ICIC Express Letters
Part B: Applications
Volume **15**, Number **2**, February **2024**

ICIC International ©2024 ISSN 2185-2766
pp. 195–205

SATELLITE IMAGE DATASET TAXONOMY: A MULTIDIMENSIONAL FRAMEWORK FOR EFFECTIVE DATASET UTILIZATION

Sehyoung Kim[1], Jaehyeong Park[1], Seyeon Cheon[1], Donggeon Kim[1]
and Juyoung Kang[2,*]

[1]Department of Business Analytics
[2]Department of e-Business
School of Business
Ajou University
206 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea
{ sehyoung66; qkrwogud; cjstpdus9; kevin980526 }@ajou.ac.kr
*Corresponding author: jykang@ajou.ac.kr

Received May 2023; accepted July 2023

Abstract. *Satellite imagery datasets are employed in research and industry. A classification system must apply the dataset according to its purpose and intention to increasing the utilization value of satellite imagery datasets. Therefore, we propose a method to organize and classify satellite imagery datasets, departing from existing research trends. Thus, this study developed a classification system using currently available satellite image datasets called a taxonomy, consisting of eight dimensions to describe satellite imagery (image quality, satellite type, label information, dataset acquisition channel, coverage area, time, applicable models, and available fields) and a table containing characteristics comprising each dimension. This paper details the iteration process of the development stage of the taxonomy and the theoretical background and evidence used to create the content for each dimension. Examples of how the taxonomy classifies actual datasets and how it can be applied in practice are included to provide practical guidelines for using satellite image datasets.*

Keywords: Taxonomy, Satellite imagery, Deep learning, Information systems, Satellite image datasets

1. **Introduction.** With the advancement of computing technology and deep learning algorithms, research using satellite imagery datasets has increased, and the applications have become more diverse [1]. Technological advancements influence these changes in the satellite and space industry. Small satellite launches have increased significantly, from 389 in 2019 to 1,202 in 2020 and 2,304 in 2022 [2]. The development of the space industry accelerated in 2022, with 180 successful rocket launches into space orbit, 44 more than in 2021 [3]. The growing interest in the satellite industry has also increased related business models and services in the private sector. Especially in recent years, as available satellite image datasets, computing technology, and deep learning algorithms have been rapidly developing, efforts to improve the performance of object detection, image/target object classification, and semantic analysis have been accumulating as research results. Many research cases have been identified that employ these results for observation activities. Figure 1 illustrates research topics related to satellite images, primarily employed to observe a wide area, such as remote sensing, but also used in various fields, including natural sciences.

The scope and frequency of research using satellite imagery datasets have rapidly increased, but research on applying them to real-world industries or developing business models is still lacking.

DOI: 10.24507/icicelb.15.02.195

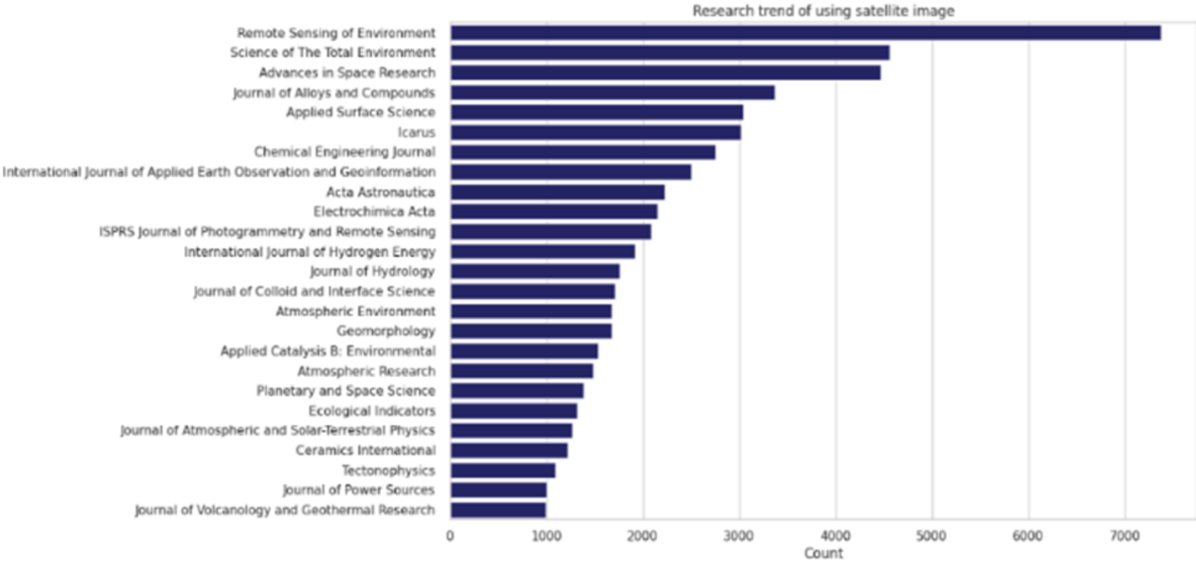

FIGURE 1. Research trends in satellite imagery

The use of satellite imagery datasets for academic research or industry should be done using a systematic approach. In addition, satellite imagery data are expected to be used more in the future than in the present, where access is relatively limited compared to other imagery data; thus, a tool is needed that can easily access information from the user perspective, including researchers unfamiliar with the domain. This study proposes the following research question to design such a concept: *What analytical techniques can be used for satellite image datasets and in which fields?*

To answer this question, we identified and organized the types of satellite datasets currently available, creating a classification scheme. A consideration in creating a taxonomy is to develop a general-purpose taxonomy that can classify new datasets, even if new objects are added. Therefore, we used the method of creating a taxonomy proposed by Nickerson et al. [8] to address the research question. Many studies have been conducted to develop taxonomies because they can be useful in all stages of knowledge discovery, management, and utilization.

This study developed a real-world taxonomy for using satellite image datasets through an example that details our considerations when creating a taxonomy for practical application. Previous studies on satellite imagery dataset taxonomy focus on providing recommendations on the appropriate algorithms to use [8]. In contrast, our work evaluates the research dimensions that can expand the work conducted with this data, extending the aspects of satellite imagery datasets. Thus, researchers can apply our study's satellite image dataset taxonomy to develop business models or create value in more fields. The motivations and contributions from the research results are summarized as follows.

1) The number of public and private open datasets related to satellites and the range of information they provide will increase, so a classification system is necessary to improve utilization.

2) This can reduce the time spent searching for information, as academia and industry must use each dataset for their purposes. If one aims to analyze information but does not know which dataset to select, the taxonomy can provide a select option for all satellite image datasets.

3) The satellite utilization industry is a promising field that can be further developed, so organizing or discussing the taxonomy at this stage is necessary.

4) Research trends using satellite image datasets, a high-demand research area, include review papers that synthesize the performance of special-purpose tasks, such as remote

sensing, indicating the need for a systematic approach to using satellite imagery. More efficient research could be conducted if an information system is established through taxonomy.

This study is organized as follows. Section 2 introduces research cases using satellite imagery and how to develop taxonomies in information systems. Next, Section 3 presents the research framework and describes the process of creating the satellite image dataset taxonomy according to the development sequence of the taxonomy, and Section 4 discusses the developed taxonomy and utilization cases. Section 5 presents taxonomy application. The final chapter summarizes the contributions and limitations of this research.

2. **Literature Review.**

2.1. **Studies on satellite image dataset.**

As the number of satellites worldwide has increased and sensor technology has advanced, high-resolution satellite imagery datasets have been built for various research studies and industries. With the abundance of satellite imagery datasets and the development of deep learning technology, satellite imagery-related research has been actively conducted in remote sensing. Ma et al. [4] conducted a meta-analysis of remote sensing studies with satellite imagery and deep learning techniques. Their research classified, visualized, and summarized specific models and applications and included pixels of the satellite image dataset. However, they do not clearly define how to use satellite imagery datasets for research or real-world applications.

In addition to this information, we present the idea of taxonomy with deep learning in remote sensing as the top layer but do not provide details on the classification process between the layers or the process of creating the taxonomy. This study implies that various satellite imagery datasets are used to apply remote sensing, and the taxonomy organizes the review process of numerous related studies. In addition, detailed examples of use, such as land-cover classification and night-light observation, are broadly part of remote sensing and are considered good research topics. Do et al. [5] conducted a study to measure the economic effects of mega-sporting events, such as the Olympic Games, based on unit-time regional changes in Olympic venues. They presented a methodology to estimate regional development based on changes in land-cover classification in the target area using annually collected satellite image datasets. Khan et al. [6] examined long-term land-cover changes in disaster-affected areas to investigate the combined effects of land-use change and natural disasters using land cover. Similar methodologies can be applied to many fields and topics, and many datasets can address the same problem. Therefore, classifying satellite imagery datasets and suggesting appropriate uses in research is essential for future research fields related to satellite imagery. Satellite imagery datasets have been used for various tasks, such as national defense, agriculture, navigation, gas-leak detection, and disaster surveillance [7]. However, if a classification system is established for these satellite imagery-related datasets, they can be more accessible to industries and applications and widely used.

2.2. **Taxonomy development in information systems.**

Taxonomies make the complex world easier to understand by categorizing complex phenomena. In this study, we applied the taxonomy creation sequence proposed by Nickerson et al. [8] to develop taxonomies in information systems. A taxonomy consists of dimensions and characteristics, where each dimension has at least two characteristics:

$$T = \{D_i, i = 1, \ldots, n | D_i = \{C_{ij}, j = 1, \ldots, k_i, k_i \geq 2\}\}.$$

The sequence for developing a taxonomy is presented in Figure 2. First, a meta-characteristic is defined, reflecting the intent and purpose of the user applying the taxonomy. Meta-characteristics should include the purpose of the taxonomy (i.e., the most fundamental characteristics that identify the purpose). Second, ending conditions are

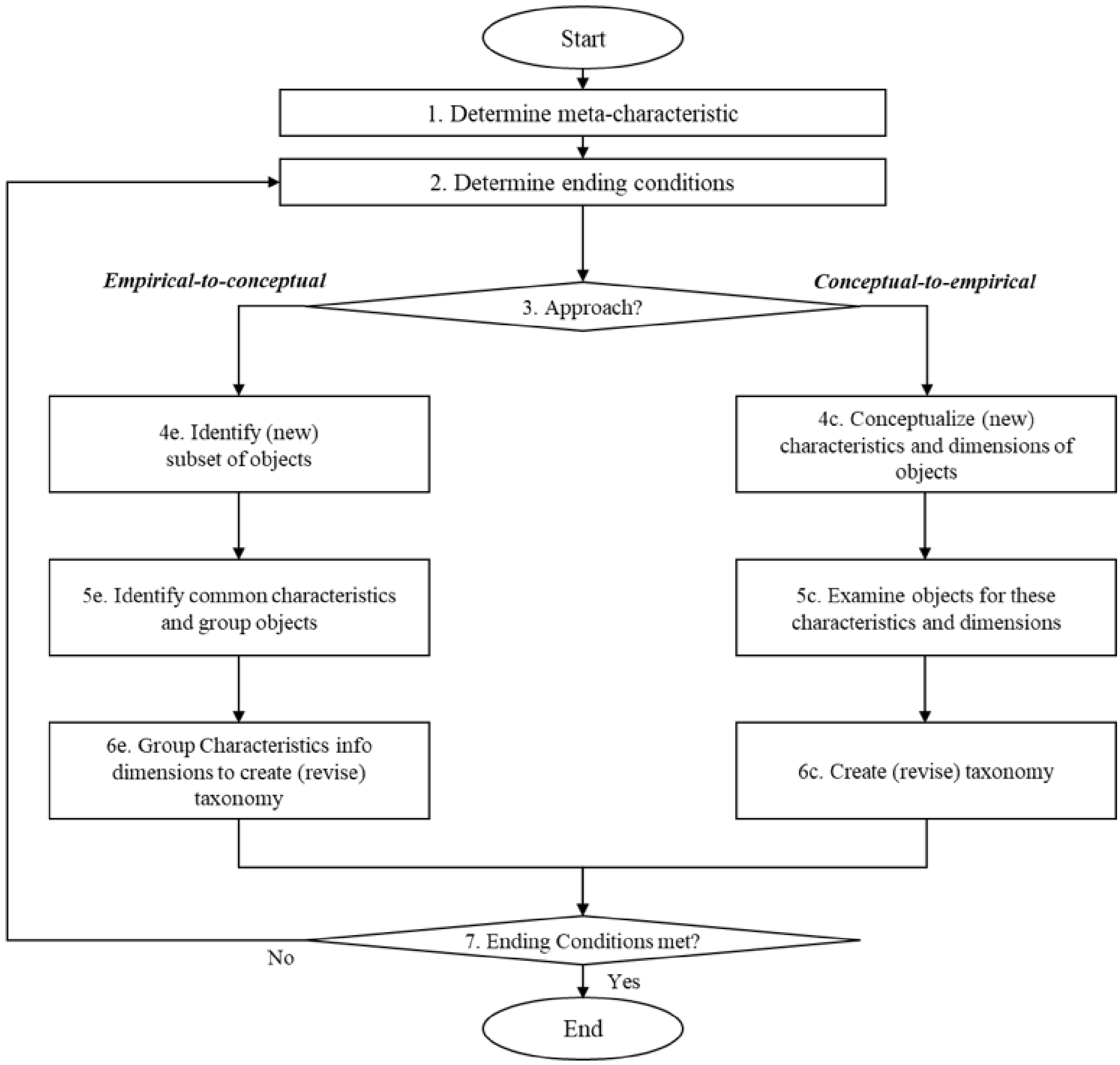

Figure 2. Taxonomy development process [8]

defined and broadly divided into objective and subjective ending conditions, which must be met to finalize the taxonomy creation process. Third, the taxonomy should be iterated to meet the ending conditions, dividing each iteration into empirical-to-conceptual and conceptual-to-empirical methods. Empirical-to-conceptual methods deductively add taxonomical dimensions and characteristics in the following sequence: 1) identify object subsets, 2) identify shared characteristics, then group the objects, and 3) group the characteristics into dimensions to create the taxonomy. In contrast, conceptual-to-empirical methods are inductive, proceeding as follows: 1) conceptualize characteristics and object dimensions, 2) examine objects for these characteristics and dimensions, and 3) create the taxonomy. The selection of an approach depends on the availability of data regarding the objects under study and the researcher's understanding of the domain of interest. Finally, if the ending conditions are satisfied, the iteration is terminated and the taxonomy is completed.

The taxonomy development process proposed by Nickerson et al. [8] has been used in many subfields of information systems. Remane et al. [10] created a taxonomy to categorize carsharing business models. Bräker et al. [11] created a taxonomy to classify augmented reality according to the characteristics of user usage. This study developed a taxonomy to increase its usefulness in business in the rapidly changing augmented reality market and establish a system for categorization [11].

3. Research Methodology.

3.1. **Research framework.** This study creates a taxonomy related to using satellite image datasets so that it can be easily employed. In Figure 3, we synthesize satellite image datasets and previous studies using satellite image datasets to create a taxonomy. The first step was to collect data. We used two types of data in this study: metadata about the dataset and papers written using the dataset. Metadata are used to describe the properties of a dataset and add information about how it can be used in research. In this study, we used 214 satellite imagery datasets and collected metadata and data descriptions related to them. Then, we collected research papers related to the datasets. In total, 257 research papers were used in this process.

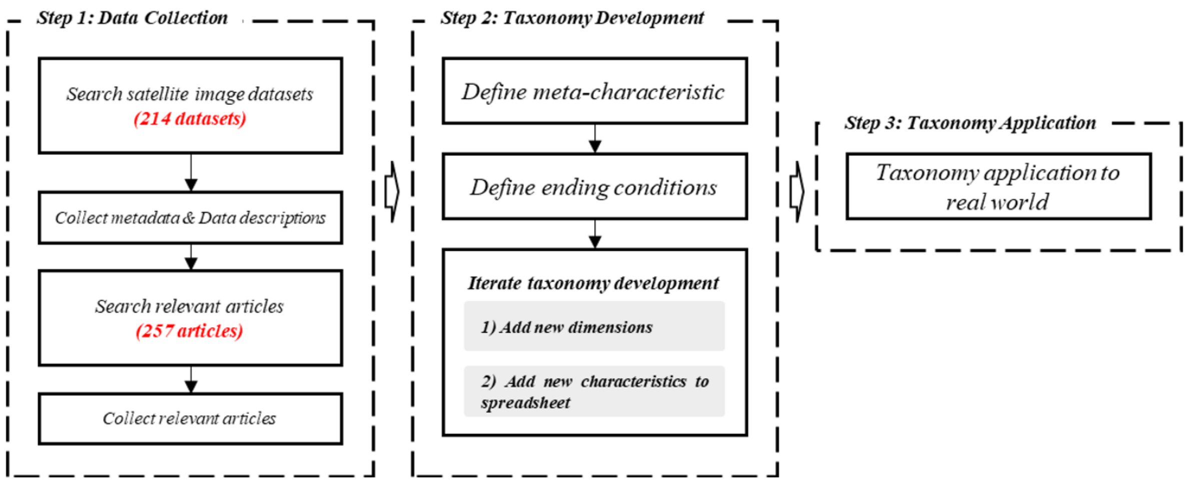

FIGURE 3. Satellite imagery dataset taxonomy development process

The second step is taxonomy development. To create the taxonomy, we first defined meta-characteristics. Next, we defined the ending conditions, and finally, we repeated the taxonomy development until the ending conditions were satisfied. We added dimensions and the determined characteristics to the spreadsheet during the iteration process. The characteristics included in the dimensions should be able to cover all satellite image datasets, so we added the characteristics of dimensions found in the 214 datasets and 257 research papers. The characteristics were organized and added to the spreadsheet as the dimensions were added. Then, after the ending conditions were met, we finalized the taxonomy creation process.

3.2. **Data collection.** This study collected metadata on satellite imagery datasets and data on research papers employing satellite imagery data. First, we identified all available satellite imagery datasets. Then, we organized the dataset types available on online platforms, such as Google dataset search, TensorFlow, GitHub, and Kaggle. Then, we collected all metadata from these datasets, identifying the dimensions and characteristics of dataset feature information using the metadata. For example, we identified and organized information on the dataset, such as the existence of labels and wavelength types. Then, we collected research papers that used the datasets and related research papers to add more information about the dataset features and how and where the data can be applied.

3.3. **Taxonomy development.** After the data collection was completed, the taxonomy creation process was based on the taxonomy development framework by Nickerson et al. [8] (Figure 4). The first step was to define meta-characteristics. This study aimed to create a satellite imagery dataset taxonomy, including dataset characteristics, information for model development, and fields that can be used. Therefore, the meta-characteristics

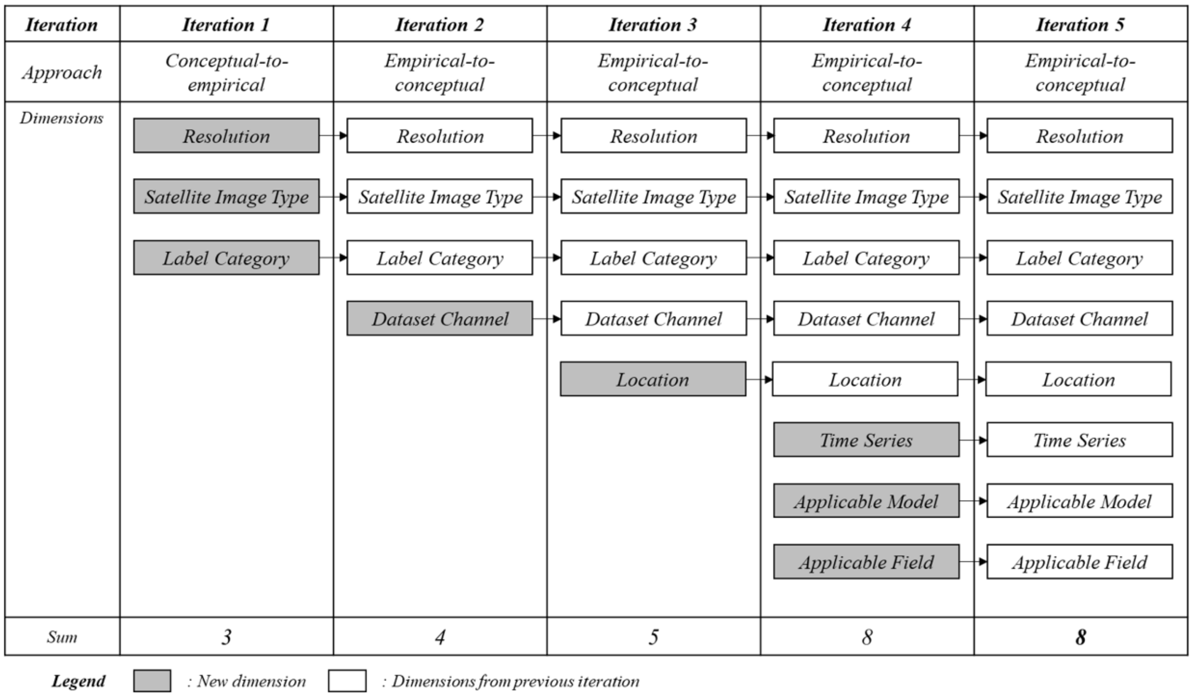

FIGURE 4. Iteration of taxonomy development

of this study were for the classification of characteristics according to the method and purpose of using satellite imagery datasets.

Next, we used the subjective and objective ending conditions proposed by Nickerson et al. [8]. We applied the following commonly used subjective endings: concise, robust, comprehensive, extendible, and explanatory [8]. Similar to various previous studies using the taxonomy development method by Nickerson et al. [8], we adapted the following objective ending conditions to suit this study [12]:

A) At least one object (satellite image dataset feature) is classified under every characteristic of every dimension;

B) No new dimensions or characteristics are added in the last iteration;

C) No dimensions or characteristics are merged or split in the last iteration;

D) Every dimension and characteristic is unique and not repeated;

E) Every known object (satellite image dataset feature) is classified in the taxonomy.

Iterative development was repeated until five subjective and five objective ending conditions were satisfied. Five iterations were conducted in this study. The first iteration was conducted using the conceptual-to-empirical method. According to the meta-characteristics in this study, we prioritized and added information to employ satellite image datasets. We added the resolution and satellite image type to reflect the image characteristics of satellite image datasets. In addition, the label category dimension was added because the dataset can be applied differently, depending on the label characteristics. Thus, three dimensions were derived through Iteration 1.

Iteration 2 was conducted through the empirical-to-conceptual method, collecting metadata from satellite image datasets. We require information on how to collect the dataset to employ the dataset, so we added the dataset channel dimension to access the dataset.

Iteration 3 added one more dimension derived from the attribute information in the metadata of the satellite imagery dataset. The dataset metadata had information about the region; some datasets provided a single region, and satellite imagery datasets provided information about multiple regions. Based on this information, the location dimension was added.

Iteration 4 was conducted through the empirical-to-conceptual method, analyzing research papers describing satellite imagery datasets. In this iteration, three dimensions were added. The time series, applicable model, and applicable field were added to the dimensions of the taxonomy. Among the studies introducing satellite imagery datasets, it was essential to introduce whether time series information is available.

Next, the studies described applicable models and algorithms using the dataset. For example, object detection methods can be used when bounding box labels are included. This study included this information in the taxonomy as an applicable model for introducing satellite image datasets. Next, we added a dimension for the relevant field. In the study introducing satellite imagery datasets, descriptions of relevant fields were added depending on what fields they can be applied in or what labels exist. To include this information in the taxonomy, we used industry classification codes. We applied the industry classification system to include various applicable fields with clear criteria.

In the last iteration, Iteration 5, the taxonomy was created to include research using satellite imagery datasets. In this iteration, we focused on relevant fields or models, which were included in the dimensions in the previous iterations. No additional dimensions were included after reviewing the studies using satellite imagery datasets. Therefore, the ending conditions for this study were met, and the taxonomy development was terminated in Iteration 5.

As new dimensions were added to this study, characteristics were generated for the satellite image dataset and organized and added to the spreadsheet. Figure 5 presents the characteristics of the dimensions added to the spreadsheet for each dataset. All new characteristics were added and organized as the dimensions were added. For example, if a dimension called "label category" was added, we summarized what was found in all datasets and added the information as a characteristic. Thus, from 214 datasets, we derived a spreadsheet with characteristics for all taxonomy dimensions from this study.

No.	Dataset_Name	Article_Name	Artide_yea	Dataset_url	Article_url	Resolutio	Dataset Siz	Image Type	gory(no label, bou	, Business Data, Sensor D	Model	Channel(API, P	Satellite(Landsat, ...	Application Category
171	38-Cloud	Cloud-Net: An End	2019	https://eod-grss-	https://ieeexplore.ieee.org/abstract/docume			Multispectral			Object detection			
172	95-Cloud	Cloud-Net+: A Clo	2020	https://eod-grss-	https://www.arxiv-vanity.com/papers/2001.0			Multispectral			Object detection			
34	95-Cloud: A Cloud Segn	Semantic segmenta	2022	https://github.co	sciencedirect.co	High	34701	SAR	No Label	Government Data	Object Detection	API		Enviornment Manageme
92	AERIAL PERUVIAN ANDE	Development and A	2020	https://eod-grss-	https://core.ac.u	-	-	Hyperspectral	Semantic Label	Public Data	Classification	Platform	Aerial	Remote Sensing
143	Aerial Peruvian Andes C	Development and A	2020	https://eod-grss-	https://core.ac.u	-	-	Hyperspectral	Semantic Label	Public Data	Classification	Platform	Aerial	Remote Sensing
170	Aerial Seabirds West Afr	21 000 birds in 4.5	2021	https://eod-grss-	https://zslpublications.onlinelibrary.wiley.com			Optical			Object detection			
18	AFO - Aerial dataset of	An ensemble deep	2021	https://www.kagg	https://content.i	Low	3647	Optical	Bounding Box	Individual Data	Object Detection	Platform		Monitoring
43	Agricultural Crop Cover	Deep learning tech	2022	https://crowdana	https://link.springer.com/artid		-	SAR	Bounding Box	Public Data	Object Detection	Platform		Agriculture

FIGURE 5. Example of satellite imagery dataset taxonomy spreadsheet

4. **Taxonomy of the Satellite Imagery Dataset.** The finalized taxonomy of the satellite image dataset is provided in Table 1.

D_1 Resolution: The first dimension is the resolution of the satellite image dataset, which has the characteristics of high (finer pixels, ≤ 2 m), medium (medium pixels, 2-30 m), and low (coarser pixels, > 30 m) resolution [13,14].

D_2 Satellite Image Type: The second dimension is the satellite image type, with five types in total. An optical satellite image is typically an image of the Earth's surface using visible and infrared wavelengths [15,16]. Synthetic aperture radar is an image of the Earth's surface using radar signals and can be acquired regardless of weather or time of day [17,18]. Light detection and ranging (LiDAR) is a system that measures the coordinates of a reflector by emitting a laser and measuring the time it takes to bounce back [19], which creates high-resolution three-dimensional (3D) maps and more [20]. Hyperspectral satellite imagery is a remote-sensing technique that measures many continuous visible and infrared spectrum bands. The spectra obtained from each pixel can be analyzed to determine what comprises the pixel [21]. Multispectral satellite imagery uses light in the visible and other wavelength ranges, such as infrared, near-infrared, or ultraviolet [22]. These images typically capture each band separately and combine them to create the final image [23].

TABLE 1. Taxonomy of satellite image dataset

Dimension	Characteristics				
D_1: ***Resolution***	$C_{1,1}$: ***Low (Coarser Pixel)***	$C_{1,2}$: ***Medium (Medium Pixel)***	$C_{1,3}$: ***High (Finer Pixel)***		
D_2: ***Satellite Image Type***	$C_{2,1}$: ***Optical***	$C_{2,2}$: ***SAR (Synthetic aperture radar)***	$C_{2,3}$: ***LiDAR (Light detection and ranging)***	$C_{2,4}$: ***Hyperspectral***	$C_{2,5}$: ***Multispectral***
D_3: ***Label Category***	$C_{3,1}$: ***Bounding box***	$C_{3,2}$: ***Polygonal segmentation***	$C_{3,3}$: ***Semantic label segmentation***	$C_{3,4}$: ***No label***	
D_4: ***Dataset Channel***	$C_{4,1}$: ***API***	$C_{4,2}$: ***Platform***	$C_{4,3}$: ***Home page***	$C_{4,4}$: ***Plug-in***	
D_5: ***Location***	$C_{5,1}$: ***One location***	$C_{5,2}$: ***Multiple locations***			
D_6: ***Time Series***	$C_{6,1}$: ***One period***	$C_{6,2}$: ***Multi period***			
D_7: ***Applicable Deep Learning Model***	$C_{7,1}$: ***Classification***	$C_{7,2}$: ***Object detection***	$C_{7,3}$: ***Semantic segmentation***	$C_{7,4}$: ***Instance segmentation***	$C_{7,5}$: ***GAN (Generative adversarial network)***
D_8: ***Applicable Field***	$C_{8,1}$: ***Agriculture, forestry, fishing and hunting*** $C_{8,2}$: ***Mining, quarrying, and oil and gas extraction*** $C_{8,3}$: ***Utilities*** $C_{8,4}$: ***Construction*** $C_{8,5}$: ***Manufacturing*** $C_{8,6}$: ***Wholesale trade*** $C_{8,7}$: ***Retail trade*** $C_{8,8}$: ***Transportation and warehousing*** $C_{8,9}$: ***Information*** $C_{8,10}$: ***Finance and insurance*** $C_{8,11}$: ***Real estate and rental and leasing*** $C_{8,12}$: ***Professional, scientific, and technical services*** $C_{8,13}$: ***Management of companies and enterprises*** $C_{8,14}$: ***Administrative and support and waste management and remediation services*** $C_{8,15}$: ***Educational services*** $C_{8,16}$: ***Health care and social assistance*** $C_{8,17}$: ***Arts, entertainment, and recreation*** $C_{8,18}$: ***Accommodation and food services*** $C_{8,19}$: ***Other services (except public administration)*** $C_{8,20}$: ***Public administration***					

D_3 Label Category: The label category's third dimension identifies four types of deep learning training for satellite image datasets. These include the bounding box for object detection, polygonal segmentation, semantic label segmentation, and one dataset with a class and no annotation label [28].

D_4 Dataset Channel: This dimension is the channel to collect datasets. To employ the satellite image dataset, we must know from where to collect it. Application programming interfaces (APIs) provide satellite imagery datasets. Platforms, such as Kaggle, provide collected datasets. Home pages provide datasets and make them available for download. Plug-ins offer satellite image datasets in various software, including geographical information system (GIS) software.

D_5 Location: The fifth dimension is the location of the satellite image dataset, which has two characteristics. The satellite image dataset consists of one location, providing only one region, and multiple locations, providing satellite images of various regions.

D_6 Time Series: The sixth dimension is the time series with two characteristics, those that only exist at one point in time and those over various time points.

D_7 Applicable Deep Learning Model: The seventh dimension is for the applicable deep learning model. The variety of satellite imagery datasets available today can be used for various deep learning techniques. Classification, object detection, semantic segmentation, instance segmentation, and generative adversarial network algorithms have been used in studies using satellite imagery datasets [24,25].

D_8 Applicable Field: The last dimension is the applicable field, employing the North American Industry Classification System (NAICS) [26]. The relevant field varies depending on the region where the satellite image dataset is used and what objects are labeled; thus, this study added the industry code as a dimension that can be classified.

5. **Taxonomy Application.** This study details a method of selecting and applying taxonomy to the PASTIS (Panoptic Agricultural Satellite Time Series) dataset [27]. Figure 6 illustrates this application on a multispectral satellite image with 10 m/pixel in the medium pixel category. And since the dataset contains semantic segmentation labeling,

Dimension	Characteristics				
D_1: ***Resolution***	$C_{1,1}$: ***Low (Coarser Pixel)***	$C_{1,2}$: ***Medium (Medium Pixel)*** ★	$C_{1,3}$: ***High (Finer Pixel)***		
D_2: ***Satellite Image Type***	$C_{2,1}$: ***Optical***	$C_{2,2}$: ***SAR (Synthetic aperture radar)***	$C_{2,3}$: ***LiDAR (Light detection and ranging)***	$C_{2,4}$: ***Hyperspectral***	$C_{2,5}$: ★ ***Multispectral***
D_3: ***Label Category***	$C_{3,1}$: ***Bounding box***	$C_{3,2}$: ***Polygonal segmentation***	$C_{3,3}$: ***Semantic label segmentation*** ★	$C_{3,4}$: ***No label***	
D_4: ***Dataset Channel***	$C_{4,1}$: ***API***	$C_{4,2}$: ***Platform***	$C_{4,3}$: ***Home page*** ★	$C_{4,4}$: ***Plug-in***	
D_5: ***Location***	$C_{5,1}$: ***One location***	$C_{5,2}$: ***Multiple locations*** ★			
D_6: ***Time Series***	$C_{6,1}$: ***One period***	$C_{6,2}$: ***Multi period*** ★			
D_7: ***Applicable Deep Learning Model***	$C_{7,1}$: ***Classification***	$C_{7,2}$: ***Object detection***	$C_{7,3}$: ***Semantic segmentation*** ★	$C_{7,4}$: ***Instance segmentation***	$C_{7,5}$: ***GAN (Generative adversarial network)***
D_8: ***Applicable Field***	$C_{8,1}$: ***Agriculture, forestry, fishing and hunting*** ★ $C_{8,2}$: ***Mining, quarrying, and oil and gas extraction*** $C_{8,3}$: ***Utilities*** $C_{8,4}$: ***Construction*** $C_{8,5}$: ***Manufacturing*** $C_{8,6}$: ***Wholesale trade*** $C_{8,7}$: ***Retail trade*** $C_{8,8}$: ***Transportation and warehousing*** $C_{8,9}$: ***Information*** $C_{8,10}$: ***Finance and insurance*** $C_{8,11}$: ***Real estate and rental and leasing*** $C_{8,12}$: ***Professional, scientific, and technical services*** $C_{8,13}$: ***Management of companies and enterprises*** $C_{8,14}$: ***Administrative and support and waste management and remediation services*** $C_{8,15}$: ***Educational services*** $C_{8,16}$: ***Health care and social assistance*** $C_{8,17}$: ***Arts, entertainment, and recreation*** $C_{8,18}$: ***Accommodation and food services*** $C_{8,19}$: ***Other services (except public administration)*** $C_{8,20}$: ***Public administration***				

FIGURE 6. Taxonomy application

we organized it into multiple periods for several locations. This labeling is also why the dataset has been heavily used in semantic segmentation research. In addition, this agriculture labeling makes this valuable dataset in fields such as agriculture, forestry, fishing, and hunting.

With this dataset, Garnot et al. [27] proposed a deep learning technique that can be used for a broader range of tasks in crop mapping, which can be used in various fields such as crop production management, harvest prediction, and damage detection in agricultural fields. If a user aims to use this dataset in the real world, it would be helpful to know what labels are present and what deep learning techniques can be used and to have information about the applicable fields. We created a system to classify satellite imagery datasets according to the characteristics of the datasets and the fields in which they can be used to enhance their usefulness. In the future, the satellite image dataset taxonomy developed in this study is expected to be helpful in various fields by systematically selecting appropriate datasets.

6. **Conclusion.** This study analyzed the metadata of satellite image datasets and studies that used them to create a satellite image dataset taxonomy to classify the types of satellite image datasets and studies, which are increasing at an unprecedented rate. Thus, the types of satellite image datasets currently available can be categorized, making it easier to apply them to the real world.

Next, there are two main contributions of this study. First, this study contributes to management. Most existing studies using satellite image datasets have been conducted in deep learning model development or remote sensing. However, there has not been much research or commercialization related to business models that apply them in the real world. Therefore, if this study can manage various satellite image datasets in terms of usability, it creates business value by using them in various fields. Second, this study has practical contributions. We collected metadata and papers using the datasets to add valuable information on using satellite imagery datasets. Through the metadata, various characteristics of the dataset were reflected in the taxonomy. We summarized the basic characteristics of satellite image datasets and how they can be used and applied. Therefore, the satellite image dataset taxonomy derived from this study can be used to effectively manage vast satellite image datasets in terms of real-world usability and application. There is a practical contribution in effectively managing the information of such diverse data.

In future research, experts should validate, evaluate, and supplement the taxonomy to compensate for these limitations. However, this study was conducted qualitatively to create the taxonomy; thus, the dimensions were arbitrarily added. Although this study followed the procedure of taxonomy development, the arbitrary addition of dimensions must be supplemented. Therefore, future studies should use interviews with experts and the Delphi method to supplement this process to validate the taxonomy. For example, "Was the selection of dimensions appropriate"? and "Were enough characteristics added"?

Acknowledgment. This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2017-0-01637) supervised by the IITP (Institute for Information & Communications Technology Promotion).

REFERENCES

[1] S. H. Kim, J. W. Chae and J. Y. Kang, Research trends and datasets review using satellite image, *Smart Media Journal*, vol.11, no.1, pp.17-30, 2022.
[2] BryceTech, *Smallsats by the Numbers 2022*, 2023.
[3] A. Witze, 2022 was a record year for space launches, *Nature*, 2023.
[4] L. Ma et al., Deep learning in remote sensing applications: A meta-analysis and review, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.152, pp.166-177, 2019.

[5] J. Do, S. Ahn and J. Kang, Urbanization effect of mega sporting events using Sentinel-2 satellite images: The case of the Pyeongchang Olympics, *Sustainable Cities and Society*, vol.74, 103158, 2021.

[6] M. M. H. Khan et al., Natural disasters and land-use/land-cover change in the southwest coastal areas of Bangladesh, *Regional Environmental Change*, vol.15, pp.241-250, 2015.

[7] A. Asokan and J. Anitha, Change detection techniques for remote sensing applications: A survey, *Earth Science Informatics*, vol.12, pp.143-160, 2019.

[8] R. C. Nickerson, U. Varshney and J. Muntermann, A method for taxonomy development and its application in information systems, *European Journal of Information Systems*, vol.22, no.3, pp.336-359, 2013.

[9] C. Pawlowski and H. Scholta, A taxonomy for proactive public services, *Government Information Quarterly*, vol.40, no.1, 101780, 2023.

[10] G. Remane et al., A taxonomy of carsharing business models, *ICIS*, 2016.

[11] J. Bräker, J. Hertel and M. Semmann, Conceptualizing interactions of augmented reality solutions, *HICSS*, 2022.

[12] D. Rau et al., Pushing the frontiers of service research – A taxonomy of proactive services, *International Conference on Information Systems, ICIS 2020 – Making Digital Inclusive: Blending the Local and the Global*, 2021.

[13] S. Abburu and S. B. Golla, Satellite image classification methods and techniques: A review, *International Journal of Computer Applications*, vol.119, no.8, 2015.

[14] A. N. Soni, Spatical context based satellite image classification-review, *International Journal of Scientific Research and Engineering Development*, vol.2, no.6, pp.861-868, 2019.

[15] Z. Li et al., Cloud and cloud shadow detection for optical satellite imagery: Features, algorithms, validation, and prospects, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.188, pp.89-108, 2022.

[16] T. Heid and A. Kääb, Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery, *Remote Sensing of Environment*, vol.118, pp.339-355, 2012.

[17] C. Henry, S. M. Azimi and N. Merkle, Road segmentation in SAR satellite images with deep fully convolutional neural networks, *IEEE Geoscience and Remote Sensing Letters*, vol.15, no.12, pp.1867-1871, 2018.

[18] B. Fan et al., Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT, *IEEE Geoscience and Remote Sensing Letters*, vol.10, no.4, pp.657-661, 2012.

[19] Q. Ma, Y. Su and Q. Guo, Comparison of canopy cover estimations from airborne LiDAR, aerial imagery, and satellite imagery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.10, no.9, pp.4225-4236, 2017.

[20] J. Castagno and E. Atkins, Roof shape classification from LiDAR and satellite image data fusion using supervised learning, *Sensors*, vol.18, no.11, 3960, 2018.

[21] P. Arellano et al., Detecting the effects of hydrocarbon pollution in the Amazon forest using hyperspectral satellite images, *Environmental Pollution*, vol.205, pp.225-239, 2015.

[22] E. McAllister et al., Multispectral satellite imagery and machine learning for the extraction of shoreline indicators, *Coastal Engineering*, 104102, 2022.

[23] K. Yuan et al., Deep-learning-based multispectral satellite image segmentation for water body detection, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.14, pp.7422-7434, 2021.

[24] Y. Gu, Y. Wang and Y. Li, A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection, *Applied Sciences*, vol.9, no.10, 2019.

[25] Y. Li et al., Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.146, pp.182-196, 2018.

[26] C. O. Kile and M. E. Phillips, Using industry classification codes to sample high-technology firms: Analysis and recommendations, *Journal of Accounting, Auditing & Finance*, vol.24, no.1, pp.35-58, 2009.

[27] V. S. F. Garnot, L. Landrieu and N. Chehata, Multi-modal temporal attention models for crop mapping from satellite time series, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.187, pp.294-305, 2022.

[28] A. A. S. Gunawan, Fanny and E. Irwansyah, Development of paddy field mapping from satellite image using semantic segmentation method in Central Borneo, *ICIC Express Letters*, vol.15, no.9, pp.941-949, DOI: 10.24507/icicel.15.09.941, 2021.