information into its models. By considering spatial information, such as the relationships between neighboring pixels within an image, and temporal information, such as changes in the object or scene being imaged over time, deep learning models can identify patterns that take into account changes in the spectral and spatial characteristics of the input data over time^[16]. As a result, deep learning has become the fastest-growing trend in big data analysis, enabling the exploitation of feature representations learned exclusively from data^[17, 18].

Over the past few years, there has been a noticeable increase in interest in the classification of both hyperspectral and multispectral images by using deep learning methods. While many articles have been published on this topic, only a limited number of review articles offer a comprehensive overview of the current state of research in this area. Specifically, the majority of review papers focus on hyperspectral image classification and examine various approaches, algorithms, techniques, and applications in this field^[19–22]. As an illustration, in Ref. [23], the authors conducted a comprehensive comparison of several Convolutional Neural Network (CNN) architectures, including 1D-CNN, 2D-CNN, 3D-CNN, and Feature fusion based CNN (FCNN), for the purpose of hyperspectral image classification. In Ref. [24], the authors provided a brief overview of several deep learning models that can be utilized for hyperspectral image classification, such as CNN, Stacked AutoEncoders (SAE), and Deep Belief Neural Networks (DBN). They systematically analyzed the state-of-the-art deep learning approaches from two perspectives: pixel-wise image classification and scene-wise image classification. However, to the best of our knowledge, there is currently no systematic review article that focuses solely on the application of deep learning techniques to MSI, even though there are several review articles that address the broader topic of deep learning and remote sensing images, which

encompasses hyperspectral, multispectral, Unmanned Aerial Vehicle (UAV), and Synthetic Aperture Radar (SAR) imagery^[16, 25, 26]. In addition, Refs. [17, 27] provides a comprehensive review and resources for high-resolution multispectral imagery in the context of scene classification, object detection, segmentation, and image retrieval. Although there are multiple review articles that cover the topic of MSI classification, a comprehensive analysis of the CNN-based approach has not yet been provided. This implies that more in-depth exploration and analysis of CNN-based MSI classification is necessary. To address this inquiry, this paper focuses specifically on different types of CNNs based on the dimensionality of input data. Specifically, we examine 1D-, 2D-, 3D-, and 4D-CNNs, which respectively cover spectral, spatial, spectral-spatial, and spectral-spatial through-time classification. To identify the type of CNN (1D, 2D, 3D, or 4D) we introduce the term “degree of convolution” to specify the dimensionality of the input data in a CNN. In CNNs, the term “dimension” can refer to the number of dimensions in the input data or feature map, as well as the number of filters or channels in a layer that do not necessarily match the input data or feature map’s dimension. To avoid ambiguity, this paper uses the term “degree” to identify the type of CNN (1D, 2D, 3D, or 4D) based on the dimensionality of the input data. By doing so, the systematic review reported in this paper aims to provide a more detailed analysis of CNN-based MSI classification in various applications.

This review follows the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines, which provide a framework for conducting and reporting systematic reviews and meta-analyses^[28]. The completed PRISMA 2020 checklist and PRISMA 2020 abstract checklist are available in Tables S1 and S2, respectively, which are in the Electronic Supplementary Material (ESM) of the online version of this article.

The objectives of this systematic review are to answer the following Research Questions (RQ):

- RQ1: In which application domains different CNN models have been successfully applied for processing MSI data?
- RQ2: What are the commonly utilized MSI datasets for training CNN models in the context of processing multispectral satellite imagery?
- RQ3: How does the degree of CNN impact the performance of classification, regression or

segmentation tasks for multispectral satellite imagery?

This paper is organized as follows. In Section 2, we provide an overview of 1D, 2D, 3D, and 4D convolutional neural architectures. Section 3 outlines the methods used for the systematic literature review, which is organized by the specific PRISMA guideline. In Section 4, we present the findings of our review and discuss the implications of these results. Finally, in Section 5, we present our conclusions and future recommendations regarding the use of different convolutional neural network architectures for processing multispectral images.

2 CNN Architecture

CNNs are a type of neural network designed for processing data that has a grid-like topology. The name “convolutional neural network” indicates that the network leverages a mathematical operation known as convolution, which is a type of linear operation. In essence, convolutional networks are neural networks that use convolution instead of general matrix multiplication in at least one of their layers^[29]. Figure 1 illustrates the basic architecture of a CNN. This architecture consists of five primary components: an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer^[30]. The convolutional and pooling layers are alternately stacked to extract deep features from the input data before passing them onto the fully connected layer, which is responsible to classify the data into various classes. A convolutional layer is a key component of the CNN architecture that consists of a collection of convolutional filters (kernels) that are convolved with the input data to generate the output feature map, where the kernel weights are adjusted through training to extract significant features. A pooling layer shrinks large-size feature maps to create smaller ones while maintaining the dominant information by subsampling, using methods, such as max, min, or global

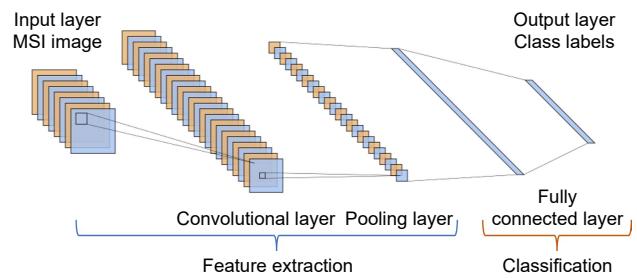


Fig. 1 Basic architecture of a convolutional neural network.

average pooling. However, this process can lead to a loss of relevant information. A Fully Connected (FC) layer in a CNN is the last layer that connects all neurons to the neurons of the previous layer, and it is used as the CNN classifier to produce the final output. The input to the fully connected layer is a flattened vector from the feature maps generated by the last pooling or convolutional layer^[31].

2.1 1D-CNN model

Multispectral and hyperspectral imagery have a high spectral resolution, where spectral information of each pixel is important for identifying small objects and capturing the unique reflectance properties of different materials and targets. 1D-CNN models can analyze spectral data, which are represented as a one-dimensional array of spectral values (see Fig. 2). While spectral information can improve classification performance, it may not always capture the full structure of an object without considering the spatial arrangement of reflectance properties across different wavelengths^[32]. Therefore, 1D-CNN models that rely solely on spectral information may struggle to distinguish between objects with similar spectral signatures but different spatial structures, leading to misclassifications or poor performance. In addition, 1D-CNN models can also be applied to temporal or spatial data at the pixel level^[33], which are less commonly utilized in remote sensing research compared to spectral data.

2.2 2D-CNN model

In the context of analyzing multispectral remote sensing imagery, the 2D-CNN is a powerful approach for feature extraction and classification. Unlike traditional machine learning algorithms that use pixel-based inputs, 2D-CNN models leverage patch-type inputs to identify and extract optimal features for classification. Incorporating spatial information from neighboring pixels through the use of patches has been shown to be effective in classifying different types that

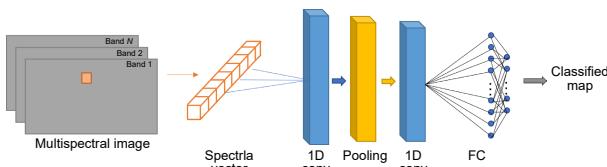


Fig. 2 1D-CNN model architecture for spectral feature extraction in MSI data.

share similar spatial characteristics, such as crops^[34]. In general, the 2D-CNN structure includes two typical layers: a convolutional layer and a pooling layer. The convolutional layer extracts spatial features by applying a convolution operator to each pixel and generating various feature maps. The pooling layer reduces the dimensions of feature maps generated by convolutional filters or abstractly emphasizes certain features (see Fig. 3)^[9, 35].

2.3 3D-CNN model

The use of a 2D-CNN can enable the extraction of spatial information from one spectral band at a time, but fails to exploit the additional information present in other bands. On the other hand, 3D-CNNs can work with an entire 3D data cube formed by stacking multispectral images, with each channel representing a different spectral band (see Fig. 4). The effectiveness of model training can be impacted by the size of the patches used as input, where too much data can introduce noise and too little data may limit the receptive field^[36]. This approach allows the extraction of spatial features over multiple bands simultaneously, which can provide better classification accuracy compared to 2D-CNNs^[37]. In addition, 3D-CNNs can also take advantage of the temporal dimension, allowing the classification of images taken at different times^[38].

2.4 4D-CNN model

4D-CNNs offer the advantage of jointly exploiting the spatial, spectral, and temporal information present in multispectral imagery by using 4D convolutional filters (see Fig. 5). These filters operate across all four

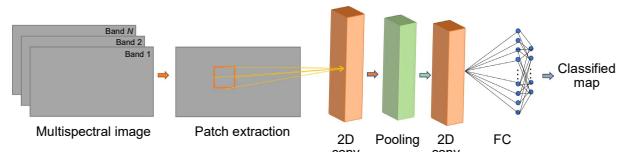


Fig. 3 2D-CNN model architecture for spatial feature extraction in MSI data.

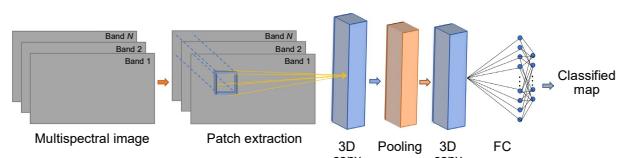


Fig. 4 3D-CNN model architecture for spatial-spectral feature extraction in MSI data.

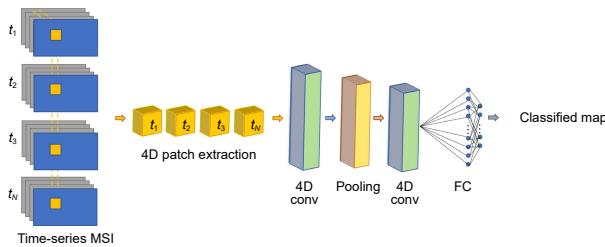


Fig. 5 4D-CNN model architecture for spatial-spectral feature extraction in time-series MSI data.

dimensions, namely the spatial dimensions (height and width), and the spectral and temporal dimensions (band and time)^[39]. This allows 4D-CNNs to capture complex relationships between spatial, spectral, and temporal features that are important for accurate image classification and analysis.

Unfortunately, 4D convolutional layers are not natively available in popular deep learning frameworks like Tensorflow^[40] or PyTorch^[41]. As a result, 4D-CNNs must be implemented using a sequence of 3D convolutions along the temporal dimension. This involves stacking multiple 3D convolutional layers that operate on the multispectral images at different points in time, allowing the network to learn the temporal dynamics of the scene^[42]. While this approach is not as computationally intensive as using native 4D convolutional layers, it still enables the sharing of spatial, spectral, and temporal information in multispectral images.

3 Material and Method

This section, in accordance with the rationale and objectives stated in the introduction part (Section 1), provides a comprehensive description of the study's approach, including eligibility criteria, information sources, search strategies, selection process, data extraction, risk of bias assessment, effect measures, and data synthesis.

3.1 Eligibility criteria

The main eligibility criterion for inclusion in this study is scientific publications that apply 1D-, 2D-, 3D-, or 4D-CNNs in the field of remote sensing. The broader term “remote sensing” is employed to capture studies that may not explicitly mention multispectral imagery but rather refer to satellite names or datasets relevant to the field. Studies not focus on multispectral satellite imagery (e.g., hyperspectral or UAV imagery) are excluded. Additionally, publications that focus

exclusively on already well-known CNN models, such as AlexNet, ResNet, and VGG, are not considered eligible, but studies that originally use different degrees of CNNs or adapt them in their models are included. Only articles and conference papers published in English until March 2024 are considered. The detailed eligibility criteria for this review are summarized in Table 1.

3.2 Information sources

On May 10th, 2024, we searched the Web of Science, IEEE Xplore, and Scopus databases to identify relevant studies. These databases are chosen for their comprehensive coverage of scientific literature in various fields. The search encompasses publications available until March 31st, 2024.

3.3 Search strategy

The search strategy utilizes a combination of keywords related to convolutional neural networks and remote sensing. Specifically, the query included variations of “Nd-cnn” OR “Nd cnn” OR “Nd convolutional” AND “remote sensing”, where “N” represents the degree of the CNN ranging from one to four. All publications are limited to those published in the English language and with source types “journal” or “conference proceeding”. The same query is applied across all three databases: Web of Science, IEEE Xplore, and Scopus. An exception is made for the Scopus database, where the search is limited to publication titles, keywords, and abstracts due to the very large number of

Table 1 Eligibility criteria.

Type of data	MSI
Algorithm or technique	CNNs (1D, 2D, 3D, and 4D) for image classification, object detection, semantic segmentation, or regression
Comparator	RQ1: e.g., Vegetation, Urban; RQ2: e.g., Sentinel, Landsat; RQ3: Evaluation metrics for classification, regression, and segmentation
Outcome	Classification and characterization of the diverse applications of CNNs within the domain of remote sensing
Timing	All
Environmental or geographical context	All
Study design	Original data, relevant articles, and conference papers
Publication	Studies published in English only

publications indexed in this database. Additionally, to manage the large volume of search results in the Scopus database, filters are applied to keywords to exclude irrelevant content (e.g., electrocardiogram and hyperspectral images). Given the limitations in specifying search dates in IEEE Xplore, publications are manually reviewed up to the search date to exclude those published after March 31st, 2024, ensuring the relevance and timeliness of the search results. Table 2 presents the basic queries for all three databases, where Nd is related to 1D-, 2D-, 3D-, and 4D-CNNs. The detailed systematic database search strategy for each query line by line is provided in Tables S3–S5 in the ESM.

3.4 Selection process

Two researchers of this paper (Antonia Ivanda and

Table 2 Search strategy for different databases and CNNs.

Database	Search query
Web of Science	(ALL=(Nd-cnn)) OR ALL=(Nd cnn) OR ALL=(Nd convolutional))) AND ALL=(remote sensing) AND (LA==("ENGLISH")) AND DT==("ARTICLE" OR "PROCEEDINGS PAPER")) AND DOP=1990-01-01/2024-03-31
IEEE Xplore	((("All Metadata": "Nd-cnn") OR "All Metadata": "Nd cnn" OR "All Metadata": "Nd convolutional") AND ("All Metadata": "remote sensing")))) AND ("ContentType": "Journals" OR "ContentType": "Conferences")
Scopus	((TITLE-ABS-KEY (1d-cnn) OR TITLE-ABS-KEY (1d AND cnn) OR TITLE-ABS-KEY (1d AND convolutional)) AND ALL (remote AND sensing)) AND ((PUBYEAR > 1992 AND PUBYEAR < 2024) OR PUBDATETXT (January 2024) OR PUBDATETXT (February 2024) OR PUBDATETXT (March 2024)) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j") OR LIMIT-TO (SRCTYPE, "p"))

Ljiljana Šerić) independently reviewed the titles and abstracts of all publications retrieved from the databases. They then compared their results and identified publications that were not initially included in their individual reviews. These inconsistencies were discussed to reach a consensus on whether these publications should be included or excluded. In cases where consensus could not be reached, a third researcher (Maja Braović) was consulted to make the final decision. Then, the agreed-upon list of publications underwent full-text screening to determine their eligibility based on the inclusion criteria.

3.5 Data extraction

The extracted data from the included studies follow standardized forms: publication type (journal or conference), authors, article title, source title, publication year, author keywords, abstract, research areas, open access, satellite, domain, application, machine learning technique, algorithm complexity, accuracy, F1-score, precision, recall, producer accuracy, user accuracy, pixel accuracy, Kappa, Intersection over Union (IoU), Mean IoU (MIoU), Dice, *R*-squared (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and parameter description. One reviewer performs data extraction, after which a second reviewer assesses to ensure accuracy and completeness of it. If there are disagreements about data extraction, they are resolved by consensus discussion.

3.6 Risk of bias assessment

Due to limitations in data availability in this research, we encounter challenges in extracting all relevant evaluation metrics for each CNN. These obstacles may also have impacted the evaluation of risk of bias. To reduce bias, we document our data extraction procedure and systematically review studies. Additionally, reviewers collaborate to ensure accuracy and consistency in information extraction. Rather than making use of a "Risk of Bias" tool, we trust comprehensive documentation and reviewer consensus to assess bias inside the scope of this systematic review.

3.7 Effect measures

The performance of CNNs includes evaluation metrics across different tasks, such as classification, segmentation, and regression. For classification and

segmentation tasks, we collect data on the following metrics: accuracy, F1-score, precision, recall, producer accuracy, user accuracy, pixel accuracy, Kappa, IoU, MIoU, and Dice coefficient. These metrics are expressed as percentages (%). For regression tasks, we extract data from the collected publications for metrics R^2 , RMSE and MAE, all of which are expressed according to the predicted unit (e.g., chlorophyll-a concentration as mg/m³). It is important to note that in some publications, it is not possible to access every measure for the performed tasks. Nevertheless, we have tried to collect and report available data to provide insights into the performance of CNNs in various remote sensing applications.

3.8 Data synthesis

In order to synthesize the findings from the included studies, data are organized into tables based on the tasks performed by convolutional neural networks. Specifically, publications are categorized according to the tasks of classification, segmentation, and regression, with separate tables created for each CNN degree (1D, 2D, 3D, and 4D). Tables detailing the extracted metrics for each CNN degree are provided in Tables S6–S9 in the ESM.

4 Result and Discussion

4.1 Study selection

A systematic review of the literature from the Web of Science, IEEE Xplore, and Scopus databases retrieves 678 articles for 1D-CNN related studies, 1213 articles for 2D-CNN related studies, 2013 articles for 3D-CNN related studies, and 48 articles for 4D-CNN related studies. The detailed query for each CNN-related study and for each database can be found in the ESM (Tables S3–S5). Studies included in this review were published until March 31st, 2024. Due to the limitation of the IEEE Xplore database, which does not allow filtering by date or month but only by year, we had to manually review the publications and remove those published after March 31st, 2024. After that, we screened the title of each publication, resulting in 405 studies related to 1D-CNNs, 765 studies related to 2D-CNNs, 1219 studies related to 3D-CNNs, and 37 studies related to 4D-CNNs. In the next stage of abstract screening and full-text assessment (if applicable), we excluded studies that were not relevant to this systematic review. Thus, in the final review process, 49 studies were

included for 1D-CNNs, 58 for 2D-CNNs, 66 for 3D-CNNs, and 3 for 4D-CNNs (see Fig. 6). More information about all included studies in this systematic review can be found in Tables S6–S9 in the ESM.

4.2 Overview of remote sensing publications

The systematic review includes a total of 143 unique publications, comprising approximately three quarters (110) of journal articles and one quarter (33) of conference articles (see Fig. 7).

In examining the distribution of articles using convolutional neural networks, it is evident that publications using 3D-CNNs dominate (see Fig. 8).

It is interesting that 3D-CNN allows the extraction of spatial features over multiple bands simultaneously, potentially providing better classification accuracy compared to 1D-CNNs and 2D-CNNs^[37]. In contrast, 4D-CNNs show the lowest representation with less than 2% of the total publications included in our systematic review. 4D-CNNs are underexplored in terms of their application across diverse domains. While we find applications in “Vegetation” and “Water” domains, there is a lack of studies in “Agriculture”, “Urban”, and “Geohazards” domains. Additionally, the temporal aspect of 4D-CNNs is not fully utilized in most studies. The reason for this could be the complexity of 4D-CNNs and their unavailability in some popular deep learning frameworks, such as Tensorflow or PyTorch. As a result, 4D-CNNs must be implemented using a sequence of 3D convolutions along the temporal dimension, involving stacking multiple 3D convolutional layers that operate on the multispectral images at different points in time. This allows the network to learn the temporal dynamics of the scene^[42]. While this approach may be computationally complex and not be as efficient as using native 4D convolutional layers, it still allows for the joint exploitation of spatial, spectral, and temporal information in multispectral imagery.

Figure 9 presents the heatmap illustrating the distribution of the most common journals used over the years. It is interesting to note that the majority of publications are published in journals associated with the term “remote sensing” in their names. The most prevalent journal sources for the publications are *Remote Sensing* and *ISPRS Journal of Photogrammetry and Remote Sensing*.

Similarly, Fig. 10 presents a heatmap for the most

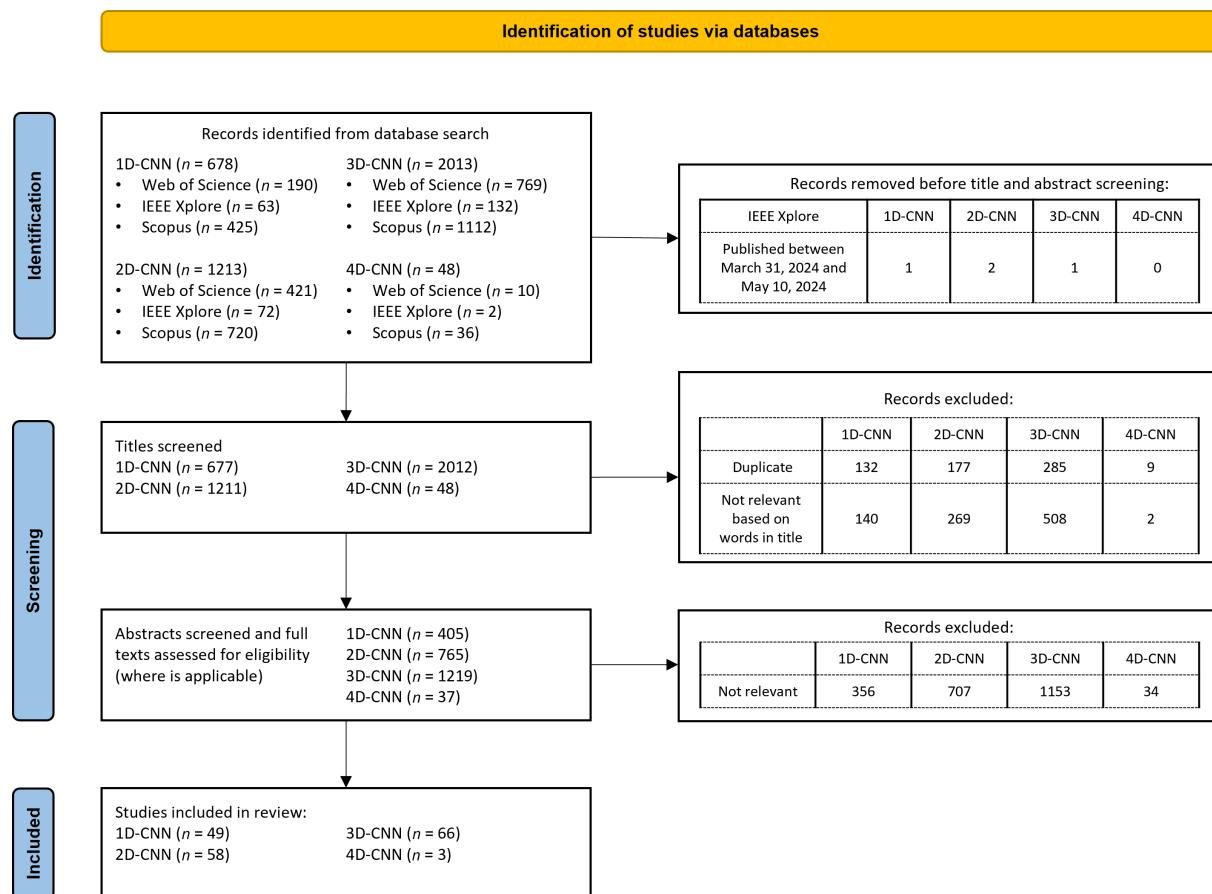


Fig. 6 Flow diagram illustrating the publication identification and screening process following the PRISMA guidelines (template is re-used from Page et al.^[28] with CC BY 4.0)

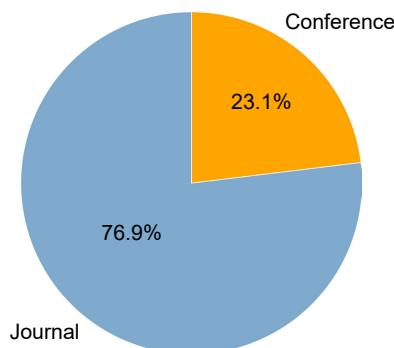


Fig. 7 Proportion of publications by type.

selected conferences over years for publishing remote sensing related articles. As for the conference sources, the International Geoscience And Remote Sensing Symposium (IGARSS) has the highest number of published articles. The label “Others” is assigned to journals or conferences that have been identified as sources with only one publication. Over the past decade, CNNs have become increasingly popular in remote sensing applications, particularly in processing

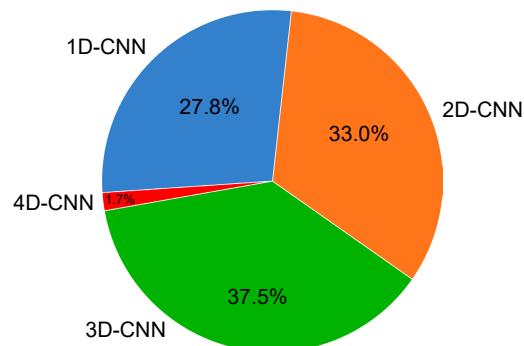


Fig. 8 Proportion of publications by CNN.

MSI data which suggest these numbers of published articles. A notable increase in research output can be observed from 2020 to 2023, with 5 or more conference papers published each year and at least 17 journal articles released annually, indicating a consistent trend in research output. This is further illustrated by the overall number of publications for each CNN degree over the years in Fig. 11. Moreover, while the number of publications peaked in 2022, there is a drastic decrease in publications for 3D-CNNs

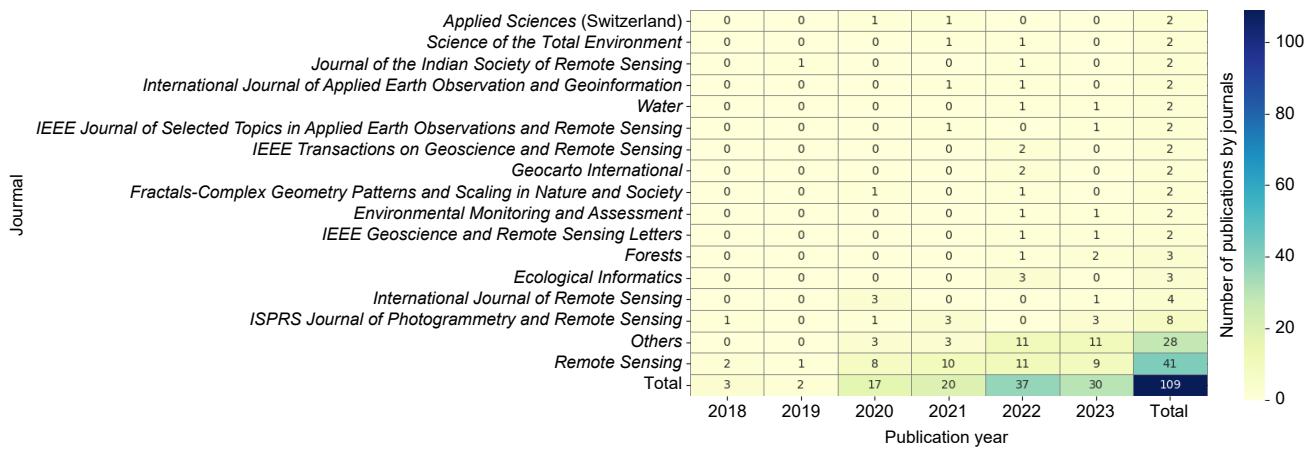


Fig. 9 Heatmap of the most frequent journals of publications focusing on CNN applications in multispectral imagery.

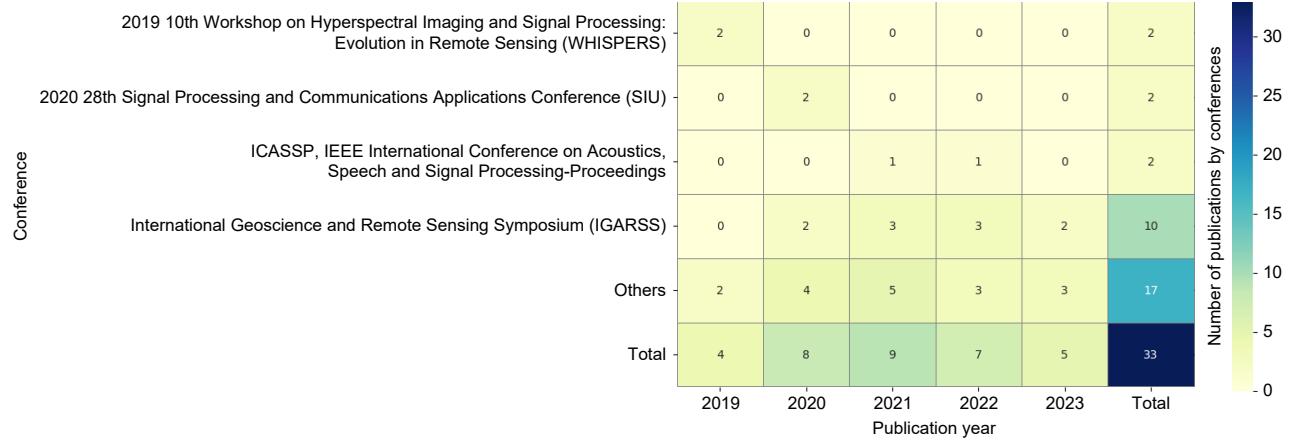


Fig. 10 Heatmap of the most frequent conferences of publications focusing on CNN applications in multispectral imagery.

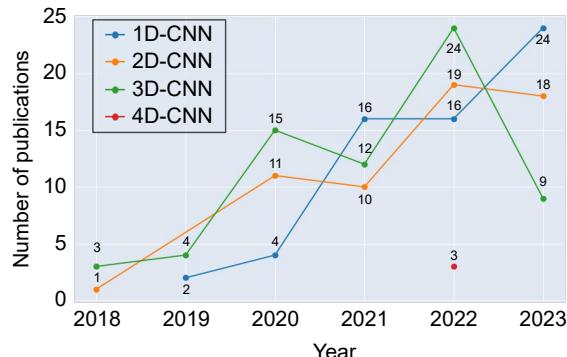


Fig. 11 Publication yearly trends over time for each 1D-, 2D-, 3D-, and 4D-CNN.

thereafter (24→9), but an increase is observed for 1D-CNNs (16→24).

Figure 12 provides a visual representation of the highest-frequency terms found in the titles and abstracts of peer-reviewed literature with larger font sizes indicating higher frequency. It can be noticed that the terms with the highest frequency are “data” and

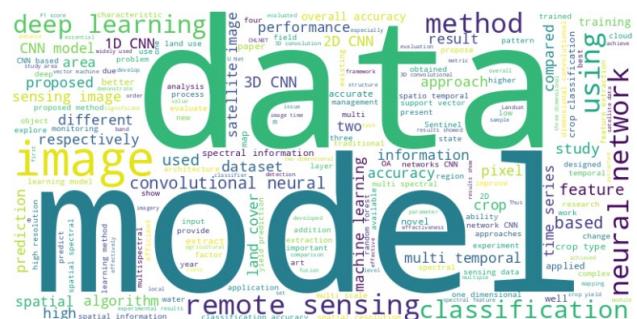


Fig. 12 Word cloud of the most frequently used words in abstracts.

“model”, which suggest the importance of selecting quality data and models in remote sensing research. Furthermore, terms that have a slightly lower frequency are “method”, “classification”, “deep learning”, and “neural network”, which suggest common approaches used in publications. This observation underlines the importance of methodological choices in analyzing and interpreting remote sensing data.

Figure 13 shows a publication distribution for each CNN type classified in one of six different domains: “Agriculture” (e.g., crop classification), “Geohazards” (e.g., wildfire prediction), “Urban” (e.g., urban land cover maps), “Vegetation” (e.g., forest classification), “Water” (e.g., estuary water quality classification), and “Others” (e.g., spatiotemporal image fusion). Most publications use first three degree of CNNs to solve problems in the domain of “Agriculture”, specifically for crop classification and crop yield prediction. The second most used domain is “Vegetation”, where 4D-CNNs are applied. “Vegetation” usually includes studies involving the analysis of plants, such as trees, shrubs, grasslands, and natural vegetation^[43]. “Agriculture”, on the other hand, includes studies focused on crop breeding, agricultural land use monitoring, and crop yield forecasting, often relying on vegetation data to assess crop health, productivity, and land use^[44]. Thus, both domains involve monitoring and analysis of plants life.

The distribution of publications of each CNN categorized by satellite data sources is presented in Fig. 14. The plot shows that the majority of publications utilized data from the Sentinel-2 satellite when implementing 1D-, 2D-, and 3D-CNNs. Sentinel-2 satellites, operated by European Space Agency (ESA), have 13 bands with spatial resolutions ranging from 10 m to 60 m and a temporal resolution of 5 days^[45]. Notably, satellites Landsat-8 and MODerate-resolution Imaging Spectroradiometer (MODIS) have been preferred across all degrees of CNNs, even though there are almost half as many publications compared to Sentinel-2. Landsat-8 carries a two-sensor

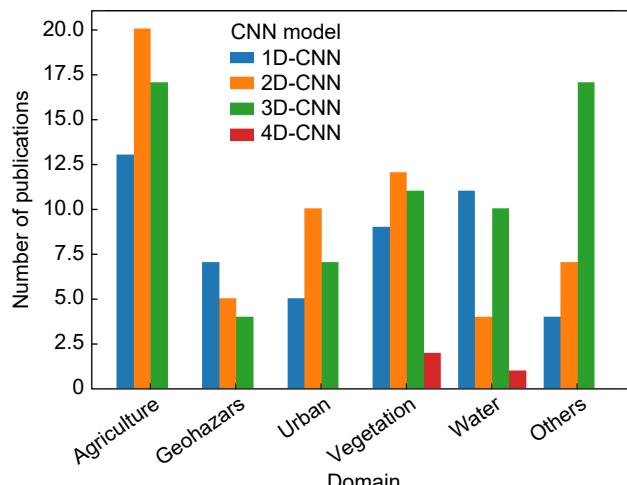


Fig. 13 Publication distribution for each CNN by domain.

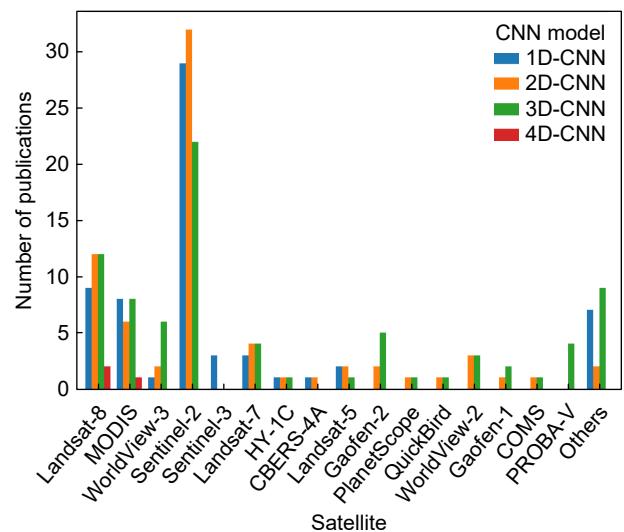


Fig. 14 Publication distribution for each CNN by satellite.

payload, the Operational Land Imager (OLI), and the Thermal InfraRed Sensor (TIRS), where OLI captures nine spectral bands including one panchromatic band while TIRS has two thermal bands. It has a temporal resolution of 16 days, with a 15-meter spatial resolution for the panchromatic band and a 30-meter spatial resolution for the multispectral bands^[46]. On the other hand, MODIS is an NASA instrument launched on two satellites—MODIS Terra and MODIS Aqua. It has 36 spectral bands and a 1–2-day temporal resolution. It collects data at three spatial resolutions: 250 m, 500 m, and 1000 m^[47].

4.3 Meta-analysis of publications

As can be seen in Fig. 14, the highest representation is of the Sentinel-2, Landsat-8, and MODIS satellites across different levels of CNNs. If we look at the individual usage of satellites for each CNN degree in particular application domains, we can observe a similar situation. Specifically, for 1D-CNNs, in Fig. 15, it can be seen that Sentinel-2 is the most represented in all domains, with the highest application in the “Agriculture” domain. Landsat-8 is most represented in the “Water” domain, followed by “Vegetation”, “Others”, and “Urban” domains. As for MODIS, it is equally used in the “Agriculture”, “Geohazards”, and “Water” domains.

Regarding the use of 2D-CNNs, if we look at the individual application domains in Fig. 16, we can see that in the “Agriculture” domain, the most represented satellite is Sentinel-2, followed by Landsat-8 and MODIS. In the “Geohazards” domain, for example, it

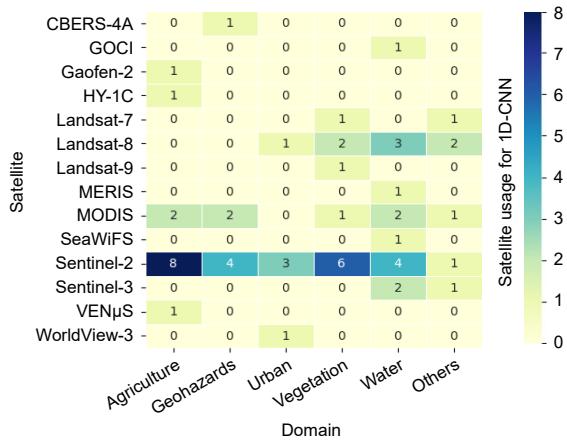


Fig. 15 Satellite usage across different domains for 1D-CNN.

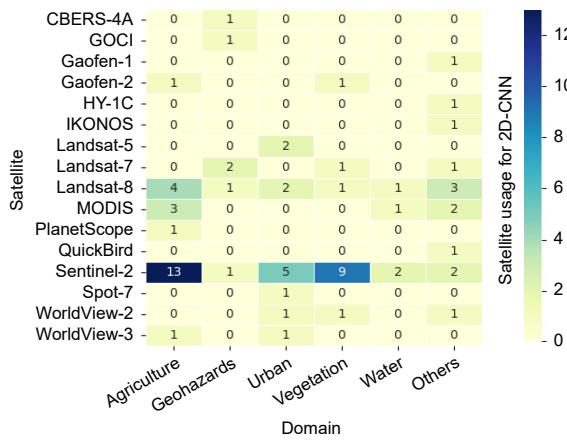


Fig. 16 Satellite usage across different domains for 2D-CNN.

is Landsat-7, while in the “Urban” domain, besides the three most common satellites (Sentinel-2, Landsat-8, and MODIS), Landsat-5 is also represented.

Figure 17 shows that more different satellites are used in the study for 3D-CNNs compared to 1D- and 2D-CNNs. For example, in the “Agricultural” domain, in addition to the three most commonly used satellites, the Gaofen-2 satellite is also used. The “Others” domain is also interesting, where many publications have used satellites, such as Gaofen-1, Landsat-8, PROBA-V, and WorldView-3. For instance, studies that use Gaofen-1 data are applied in cloud masking^[48, 49]. As for Landsat-8 data, they are used for image preprocessing and enhancement purposes^[50–53]. PROBA-V data are mainly utilized for improving satellite image quality^[54–57], while data from the WorldView-3 satellite are used for image analysis and processing^[58–60] and image compression^[51].

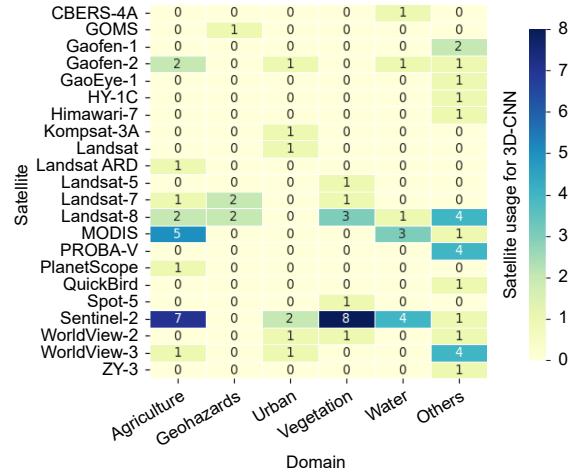


Fig. 17 Satellite usage across different domains for 3D-CNN.

The number of publications that use 4D-CNNs is too small, so if we look at the satellite usage for 4D-CNNs in Fig. 18, we can see that there is not enough data for general conclusions. There are a total of three publications where 4D-CNNs are applied, of which two use Landsat-8 data for problem-solving in the “Vegetation” domain and one uses MODIS data for problem-solving in the “Water” domain. Both publications related to using 4D-CNNs on Landsat-8 data in the “Vegetation” domain are focused on land cover classification^[39, 42], while the publication related to MODIS data in the “Water” domain uses 4D-CNNs for cyanobacteria bloom prediction^[61].

As shown in Fig. 15, Sentinel-2 is the most frequently used satellite data source across different CNN architectures. Further analysis reveals that a significant portion of these Sentinel-2 based studies

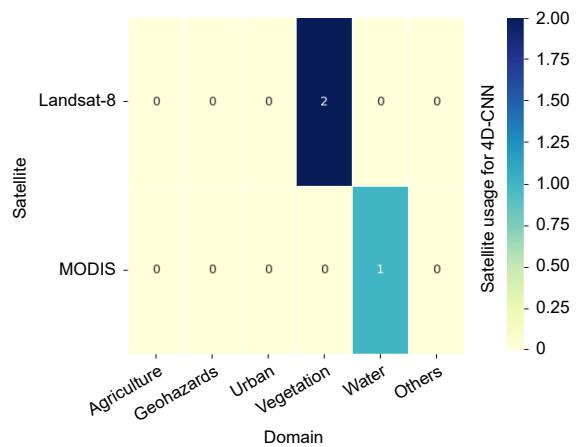


Fig. 18 Satellite usage across different domains for 4D-CNN.

focus on agricultural applications. For instance, in the 1D-CNNs category (see Fig. 16), Sentinel-2 data are predominantly used in the Agriculture domain. This trend is consistent across 2D-CNNs (see Fig. 17) and 3D-CNNs (see Fig. 18) applications as well, highlighting the particular suitability of Sentinel-2 data for agricultural studies using various CNN architectures.

CNN architectures can often be complex, differ in the number of layers or activation functions, but the final classification is achieved from the output layer. The output layer represents the final layer of the CNN architecture which usually utilizes loss function. This loss function aims to calculate an error using two parameters: the predicted value calculated by a CNN and the actual value. Different types of loss functions are used mainly in machine learning techniques, such as classification (e.g., Softmax loss function), regression (e.g., Euclidean loss function), and segmentation (e.g., cross-entropy loss)^[31]. In this systematic review, from each publication, we extract the type of the three machine learning techniques used and metrics related to each technique where it is possible. The distribution of publications using machine learning techniques over different CNNs is shown in Fig. 19. We also include compression and image enhancement, which are not traditionally categorized as machine learning techniques. These are techniques related to image processing that often use CNNs to achieve tasks and in our systematic review these techniques are used in publications related to 3D-CNNs. Our findings indicate that classification is the most used machine learning technique, with more than 30 publications for 1D-, 2D-, and 3D-CNNs. Regression is the second most used machine learning technique, and segmentation is the least used. Because

most publications use classification, we perform further analysis of metrics for publications using classification. There are not enough publications related to regression to make a constructive analysis, and in addition, the unavailability of data for metrics makes the sample very small for any concrete conclusions. Therefore, we provide full dataset for regression and segmentation metrics where we could extract them from publications and analyze it in the sections below.

4.3.1 Analysis of publications implementing classification

Classification involves examining the connections among a group of “objects” to determine whether the data can be accurately summarized by a limited number of categories representing similar objects^[62]. From the perspective of remote sensing imagery, it is a process of categorizing pixels in an image of raw satellite data to obtain a given set of labels. There are different types of classification, of which the most popular are supervised and unsupervised^[63]. In this systematic review, we do not analyze the types of classifications used in publications, only performance metrics accuracy and F1-score.

Accuracy is a measure that refers to the total data accurately predicted by the trained classifier when tested on unseen data. It ranges from 0 to 1, or in percentage terms from 0% to 100%, where accuracy closest to 1 or 100% suggests that the classifier is more accurate^[64]. Figure 20 shows a box and whisker plot pointing the central tendency, dispersion, and potential outliers of accuracy for different CNN degrees over different domains. If we look at the accuracy of 1D-CNNs for each domain, it can be seen that 1D-CNNs tend to have an accuracy of around 90% in the examined publications. The median accuracy values for 1D-CNNs in each domain are as follows: “Agriculture” – 92%, “Geohazards” – 90.59%, “Urban” – 91.3%, “Vegetation” – 93.78%, “Water” – 91.38%, and “Others” – 90.59%. For the “Urban” domain, the plot shows slightly negative skewness, suggesting that there are some publications with less than 90% accuracy, which is further confirmed by one publication falling outside the whisker. For the “Agriculture” domain, the plot shows slightly positive skewness, indicating that there are several publications with higher accuracy compared to other publications. Examining the accuracy of 2D-CNNs for each domain, it can be noticed that the boxes for 2D-CNNs across domains

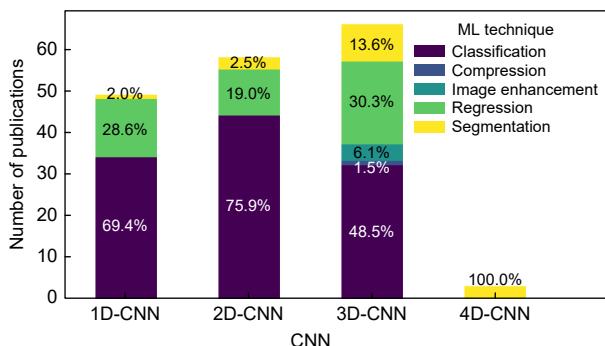


Fig. 19 Number of publications using different machine learning techniques by each CNN.

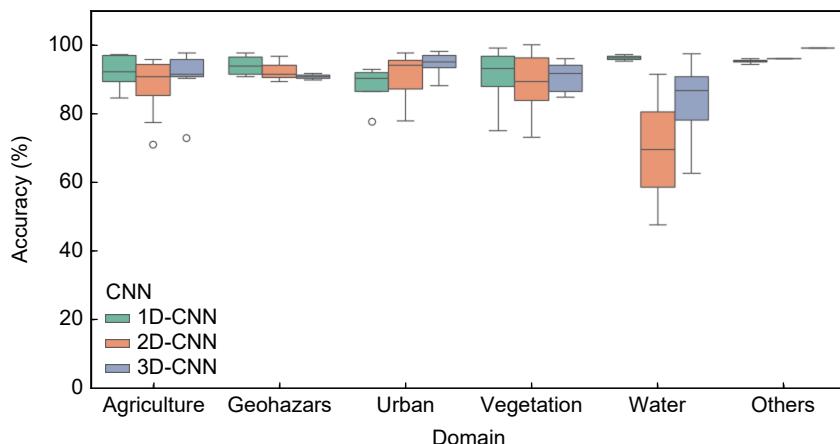


Fig. 20 Accuracy of CNNs using classification across different domains.

are taller compared to those for the other two CNNs. This suggests that there is greater variability in the accuracies reported in publications for 2D-CNNs compared to the other CNNs. The median accuracy values for 2D-CNNs in each domain are as follows: “Agriculture” – 90.28%, “Geohazards” – 94.09%, “Urban” – 95%, “Vegetation” – 93%, “Water” – 89.18%, and “Others” – 91.57%. For the “Agriculture” and “Urban” domains, the plot shows negative skewness, and in the “Agriculture” domain, one dot is outside the whisker, indicating that one publication falls outside the range. Positive skewness is observed in the “Geohazards” domain, indicating that most accuracies are concentrated on the lower end of the value range, but there are a few high-value accuracies pulling the distribution tail to the right. Accuracies of 3D-CNNs do not have median accuracy values around 90% for all domains as are the cases for 1D- and 2D-CNNs. Instead, the median accuracies are 96.17% for “Agriculture”, 69.44% for “Geohazards”, 86.53% for “Urban”, 95.18% for “Vegetation”, 95.94% for “Water”, and 98.78% for “Others”. Moreover, it shows slightly negative skewness in the “Vegetation” and “Water” domains, while it exhibits positive skewness in the “Agriculture” domain, with one accuracy as an outlier.

Based on the available median accuracy values, 3D-CNNs are proved to be the best in the “Agriculture” domain with a median accuracy of 96.17%. In contrast, in the “Geohazards” domain, it achieves a median accuracy of only 69.44%. In the “Geohazards”, “Urban”, and “Water” domains, 2D-CNNs achieve higher median accuracy compared to the other CNNs. For the “Vegetation” domain, all three CNNs show high median accuracy values. Based on the results, 1D-

CNNs and 2D-CNNs can be used for problems that require consistent performance across all domains, while 3D-CNNs are recommended for solving problems that require high accuracy in the “Agriculture” and “Water” domains.

The F1-score is the harmonic mean of the precision and recall metrics. Precision is the ratio of correctly predicted positive observations to the total number of predicted positive observations. Recall is the ratio of correctly predicted positive observations to all observations in the actual class. The best value achieved by the F1-score is 1 (perfect precision and recall) and the worst value is 0^[65]. Figure 21 displays a box and whisker plot of the F1-score metric for each CNN and application domain. Due to the unavailability of F1-score data for every individual publication, generalizing conclusions for all application domains to avoid bias is not feasible. However, domain-level analysis is conducted where data are available for all three CNNs. In the “Agriculture” domain, the median F1-scores are 85.33% for 1D-CNNs, 91% for 2D-CNNs, and 90.08% for 3D-CNNs. In the “Geohazards” domain, the median F1-scores are 93.96% for 1D-CNNs, 83.66% for 2D-CNNs, and 74.81% for 3D-CNNs. In the “Vegetation” domain, the median F1-scores are 94% for 1D-CNNs, 83.68% for 2D-CNNs, and 83.21% for 3D-CNNs. Based on these findings, we can single out 2D-CNNs and 3D-CNNs with F1-scores around 90% for “Agriculture”, suggesting that 2D-CNNs and 3D-CNNs are likely to capture spatial patterns in crop classifications. On the other hand, 1D-CNNs show high F1-scores (90%) compared to 2D-CNNs and 3D-CNNs in the “Vegetation” and “Geohazards” domains. These findings need to be

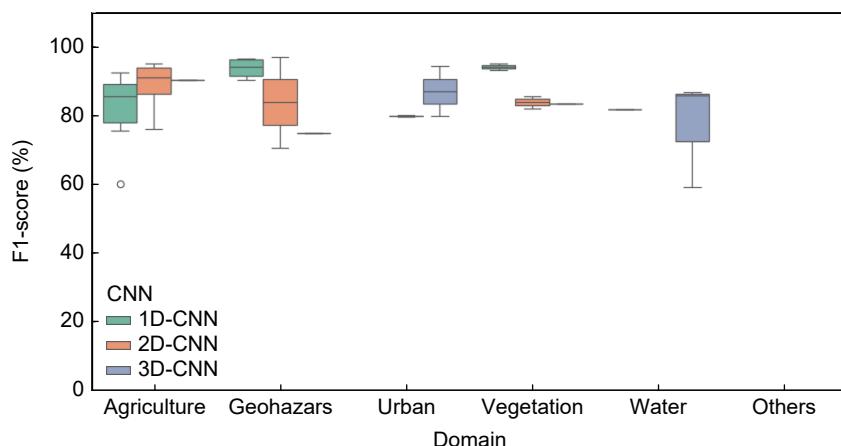


Fig. 21 F1-scores of CNNs using classification across different domains.

interpreted with caution because the F1-scores are imbalanced, preventing us from making a general conclusion.

Considering the large number of publications that used CNNs for classification, for better visibility of the work, details about individual publications as well as metrics describing CNN performance have been added to the ESM.

4.3.2 Analysis of publications implementing regression

Regression is a statistical method used to investigate the relationship between species and the environment, based on observations of species and environmental variables at a series of locations. In regression analysis, data on a particular species are analyzed separately and how they are related to environmental variables. The goal of regression analysis is to describe the response variable (abundance of species) as a function of one or more environmental explanatory variables. The response function cannot predict responses without errors, but efforts should be made to minimize those errors^[66]. In this systematic review, we extract certain metrics from publications used to evaluate models, such as R^2 , RMSE, and MAE. R^2 , also known as the coefficient of determination, determines the proportion of variance in the dependent variable that can be explained by the independent variables. It provides information on how well the observed values match the predicted values and can be expressed as a value or percentage, ranging from 0 to 1 or 0% to 100%. A value closer to 1 or 100% is desirable, indicating a better fit between the observed and predicted values. RMSE is the square root of the mean of the squares of all the errors. It indicates how close the line of best fit

is to the set of points. MAE provides the average value of the absolute difference between the observed values and the predicted values^[67].

Since regression predicts the value of the response variable for specific species, a general analysis is not possible. For example, although we have grouped publications by application domains, there is heterogeneity within the same domain (e.g., within the Water domain, parameters such as secchi and temperature can be analyzed). Therefore, averaging metric values is not recommended as it could mask important application-specific differences and lead to misinterpretation, resulting in a lack of validity in the comparison. Tables 3–5 highlight publications that have used regression with 1D-, 2D-, and 3D-CNNs, respectively. For each study, key information (e.g., publication type, year, satellite, and metrics such as R^2 , RMSE, and MAE) is presented and grouped by domain. Many values are missing. For those shown in Table 3, it can be seen that for the 1D-CNNs in the “Water” domain, the prediction of chlorophyll value expressed in milligrams per cubic meter is dominant in the regression analysis. For publications related to 2D-CNNs using regression (Table 4), most studies are in the “Others” domain, focusing on parameters related to image quality such as reflectance. As for the publications related to 3D-CNNs and their use of regression (Table 5), in addition to the “Others” domain, most of them are related to the “Agriculture” domain. The studies mainly focused on predicting soybean and crop yield parameter values.

4.3.3 Analysis of publications implementing segmentation

Image segmentation is a technique defined as the

Table 3 1D-CNN publications that implement regression.

Domain	Study	Publication type	Satellite	R^2	RMSE	MAE	Parameter
Agriculture	Jeong et al. ^[68]	Journal	MODIS	0.860	0.61	—	Rice yield (Mg/ha)
	Sabo et al. ^[69]	Journal	MODIS	—	—	—	—
Geohazard	Xu et al. ^[70]	Journal	MODIS	0.907	0.31	—	Drought
Urban	Vulova et al. ^[71]	Journal	Landsat-8	0.824	25.00	—	Reference evapotranspiration (mm/h)
Vegetation	Zhou et al. ^[72]	Journal	MODIS	—	—	—	Rice yield (Mg/ha)
	Fathi et al. ^[73]	Journal	Sentinel-2	0.745	6.09	4.90	Soybean yield (Bu/Ac)
Water	Maier et al. ^[74]	Journal	MODIS	0.624	19.30	14.60	Chl-a (mg/m ³)
			SeaWiFS	0.874	8.18	1.49	
			MERIS	0.915	12.52	1.47	Chl-a (mg/m ³)
			MODIS	0.894	6.73	1.45	
Others	Mukonza and Chiang ^[76]	Journal	Landsat-8	0.930	0.15	—	Temperature (degree)
			Sentinel-3	0.910	0.20	—	
	Salah et al. ^[77]	Journal	Sentinel-2	—	4.12	1.06	Chl-a (mg/m ³)
	Ivanda et al. ^[78]	Journal	Sentinel-3	0.890	2.30	1.40	Secchi (m)
	Zeng et al. ^[79]	Journal	MODIS	0.874	18.97	1.50	Chl-a (mg/m ³)
Others	Salah et al. ^[80]	Conference	Sentinel-2	—	11.87	6.71	Chl-a (mg/m ³)
	Ojaghi et al. ^[33]	Journal	Sentinel-3	0.970	3.00	—	Reflectance

Note: In the “Parameter” column, Mg/ha denotes metric tons per hectare, Bu/Ac denotes bushels per acre, and mg/m³ denotes milligrams per cubic meter.

Table 4 2D-CNN publications that implement regression.

Domain	Study	Publication type	Satellite	R^2	RMSE	MAE	Parameter
Agriculture	Sagan et al. ^[81]	Journal	WorldView-3, PlanetScope	—	—	—	—
	Nejad et al. ^[37]	Journal	MODIS	0.730	6.25	5.09	Crop yield (Bu/Ac)
	Sabo et al. ^[69]	Journal	MODIS	—	—	—	—
Geohazards	Lee et al. ^[82]	Journal	COMS	0.940	8.32	6.09	Maximum sustained wind speed (kts)
Vegetation	Fathi et al. ^[73]	Journal	Sentinel-2	0.758	—	—	Soybean yield (Bu/Ac)
Water	Zhong et al. ^[83]	Journal	Sentinel-2	0.900	1.03	—	Bathymetry (m)
Others	Papadomanolaki et al. ^[84]	Journal	WorldView-2, IKONOS	—	—	—	—
	Zhu et al. ^[52]	Journal	Landsat-8, MODIS	0.991	0.03	—	Reflectance
	Zhang et al. ^[85]	Journal	QuickBird	0.780	1.06	—	Similarity of spectral features
	Zhao et al. ^[53]	Journal	Landsat-8, HY-1C	0.792	—	—	Atmospheric correction
Others	Wang et al. ^[86]	Journal	MODIS, Sentinel-2, Landsat-8	—	—	—	—

Note: In the “Parameter” column, Bu/Ac denotes bushels per acre and kts denotes a knot that is a unit of speed equal to one nautical mile per hour.

process of dividing or partitioning an image into homogeneous parts, called segments. This is particularly useful for applications such as image compression or object detection, as processing the

entire image for these types of applications is inefficient. Therefore, image segmentation is used to segment parts of the image for further processing^[96]. According to Ref. [97], in remote sensing, it is often

Table 5 3D-CNN publications that implement regression.

Domain	Study	Publication type	Satellite	R^2	RMSE	MAE	Parameter
Agriculture	Terliksiz and Altýlar ^[87]	Conference	MODIS	–	0.81	–	Soybean yield (Bu/Ac)
	Qiao et al. ^[88]	Journal	MODIS	0.755	0.76	–	Crop yield (MT/ha)
	Sagan et al. ^[81]	Journal	WorldView-3, PlanetScope	–	–	–	–
	Nejad et al. ^[37]	Journal	MODIS	0.780	5.93	4.39	Crop yield (Bu/Ac)
	Wang et al. ^[89]	Conference	MODIS	–	5.47	–	Crop yield (kg/ha)
	Wang et al. ^[90]	Journal	MODIS	–	5.33	–	Crop yield (kg/ha)
Geohazards	Lee et al. ^[82]	Journal	COMS	0.886	11.34	8.65	Maximum sustained wind speed (kts)
Vegetation	Fernandez-Beltran et al. ^[91]	Journal	Sentinel-2	0.953	107.26	–	Crop yield (kg/ha)
Water	Fei et al. ^[92]	Journal	AVHRR	–	0.35	0.26	SST (degrees)
	Wang et al. ^[93]	Journal	MODIS	–	–	–	–
	Molini et al. ^[54]	Conference	PROBA-V	–	–	–	–
	Chen et al. ^[94]	Journal	GeoEye-1, WorldView-2	–	–	–	–
	Zhang et al. ^[49]	Journal	GaoFen-1, Sentinel-2	–	–	–	–
	He et al. ^[59]	Journal	WorldView-3	–	–	–	–
Others	Zhang et al. ^[95]	Journal	Himawari-7	0.865	0.31	–	Aerosol optical depth
	Ibrahim et al. ^[57]	Conference	PROBA-V	–	–	–	–
	Igeta and Iwasaki ^[60]	Conference	WorldView-3	–	–	–	–
	Zhu et al. ^[52]	Journal	Landsat-8, MODIS	0.991	0.03	–	Reflectance
	Zhang et al. ^[85]	Journal	QuickBird	0.787	0.10	–	Similarity of spectral features
	Zhao et al. ^[53]	Journal	HY-1C, Landsat-8	–	–	–	–

Note: In the “Parameter” column, Bu/Ac denotes bushels per acre, MT/ha denotes metric tons per hectare, kg/ha denotes kilograms per hectare, kts denotes a knot that is a unit of speed equal to one nautical mile per hour, and SST denotes sea surface temperature.

viewed as a tool for detecting landscape changes, and land use or land cover classification. In Table 6, we comprise all publications related to segmentation for different types of CNNs grouped by domains. It can be seen that even if this is a really small representative sample of sixteen publications, that the most publications are in domains “Agriculture” and “Vegetation”, which is in accordance to Ref. [97]. It is interesting that the most of these publications are using Landsat-8 data for segmentation. Also, we extract following metrics (if it is possible): producer accuracy, user accuracy, accuracy, F1-score, precision, recall, Kappa, IoU, MIoU, and Dice. Accuracy, F1-score, precision, and recall have same definition as for classification. Producer’s and user’s accuracies are commonly used in segmentation tasks. Producer’s accuracy indicates how well a training set pixel is classified for a given coverage type. User’s accuracy

indicates the probability that a pixel classified as belonging to a certain class actually represents that class on the ground. The Kappa coefficient indicates how much the classification is better compared to a classification where each pixel is randomly assigned a class value^[108]. The Dice coefficient is a metric used to compare the similarity of two samples. The Dice coefficient is twice the overlapping area of the two segmentations divided by the total number of pixels in the two images. IoU measures the overlapping area between the predicted segmentation and the true segmentation, representing the overlapping area divided by the union area of the predicted segmentation and the true segmentation. MIoU is the average of the IoU values calculated for each class^[109].

4.4 Limitations

This systematic review covers a wide range of

Table 6 Publications that implement segmentation.

CNN	Domain	Study and type	Satellite	Producer accuracy	User accuracy	Accuracy	F1-score	Precision	Recall	Kappa	IoU	MIoU	Dice	(%)
1D	Others	Bahl and Lafarge ^[98] , conference	Landsat-8	–	–	93.65	88.98	93.79	84.65	–	–	–	–	–
	Urban	Ghadorh et al. ^[99] , journal	Spot-7	–	–	–	–	–	–	–	–	–	79.51	88.65
2D	Vegetation	Li et al. ^[100] , journal	Gaofen-2	–	–	94.13	82.78	–	–	88.86	–	73.00	–	–
		Saralioglu and Gungor ^[101] , journal	WorldView-2	–	–	95.60	–	–	–	–	–	–	–	–
Agriculture		Mohammadi et al. ^[102] , conference	Landsat ARD	93.70	93.60	–	–	–	–	91.80	–	–	–	–
		Gallo et al. ^[103] , journal	Sentinel-2	–	–	70.33	70.30	70.35	70.33	64.00	–	53.02	–	–
		Mohammadi et al. ^[104] , journal	Landsat-7, Landsat-8	89.40	90.80	–	–	–	–	–	–	–	–	–
		Kalinicheva et al. ^[105] , journal	Sentinel-2, SPOT-5	–	–	–	–	–	–	–	–	–	–	–
3D		Meshkini et al. ^[106] , conference	Landsat-5, Landsat-7, Landsat-8	–	–	–	–	–	–	–	–	–	–	–
		Li et al. ^[100] , journal	Gaofen-2	–	–	97.45	89.49	–	–	94.68	–	82.45	–	–
		Meshkini et al. ^[107] , journal	Landsat-8	–	–	–	–	–	–	–	–	–	–	–
		Saralioglu and Gungor ^[101] , journal	WorldView-2	–	–	95.60	–	–	–	–	–	–	–	–
4D	Water	Wang et al. ^[61] , journal	MODIS	–	–	64.29	–	–	–	–	–	–	–	–
	Vegetation	Giannopoulos et al. ^[39] , journal	Landsat-8	–	–	61.56	61.56	61.56	61.56	–	44.47	–	–	–
		Giannopoulos et al. ^[42] , journal	Landsat-8	–	–	89.16	77.96	–	–	–	–	–	–	–
	Water	Wang et al. ^[61] , journal	MODIS	–	–	71.43	–	–	–	–	–	–	–	–

keywords related to the topic in three different databases, so it may have missed some relevant literature. For example, studies that do not explicitly mention the term “remote sensing” but only the name of the satellite could have been overlooked. Moreover, non-English studies are excluded, potentially affecting the geographic diversity of the included publications. Additionally, publications that are not open access or inaccessible could not be further examined to extract relevant information. Due to the large number of studies, specific knowledge gaps within each satellite, domain, type of CNN, and machine learning technique are not extensively explored. A future review could narrow its focus, allowing for a more detailed assessment of knowledge gaps, such as concentrating on a single type of CNN within a specific domain.

Furthermore, we do not assess the risk of bias or quality of the reviewed studies. Additional details regarding all studies reviewed in this article are available in the ESM. Moreover, CSV data containing all extracted information, sourced from the publications, are available from the authors. Finally, this study can be seen as a starting point by researchers to identify which CNNs and satellites are predominantly used in specific remote sensing applications.

5 Conclusion

The objective of this systematic review is to assess the literature on the possibilities, challenges and gaps in the use of deep learning models, specifically 1D-, 2D-, 3D-, and 4D-CNNs, on multispectral imaging data. In

particular, this research investigates the trends in the application domain of CNNs, the type of satellite data, and the machine learning technique used. As a general conclusion from all publications, we can state that most of the papers are published in peer-reviewed journals and that most of them use a 3D-CNN implementations. In addition, the majority of publications on 1D-, 2D-, and 3D-CNNs focus on solving problems in the field of agriculture and use Sentinel-2 satellite data. Regarding 4D-CNNs, it is not possible to draw general conclusions due to the small sample size (only three publications). The predominant machine learning technique used by CNNs is classification. In the “Geohazards”, “Urban”, and “Water” domains, 2D-CNNs achieve a higher median accuracy compared to the other CNNs (94.09%, 95%, and 89.18%, respectively). 3D-CNNs prove to be the best in agriculture (96.7%). Due to the small number of publications and the diversity of problems within the domains addressed by CNNs for the machine learning regression technique, no general conclusions could be drawn. The same applies to the publications that use the machine learning segmentation technique, as there are only 16 studies.

To summarize, a thorough review of the existing literature reveals a considerable gap: no previous research has compared the performance of 1D, 2D, 3D, and 4D convolutional neural networks on the same task or even on similar tasks. This lack emphasizes the need to conduct comprehensive benchmarking studies to evaluate and compare the performance of these four CNN model types. Such studies should be designed to evaluate the models on certain representative tasks with the same dataset to ensure a fair and systematic comparison. This effort is crucial for advancing our understanding of the strengths and limitations of each CNN type and for guiding future applications and developments in this field.

Electronic Supplementary Material

The full datasets are available from the GitHub repository and the following supporting materials are available in the online version of this article at <https://doi.org/10.26599/BDMA.2024.9020086>.

- Table S1 PRISMA 2020 checklist;
- Table S2 PRISMA 2020 abstract checklist;
- Table S3 Web of Science search strategy;
- Table S4 IEEE Xplore search strategy;

- Table S5 Scopus search strategy;
- Table S6 Publications used in analysis—1D-CNN;
- Table S7 Publications used in analysis—2D-CNN;
- Table S8 Publications used in analysis—3D-CNN;
- Table S9 Publications used in analysis—4D-CNN.

Acknowledgment

This research was supported through Project Coastal Auto-purification Assessment Technology (CAAT), funded by the European Union from European Structural and Investment Funds 2014–2020 (No. KK.01.1.1.04.0064).

References

- [1] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data*. London, UK: CRC Press, 2001.
- [2] Z. Adiri, R. Lhissou, A. El Harti, A. Jellouli, and M. Chakouri, Recent advances in the use of public domain satellite imagery for mineral exploration: A review of landsat-8 and sentinel-2 applications, *Ore Geol. Rev.*, vol. 117, p. 103332, 2020.
- [3] A. Jamali and M. Mahdianpari, A cloud-based framework for large-scale monitoring of ocean plastics using multi-spectral satellite imagery and generative adversarial network, *Water*, vol. 13, no. 18, p. 2553, 2021.
- [4] D. Walshe, D. McInerney, R. van de Kerchove, C. Goyens, P. Balaji, and K. A. Byrne, Detecting nutrient deficiency in spruce forests using multispectral satellite imagery, *Int. J. Appl. Earth Obs. Geoinform.*, vol. 86, p. 101975, 2020.
- [5] F. Marchese, N. Genzano, M. Neri, A. Falconieri, G. Mazzeo, and N. Pergola, A multi-channel algorithm for mapping volcanic thermal anomalies by means of sentinel-2 MSI and landsat-8 OLI data, *Remote Sens.*, vol. 11, no. 23, p. 2876, 2019.
- [6] Z. Yi, L. Jia, and Q. Chen, Crop classification using multi-temporal sentinel-2 data in the Shiyang river basin of China, *Remote Sens.*, vol. 12, no. 24, p. 4052, 2020.
- [7] C. Sun, Y. Bian, T. Zhou, and J. Pan, Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region, *Sensors*, vol. 19, no. 10, p. 2401, 2019.
- [8] P. Lynch, L. Blesius, and E. Hines, Classification of urban area using multispectral indices for urban planning, *Remote Sens.*, vol. 12, no. 15, p. 2503, 2020.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [10] L. Li, C. Solana, F. Canters, and M. Kervyn, Testing

- random forest classification for identifying lava flows and mapping age groups on a single Landsat 8 image, *J. Volcanol. Geotherm. Res.*, vol. 345, pp. 109–124, 2017.
- [11] L. Viskovic, I. N. Kosovic, and T. Mastelic, Crop classification using multi-spectral and multitemporal satellite imagery with machine learning, in *Proc. 2019 Int. Conf. Software, Telecommunications and Computer Networks*, Split, Croatia, 2019, pp. 1–5.
- [12] X. Huang and L. Zhang, An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, 2013.
- [13] A. da Penha Pacheco, J. A. da Silva Junior, A. M. Ruiz-Armenteros, and R. F. F. Henriques, Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central Portugal using landsat-8, sentinel-2, and terra imagery, *Remote Sens.*, vol. 13, no. 7, p. 1345, 2021.
- [14] A. Ivanda, L. Šerić, M. Bugarić, and M. Braović, Mapping chlorophyll-a concentrations in the Kaštela bay and Brač channel using ridge regression and sentinel-2 satellite images, *Electronics*, vol. 10, no. 23, p. 3004, 2021.
- [15] P. Das and V. Pandey, Use of logistic regression in land-cover classification with moderate-resolution multispectral data, *J. Indian Soc. Remote Sens.*, vol. 47, no. 8, pp. 1443–1454, 2019.
- [16] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS J. Photogramm. Remote Sens.*, vol. 152, pp. 166–177, 2019.
- [17] X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, Deep learning in remote sensing: A comprehensive review and list of resources, *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, 2017.
- [18] L. Mou, L. Bruzzone, and X. X. Zhu, Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery, *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, 2019.
- [19] S. M. Borzov and O. I. Potaturkin, Spectral-spatial methods for hyperspectral image classification. review, *Optoelectron. Instrum. Data Process.*, vol. 54, no. 6, pp. 582–599, 2018.
- [20] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, A survey: Deep learning for hyperspectral image classification with few labeled samples, *Neurocomputing*, vol. 448, pp. 179–204, 2021.
- [21] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, Modern trends in hyperspectral image analysis: A review, *IEEE Access*, vol. 6, pp. 14118–14129, 2018.
- [22] X. Yuan, J. Shi, and L. Gu, A review of deep learning methods for semantic segmentation of remote sensing imagery, *Expert Syst. Appl.*, vol. 169, p. 114417, 2021.
- [23] S. Bera, V. K. Shrivastava, and S. C. Satapathy, Advances in hyperspectral image classification based on convolutional neural networks: A review, *CMES Comput. Model. Eng. Sci.*, vol. 133, no. 2, pp. 219–250, 2022.
- [24] Y. Li, H. Zhang, X. Xue, Y. Jiang, and Q. Shen, Deep learning for remote sensing image classification: A survey, *WIREs Data Min. Knowledge Discovery*, vol. 8, no. 6, p. e1264, 2018.
- [25] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, Review on convolutional neural networks (CNN) in vegetation remote sensing, *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021.
- [26] L. Zhang, L. Zhang, and B. Du, Deep learning for remote sensing data: A technical tutorial on the state of the art, *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, 2016.
- [27] J. E. Ball, D. T. Anderson, and C. S. C. Sr, Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community, *J. Appl. Remote Sens.*, vol. 11, no. 4, p. 042609, 2017.
- [28] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., The PRISMA 2020 statement: An updated guideline for reporting systematic reviews, *Int. J. Surg.*, vol. 88, p. 105906, 2021.
- [29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning: Convolutional Networks*. Cambridge, MA, USA: MIT Press, 2016.
- [30] H. Gu, Y. Wang, S. Hong, and G. Gui, Blind channel identification aided generalized automatic modulation recognition based on deep learning, *IEEE Access*, vol. 7, pp. 110722–110729, 2019.
- [31] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, *J. Big Data*, vol. 8, no. 1, p. 53, 2021.
- [32] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, Deep convolutional neural networks for hyperspectral image classification, *J. Sens.*, vol. 2015, p. 258619, 2015.
- [33] S. Ojaghi, Y. Bouroubi, S. Foucher, M. Bergeron, and C. Seynat, Deep learning-based emulation of radiative transfer models for top-of-atmosphere BRDF modelling using sentinel-3 OLCI, *Remote Sens.*, vol. 15, no. 3, p. 835, 2023.
- [34] L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. M. Ramos, J. Li, and J. Marcato Junior, The segment anything model (SAM) for remote sensing applications: From zero to one shot, *Int. J. Appl. Earth Obs. Geoinform.*, vol. 124, p. 103540, 2023.
- [35] M. G. Park and N. W. Park, Application of deep learning algorithms considering spatio-temporal features for crop classification, in *Proc. 40th Asian Conf. Remote Sensing*, Daejeon, Republic of Korea, 2019, <http://acrs2019.sigongji.com/wp/pdf/ThP-87.pdf>, 2023.
- [36] X. Zhou, W. Zhou, F. Li, Z. Shao, and X. Fu, Vegetation

- type classification based on 3D convolutional neural network model: A case study of Baishuijiang national nature reserve, *Forests*, vol. 13, no. 6, p. 906, 2022.
- [37] S. M. M. Nejad, D. Abbasi-Moghadam, A. Sharifi, N. Farmonov, K. Amankulova, and M. László, Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 254–266, 2023.
- [38] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, 3D convolutional neural networks for crop classification with multi-temporal remote sensing images, *Remote Sens.*, vol. 10, no. 1, p. 75, 2018.
- [39] M. Giannopoulos, G. Tsagkatakis, and P. Tsakalides, 4D U-Nets for multi-temporal remote sensing data classification, *Remote Sens.*, vol. 14, no. 3, p. 634, 2022.
- [40] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., TensorFlow: Large-scale machine learning on heterogeneous distributed systems, arXiv preprint arXiv: 1603.04467, 2016.
- [41] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., PyTorch: An imperative style, high-performance deep learning library, in *Proc. 33rd Conf. Neural Information Processing Systems*, Vancouver, Canada, 2019, doi: 10.48550/arXiv.1912.01703.
- [42] M. Giannopoulos, G. Tsagkatakis, and P. Tsakalides, 4D convolutional neural networks for multi-spectral and multi-temporal remote sensing data classification, in *Proc. 2022 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Singapore, 2022, pp. 1541–1545.
- [43] E. O. Box and K. Fujiwara, Vegetation types and their broad-scale distribution, in *Vegetation Ecology*, E. van der Maarel and J. Franklin, eds, 2nd ed. Chichester, England: John Wiley & Sons, Ltd., 2013, pp. 455–485.
- [44] M. Weiss, F. Jacob, and G. Duveiller, Remote sensing for agricultural applications: A meta-review, *Remote Sens. Environ.*, vol. 236, p. 111402, 2020.
- [45] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, Sentinel-2 data for land cover/use mapping: A review, *Remote Sens.*, vol. 12, no. 14, p. 2291, 2020.
- [46] D. P. Roy, M. A. Wulder, T. R. Loveland, C. E. Woodcock, R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, et al., Landsat-8: Science and product vision for terrestrial global change research, *Remote Sens. Environ.*, vol. 145, pp. 154–172, 2014.
- [47] A. Savtchenko, D. Ouzounov, S. Ahmad, J. Acker, G. Leptoukh, J. Koziana, and D. Nickless, Terra and aqua MODIS products available from NASA GES DAAC, *Adv. Space Res.*, vol. 34, no. 4, pp. 710–714, 2004.
- [48] Y. Chen, L. Tang, Z. Kan, A. Latif, X. Yang, M. Bilal, and Q. Li, Cloud and cloud shadow detection based on multiscale 3D-CNN for high resolution multispectral imagery, *IEEE Access*, vol. 8, pp. 16505–16516, 2020.
- [49] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images, *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 161–173, 2021.
- [50] H. Jiang and N. Lu, Multi-scale residual convolutional neural network for haze removal of remote sensing images, *Remote Sens.*, vol. 10, no. 6, p. 945, 2018.
- [51] F. Kong, K. Hu, Y. Li, D. Li, and S. Zhao, Spectral-spatial feature partitioned extraction based on CNN for multispectral image compression, *Remote Sens.*, vol. 13, no. 1, p. 9, 2021.
- [52] Z. Zhu, Y. Tao, and X. Luo, HCNNet: A hybrid convolutional neural network for spatiotemporal image fusion, *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 2005716, 2022.
- [53] X. Zhao, Y. Ma, Y. Xiao, J. Liu, J. Ding, X. Ye, and R. Liu, Atmospheric correction algorithm based on deep learning with spatial-spectral feature constraints for broadband optical satellites: Examples from the HY-1C coastal zone imager, *ISPRS J. Photogramm. Remote Sens.*, vol. 205, pp. 147–162, 2023.
- [54] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, Deep learning for super-resolution of unregistered multi-temporal satellite images, in *Proc. 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing*, Amsterdam, the Netherlands, 2019, pp. 1–5.
- [55] F. Salvetti, V. Mazzia, A. Khaliq, and M. Chiaberge, Multi-image super resolution of remotely sensed images using residual attention deep neural networks, *Remote Sens.*, vol. 12, no. 14, p. 2207, 2020.
- [56] F. Dorr, Satellite image multi-frame super resolution using 3D wide-activation neural networks, *Remote Sens.*, vol. 12, no. 22, p. 3812, 2020.
- [57] M. R. Ibrahim, R. Benavente, F. Lumbrales, and D. Ponsa, 3DRRDB: Super resolution of multiple remote sensing images using 3D residual in residual dense blocks, in *Proc. 2022 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 322–331.
- [58] N. Li, R. Wang, H. Zhao, and W. Wei, An improved feature extraction method based on context features for multi-spectral remote sensing imagery, in *Proc. 2019 IEEE Int. Conf. Signal, Information and Data Processing*, Chongqing, China, 2019, pp. 1–4.
- [59] S. He, R. Zhou, S. Li, S. Jiang, and W. Jiang, Disparity estimation of high-resolution remote sensing images with dual-scale matching network, *Remote Sens.*, vol. 13, no. 24, p. 5050, 2021.
- [60] T. Igeta and A. Iwasaki, An unsupervised network for stereo matching of very high resolution satellite imagery, in *Proc. 2022 IEEE Int. Geoscience and Remote Sensing Symp.*, Kuala Lumpur, Malaysia, 2022, pp. 971–974.
- [61] L. Wang, X. Wang, Z. Zhao, Y. Wu, J. Xu, H. Zhang, J. Yu, Q. Sun, and Y. Bai, Multi-factor status prediction by

- 4D fractal CNN based on remote sensing images, *Fractals*, vol. 30, no. 2, p. 2240101, 2022.
- [62] A. D. Gordon, *Classification*. New York, NY, USA: CRC Press, 1999.
- [63] J. Al-Doski, S. B. Mansorl, and H. Z. M. Shafri, Image classification in remote sensing, *J. Environ. and Earth Sci.*, <http://core.ac.uk/download/pdf/234663192.pdf>, 2023.
- [64] M. Hossin and M. N. Sulaiman, A review on evaluation metrics for data classification evaluations, *Int. J. Data Min. Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, 2015.
- [65] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms: A Classification Perspective*. New York, NY, USA: Cambridge University Press, 2011.
- [66] C. J. F. T. Braak and C. W. N. Looman, Regression, in *Data Analysis in Community and Landscape Ecology*, R. H. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, eds. New York, NY, USA: Cambridge University Press, 1995, pp. 29–77.
- [67] A. V. Tatachar, Comparative assessment of regression models based on model evaluation metrics, *Int. Res. J. Eng. Technol.*, vol. 8, no. 9, pp. 853–860, 2021.
- [68] S. Jeong, J. Ko, and J. M. Yeom, Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in south and North Korea, *Sci. Total Environ.*, vol. 802, p. 149726, 2022.
- [69] F. Sabo, M. Meroni, F. Waldner, and F. Rembold, Is deeper always better? Evaluating deep learning models for yield forecasting with small data, *Environ. Monit. Assess.*, vol. 195, no. 10, p. 1153, 2023.
- [70] Z. Xu, H. Sun, T. Zhang, H. Xu, D. Wu, and J. Gao, Evaluating established deep learning methods in constructing integrated remote sensing drought index: A case study in China, *Agric. Water Manage.*, vol. 286, p. 108405, 2023.
- [71] S. Vulova, F. Meier, A. D. Rocha, J. Quanz, H. Nouri, and B. Kleinschmit, Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence, *Sci. Total Environ.*, vol. 786, p. 147293, 2021.
- [72] X. Zhou, Q. Xin, Y. Dai, and W. Li, A deep-learning-based experiment for benchmarking the performance of global terrestrial vegetation phenology models, *Global Ecol. Biogeogr.*, vol. 30, no. 11, pp. 2178–2199, 2021.
- [73] M. Fathi, R. Shah-Hosseini, and A. Moghimi, 3D-ResNet-BILSTM model: A deep learning model for county-level soybean yield prediction with time-series sentinel-1, sentinel-2 imagery, and daymet data, *Remote Sens.*, vol. 15, no. 23, p. 5551, 2023.
- [74] P. M. Maier, S. Keller, and S. Hinz, Deep learning with WASI simulation data for estimating chlorophyll a concentration of inland water bodies, *Remote Sens.*, vol. 13, no. 4, p. 718, 2021.
- [75] D. Fan, H. He, R. Wang, Y. Zeng, B. Fu, Y. Xiong, L. Liu, Y. Xu, and E. Gao, CHLNET: A novel hybrid 1D CNN-SVR algorithm for estimating ocean surface chlorophyll-a, *Front. Mar. Sci.*, vol. 9, p. 934536, 2022.
- [76] S. S. Mukonza and J. L. Chiang, Micro-climate computed machine and deep learning models for prediction of surface water temperature using satellite data in Mundan water reservoir, *Water*, vol. 14, no. 18, p. 2935, 2022.
- [77] M. Salah, H. Higa, J. Ishizaka, and S. I. Salem, 1D convolutional neural network-based chlorophyll-a retrieval algorithm for sentinel-2 multispectral instrument in various trophic states, *Sens. Mater.*, vol. 35, no. 11, pp. 3743–3761, 2023.
- [78] A. Ivanda, L. Šerića, D. Žagar, and K. Oštir, An application of 1D convolution and deep learning to remote sensing modelling of secchi depth in the northern Adriatic sea, *Big Earth Data*, vol. 8, no. 1, pp. 82–114, 2024.
- [79] Y. Zeng, T. Liang, D. Fan, and H. He, A novel algorithm for the retrieval of chlorophyll a in marine environments using deep learning, *Water*, vol. 15, no. 21, p. 3864, 2023.
- [80] M. Salah, H. Higa, J. Ishizaka, and S. I. Salem, B1D-CNN: A novel convolution neural network-based chlorophyll-a retrieval algorithm for sentinel-2 data, in *Proc. 2023 IEEE Int. Geoscience and Remote Sensing Symp.*, Pasadena, CA, USA, 2023, pp. 3950–3953.
- [81] V. Sagan, M. Maimaitijiang, S. Bhadra, M. Maimaitiyiming, D. R. Brown, P. Sidike, and F. B. Fritschi, Field-scale crop yield prediction using multi-temporal worldview-3 and planetscope satellite data and deep learning, *ISPRS J. Photogramm. Remote Sens.*, vol. 174, pp. 265–281, 2021.
- [82] J. Lee, J. Im, D. H. Cha, H. Park, and S. Sim, Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data, *Remote Sens.*, vol. 12, no. 1, p. 108, 2020.
- [83] J. Zhong, J. Sun, Z. Lai, and Y. Song, Nearshore bathymetry from icesat-2 lidar and sentinel-2 imagery datasets using deep learning approach, *Remote Sens.*, vol. 14, no. 17, p. 4229, 2022.
- [84] M. Papadomanolaki, S. Christodoulidis, K. Karantzalos, and M. Vakalopoulou, Unsupervised multistep deformable registration of remote sensing imagery based on deep learning, *Remote Sens.*, vol. 13, no. 7, p. 1294, 2021.
- [85] E. Zhang, Y. Fu, J. Wang, L. Liu, K. Yu, and J. Peng, MSAC-Net: 3D multi-scale attention convolutional network for multi-spectral imagery pansharpening, *Remote Sens.*, vol. 14, no. 12, p. 2761, 2022.
- [86] Z. Wang, S. Fang, and J. Zhang, Spatiotemporal fusion model of remote sensing images combining single-band and multi-band prediction, *Remote Sens.*, vol. 15, no. 20, p. 4936, 2023.
- [87] A. S. Terliksiz and D. T. Altýlar, Use of deep neural networks for crop yield prediction: A case study of soybean yield in lauderdale county, Alabama, USA, in *Proc. 2019 8th Int. Conf. Agro-Geoinformatics*, Istanbul, Turkey, 2019, pp. 1–4.

- [88] M. Qiao, X. He, X. Cheng, P. Li, H. Luo, L. Zhang, and Z. Tian, Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks, *Int. J. Appl. Earth Obs. Geoinform.*, vol. 102, p. 102436, 2021.
- [89] N. Wang, Z. Ma, P. Huo, and X. Liu, Predicting crop yield using 3D convolutional neural network with dimension reduction and metric learning, in *Proc. 2023 IEEE 6th Int. Conf. Pattern Recognition and Artificial Intelligence*, Haikou, China, 2023, pp. 1004–1008.
- [90] N. Wang, Z. Ma, P. Huo, X. Liu, Z. He, and K. Lu, 3D convolutional neural network with dimension reduction and metric learning for crop yield prediction based on remote sensing data, *Appl. Sci.*, vol. 13, no. 24, p. 13305, 2023.
- [91] R. Fernandez-Beltran, T. Baidar, J. Kang, and F. Pla, Rice-yield prediction with multi-temporal sentinel-2 data and 3D CNN: A case study in Nepal, *Remote Sens.*, vol. 13, no. 7, p. 1391, 2021.
- [92] T. Fei, B. Huang, X. Wang, J. Zhu, Y. Chen, H. Wang, and W. Zhang, A hybrid deep learning model for the bias correction of SST numerical forecast products using satellite data, *Remote Sens.*, vol. 14, no. 6, p. 1339, 2022.
- [93] L. Wang, W. Li, X. Wang, and J. Xu, Remote sensing image analysis and prediction based on improved Pix2Pix model for water environment protection of smart cities, *PeerJ Comput. Sci.*, vol. 9, p. e1292, 2023.
- [94] G. Chen, Q. Pei, and M. Kamruzzaman, Remote sensing image quality evaluation based on deep support value learning networks, *Signal Process Image Commun.*, vol. 83, p. 115783, 2020.
- [95] L. Zhang, P. Liu, L. Wang, J. Liu, B. Song, Y. Zhang, G. He, and H. Zhang, Improved 1-km-resolution hourly estimates of aerosol optical depth using conditional generative adversarial networks, *Remote Sens.*, vol. 13, no. 19, p. 3834, 2021.
- [96] D. Kaur and Y. Kaur, Various image segmentation techniques: A review, *Int. J. Comput. Sci. Mobile Comput.*, vol. 3, no. 5, pp. 809–814, 2014.
- [97] V. Dey, Y. Zhang, and M. Zhong, A review on image segmentation techniques with remote sensing perspective, in *Proc. of the International Society for Photogrammetry and Remote Sensing Symposium*, <https://api.semanticscholar.org/CorpusID:19014182>, 2023.
- [98] G. Bahl and F. Lafarge, Scanner neural network for on-board segmentation of satellite images, in *Proc. 2022 IEEE Int. Geoscience and Remote Sensing Symp.*, Kuala Lumpur, Malaysia, 2022, pp. 3504–3507.
- [99] H. Ghadorh, W. Boulila, S. Masood, A. Koubaa, F. Ahmed, and J. Ahmad, Semantic segmentation and edge detection—approach to road detection in very high resolution satellite images, *Remote Sens.*, vol. 14, no. 3, p. 613, 2022.
- [100] R. Li, S. Zheng, C. Duan, L. Wang, and C. Zhang, Land cover classification from remote sensing images based on multi-scale fully convolutional network, *Geo-Spat. Inf. Sci.*, vol. 25, no. 2, pp. 278–294, 2022.
- [101] E. Saralioglu and O. Gungor, Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network, *Geocarto Int.*, vol. 37, no. 2, pp. 657–677, 2022.
- [102] S. Mohammadi, M. Belgiu, and A. Stein, 3D fully convolutional neural networks with intersection over union loss for crop mapping from multi-temporal satellite images, in *Proc. 2021 IEEE Int. Geoscience and Remote Sensing Symp.*, Brussels, Belgium, 2021, pp. 5834–5837.
- [103] I. Gallo, L. Ranghetti, N. Landro, R. L. Grassi, and M. Boschetti, In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series, *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 335–352, 2023.
- [104] S. Mohammadi, M. Belgiu, and A. Stein, Improvement in crop mapping from satellite image time series by effectively supervising deep neural networks, *J. Photogramm. Remote Sens.*, vol. 198, pp. 272–283, 2023.
- [105] E. Kaliniccheva, J. Sublime, and M. Trocan, Unsupervised satellite image time series clustering using object-based approaches and 3D convolutional autoencoder, *Remote Sens.*, vol. 12, no. 11, p. 1816, 2020.
- [106] K. Meshkini, F. Bovolo, and L. Bruzzone, An unsupervised change detection approach for dense satellite image time series using 3D CNN, in *Proc. 2021 IEEE Int. Geoscience and Remote Sensing Symp.*, Brussels, Belgium, 2021, pp. 4336–4339.
- [107] K. Meshkini, F. Bovolo, L. Bruzzone, A 3D CNN approach for change detection in HR satellite image time series based on a pretrained 2D CNN, in *Proc. Int. Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, Nice, France, 2022, doi: 10.5194/isprs-archives-XLIII-B3-2022-143-2022.
- [108] M. R. Rahman and S. K. Saha, Multi-resolution segmentation for object-based classification and accuracy assessment of land use/land cover classification using remotely sensed data, *J. Indian Soc. Remote Sens.*, vol. 36, no. 2, pp. 189–201, 2008.
- [109] S. Park and N. W. Park, Effects of class purity of training patch on classification performance of crop classification with convolutional neural network, *Appl. Sci.*, vol. 10, no. 11, p. 3773, 2020.



Antonia Ivanda received the BEng and MEng degrees in computer science from University of Split, Croatia in 2017 and 2019, respectively. She is currently a PhD candidate in electrical engineering and information technology at University of Split, Croatia. Her research lies in the field of computer science, with a primary focus on artificial intelligence for data analysis and prediction of bathing water quality, as well as assessment of autopurification capabilities using a remote sensing data-driven approach. Additionally, she is an expert in using qGIS software for spatial visualization of vector and raster data obtained via satellites. She is the author or co-author of more than 20 scientific publications.



Maja Braović received the BEng and MEng degrees in computer science, and the PhD degree in artificial intelligence from University of Split, Croatia in 2008, 2010, and 2015, respectively. She is currently an associate professor at Department of Electronic and Computer Science, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Croatia. Her main research interests include artificial intelligence, computer vision, image understanding, and natural language processing. She is the author or co-author of more than 30 scientific publications.



Ljiljana Šerić is full professor at Department of Electronics and Computer Science, Faculty of Electrical Engineering, Mechanical Engineering, and Naval Architecture, University of Split, Croatia. She participates in the activities of Laboratory for Intelligent Systems and Laboratory for Advanced Internet Technologies, and a collaborator of Centre for Wildfire Research, University of Split, Croatia. She received the PhD degree in computer science from University of Split, Croatia in 2010. She has worked on scientific, specialistic, and technical projects in the domain of environmental disasters protection and environmental hazards prevention. She has participated in the projects dealing with wildfires, high winds, and coastal pollutions. Her research interests are artificial intelligence, web technologies, and distributed systems.