Geoid
Journal of Geodesy and Geomatics

Geoid Vol. 21, No. 1, 2026, 64-78

P-ISSN: 1858-2281; E-ISSN: 2442-3998

Evaluation of Google Earth Engine Embedding Dataset for Remote Sensing Image Classification

Calvin Wijaya* & Harintaka
Department of Geodetic Engineering, Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia
*Corresponding author: calvin.wijaya@ugm.ac.id

Received: 28 August 2025; Revised: 18 January 2026; Accepted: 19 January 2026; Published: 20 January 2026

Abstract: Google Earth Engine (GEE) has emerged as one of the most powerful cloud-based platforms for processing and analyzing remote sensing imagery. By integrating vast Earth observation archives with scalable computational resources, it provides an accessible environment for researchers, practitioners, and decision-makers. In 2025, Google's AlphaEarth Foundation introduced a novel embedding model trained on diverse Earth observation datasets available on the GEE server. This model, generated from annual time-series imagery and offered in an analysis-ready format, enables general-purpose applications such as classification, clustering, regression and change detection. Despite its potential, the performance and capabilities of this embedding model remain underexplored. This study evaluates the effectiveness of the embedding datasets in GEE for supervised classification method. Comparative experiments were conducted against widely used remote sensing imagery, including Sentinel-2 and Landsat 9 imagery, using multiple algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Trees (CART), and Object-Based Image Analysis (OBIA). In addition, a case study was carried out to examine the use of embedding datasets for mangrove classification. Validation using overall accuracy demonstrates that embedding datasets achieve superior results compared to conventional imagery. Classification using the embedding dataset achieved an average overall accuracy of 94%, outperforming Landsat 9 (83.1%) and Sentinel-2 (82.5%). Moreover, the embedding dataset produced a classification pattern similar to OBIA, even without the need for image segmentation. The findings highlight the potential of embedding datasets to enhance classification accuracy and broaden the scope of remote sensing applications, suggesting new opportunities for leveraging advanced machine learning representations in geospatial analysis.

Keywords : Google Earth Engine; Image Classification; Embedding Dataset; GEE; Life on land

How to cite: Wijaya, Calvin., Harintaka. (2026). Evaluation of Google Earth Engine Embedding Dataset for Remote Sensing Image Classification. *Geoid, 21*(1), 64-78.

Introduction

Google Earth Engine (GEE) (Gorelick et al., 2017) has revolutionized the way remote sensing processes are conducted. It has transformed the traditional time-consuming and inefficient processing of remote sensing imagery into a more effective and scalable workflow. Historically, remote sensing data processing relied on dedicated desktop software installed locally on the client side, with performance highly dependent on the user's hardware specifications. This approach was inefficient given the large size of remote sensing datasets. With the increasing volume of data generated daily from Earth observation satellites, remote sensing now faces a significant "Big Data" challenge (Chi et al., 2016; Dritsas & Trigka, 2025; Ma et al., 2015; Mahdianpari et al., 2019). Monitoring applications or time-series analyses involving multiple datasets have become particularly challenging under such conditions. This paradigm, however, has shifted with the emergence of GEE.

GEE is a cloud-based geospatial processing platform that enables large-scale analysis using high-performance computing resources accessible via the web (Gorelick et al., 2017). It hosts petabytes of remote sensing data and provides free access for educational and research purposes (Kumar & Mutanga, 2018; Parente et al., 2019). Its datasets encompass information from multiple sources, including Landsat, Sentinel, MODIS, climate models, temperature measurements, and geophysical parameters (Gorelick et al., 2017; Padarian et al., 2015). Leveraging these capabilities, GEE has facilitated numerous innovations and research advancements (Ahmad et al., 2020; Fattore et al., 2021; Hird et al., 2017). Velastegui-Montoya et al. (2023) report that over 2,800

articles related to GEE were published between 2011 and 2022, demonstrating its rapid adoption and growing influence in the field.

Innovation within GEE continues to evolve. In 2025, Google's AlphaEarth Foundation (AEF) introduced a novel embedding model designed to integrate diverse geospatial datasets into a unified, time-continuous embedding space, creating a rich and comprehensive information representation (Brown et al., 2025). This embedding model supports a wide range of applications, including supervised and unsupervised classification, regression, change detection, and similarity search. The model processes multi-sensor time series and condenses the information into a 64-dimensional embedding vector at 10-meter resolution (Brown et al., 2025), thus enabling efficient representation of complex datasets.

The underlying concept of the AEF embedding model is comparable to Principal Component Analysis (PCA) (Shlens, 2003), a linear transformation technique commonly used for dimensionality reduction (Li et al., 2022; Salem & Hussein, 2019). PCA is often applied to remote sensing data, such as hyperspectral imagery, to reduce the number of bands while retaining key information. However, unlike PCA, which is linear, the AEF model employs deep learning to reduce thousands of multi-sensor time-series input dimensions to a 64-band representation (Brown et al., 2025). This approach produces a dataset with highly rich semantic information, making it suitable for diverse geospatial applications. The first version of the embedding model is available within GEE, providing researchers with new flexibility and opportunities.

Regarding embedding, most applications have been in the domain of natural language processing (NLP). Embeddings capture multiple dimensions of linguistic information, such as words and sentences, and represent them as numerical vectors in a continuous space. This dimensionality reduction enables more efficient processing in deep learning models for NLP tasks. Research by Harris et al. (2024) improved embedding performance in Large Language Models (LLMs) for tasks such as text enrichment and rewriting. Similarly, Ba Alawi & Bozkurt (2025) investigated the effectiveness of embedding techniques for Turkish text analysis. Both studies support the notion that embedding techniques enhance performance in NLP-related applications.

Another practical use of embeddings is in few-shot learning models, where labeled datasets are scarce for training deep learning algorithms**.** By reducing the dependency on large, labeled datasets, embeddings facilitate efficient model training. For example, Askari et al. (2025) demonstrated improved image classification accuracy using multi-scale embeddings combined with attention mechanisms. Similarly, Kang et al. (2021) developed relational embedding networks that integrate two relational models for image classification, achieving consistent performance across benchmark datasets. Despite these advancements, the application of embedding techniques for land cover classification, particularly within GEE, remains underexplored due to the recent introduction of the AlphaEarth embedding dataset.

Given the novelty of the AEF embedding dataset, its potential applications remain largely unexplored in the context of remote sensing and geospatial analysis. While embedding techniques have shown considerable success in other domains such as NLP and computer vision, there is currently a lack of systematic evaluation of embedding-based geospatial dataset for land cover classification within large-scale geospatial platforms such as GEE. To address this research gap, this study conducts a comprehensive evaluation of the AEF embedding dataset for land cover classification. Specifically, we assess its classification accuracy and computational efficiency in comparison to widely used satellite datasets, including Landsat and Sentinel-2, across multiple machine learning algorithms. The primary objective of this research is to determine whether embedding-based representations can offer a viable alternative—or even an advantage—over conventional remote sensing imagery for thematic classification tasks. By assessing the performance of the dataset, a more accurate and representative characterization of the Earth's land cover can be obtained, which supports the Sustainable Development Goals (SDGs), particularly Life on Land.

The main contributions of this study are as follows: (i) to introduce and assess the newly released AEF embedding dataset within the GEE environment for remote sensing applications; (ii) to conduct a comparative analysis of classification performance between embedding data and traditional satellite imagery across multiple classifiers; and (iii) to provide insights into the potential of embedding models for advancing land cover mapping, which could influence future research and operational practices in geospatial science. The remainder of this paper is structured as follows: Section 2 describes the study area, datasets, and methodology employed for data processing and evaluation. Section 3 presents the classification results and provides an in-

depth discussion of findings. Finally, Section 4 summarizes the conclusions and outlines potential directions for future research.

Methodology

Study Area

This research focuses on the Special Region of Yogyakarta Province as the primary location for evaluating the performance of embedding datasets in comparison to Landsat 9 and Sentinel-2 imagery. In addition, Yogyakarta City is included for comparison with Object-Based Image Analysis (OBIA), and a section of the Cilacap area is used as a case study for mangrove classification. Although GEE allows global-scale analysis, these areas were selected to capture diverse land surface characteristics, including urban environments, coastal regions, agricultural lands, and forested areas. While the analysis is demonstrated in these specific locations, the methodology is fully reproducible elsewhere since the Earth observation datasets utilized are globally available and continuously updated. The study area for the Special Region of Yogyakarta covers approximately 317,239.159 ha, Yogyakarta City spans 3,354.706 ha, and the selected region in Cilacap encompasses 103,720.86 ha. Administrative boundaries were used for Yogyakarta Province and Yogyakarta City, while a rectangular bounding box was applied to define the Cilacap study area using GEE tools. The research locations are illustrated in Figure 1.

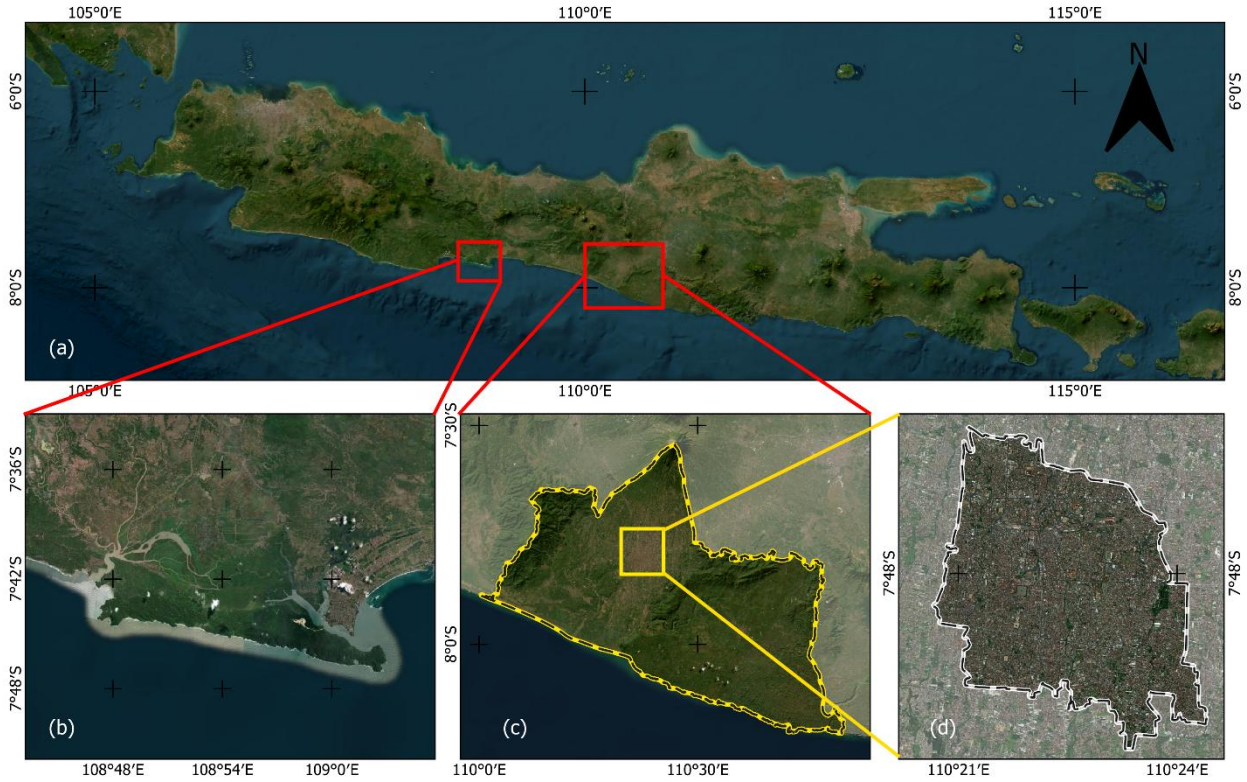


Figure 1. Research areas: (a) Java Island; (b) Cilacap area for mangrove classification case study; (c) Special Region of Yogyakarta Province as the main study area for embedding dataset comparison; and (d) Yogyakarta City for embedding vs. OBIA classification (Source: Author's processing based on ESRI basemap)

Data

The primary dataset for this study is the Google Satellite Embedding Dataset (Version 1), which serves as the basis for analysis and evaluation against traditional satellite imagery. Additional datasets include Landsat 9 and Sentinel-2, which together represent widely used Earth observation imagery for land cover classification. These three datasets form the core data sources for all processes, including the mangrove classification case study and OBIA comparison. All datasets were accessed from the GEE data catalog and are freely available for research purposes. The analysis was conducted for the year 2024, as the embedding dataset provides annual

composites. The embedding dataset offers a spatial resolution of 10 meters and consists of a 64-dimensional embedding vector that encodes temporal trajectories of surface conditions at and around each pixel (Brown et al., 2025). These embeddings are derived from multiple Earth observation instruments and datasets over a single calendar year. For the Landsat 9 collection, we utilized Collection 2 Tier 1 TOA Reflectance, and for Sentinel-2, Level-2A surface reflectance products were used. Landsat 9 imagery has a spatial resolution of 30 meters, whereas Sentinel-2 imagery offers 10-meter resolution. A visualization of the three datasets used in this study for the Special Region of Yogyakarta is provided in Figure 2.

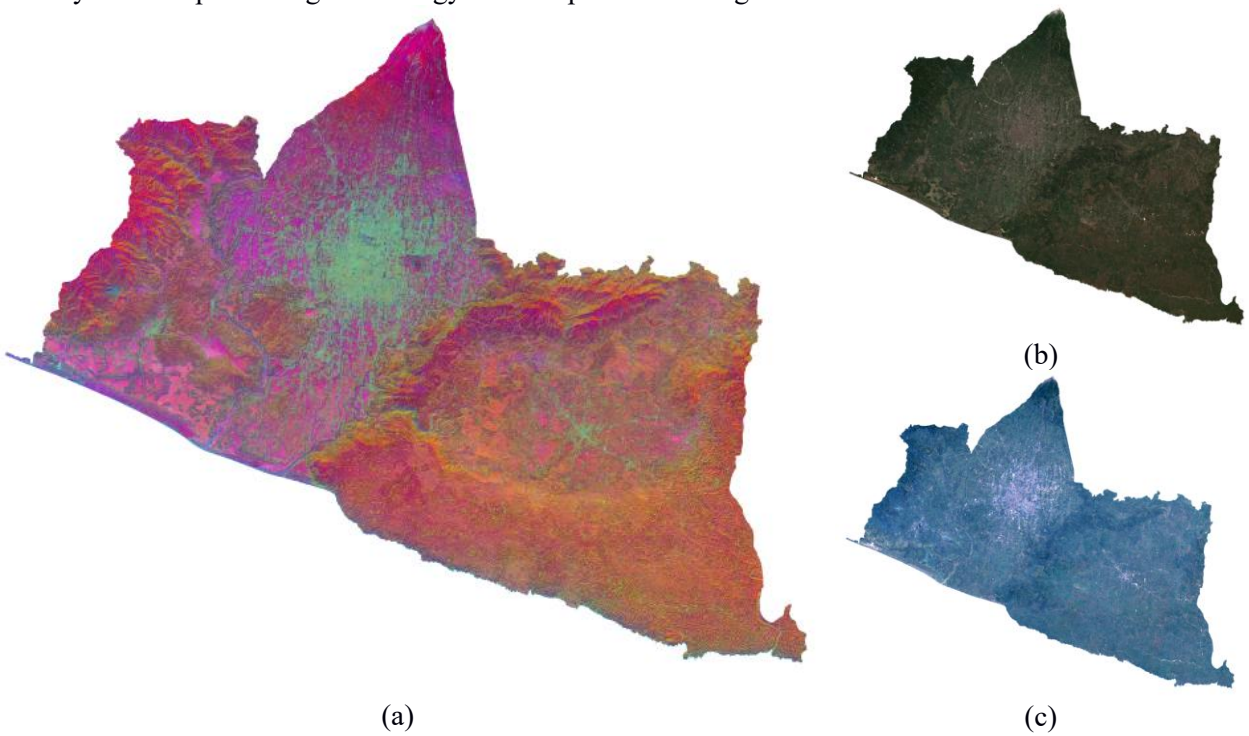


Figure 2. Visualization of datasets in the Special Region of Yogyakarta: (a) Embedding dataset, (b) Sentinel-2, and (c) Landsat 9 (Source: GEE catalogue, processed on GEE by the authors).

The rationale for including Landsat 9 and Sentinel-2 lies in their prominence as the most widely used datasets in remote sensing applications. Both have become foundational sources for Earth observation and land cover analysis (Ez-zahouani et al., 2025). According to Velastegui-Montoya et al. (2023), Landsat and Sentinel imagery have been the most frequently used satellite datasets in research over the past decade. These datasets are freely accessible, offer adequate spatial resolution for land cover mapping, and provide high applicability across various analyses (Adorno et al., 2023). It has easy access and high applicability for wide range of analysis (Yang & Suh, 2023). Furthermore, the Landsat program maintains one of the longest continuous global Earth observation records (Adrian et al., 2021). This study specifically employs Landsat 9, the most recent generation in the Landsat series. Given their common usage and reliability, Landsat 9 and Sentinel-2 serve as appropriate benchmarks to assess the performance of the embedding dataset and evaluate its potential as an alternative data source for land cover classification.

Data Pre-processing

All data pre-processing and analysis were conducted within the GEE Code Editor. The pre-processing step ensures that datasets are prepared for subsequent classification tasks. For the embedding dataset, pre-processing involved filtering the data to the desired year and study area boundaries. This step was straightforward, requiring only temporal and spatial filtering. The filtered embedding data were then merged into an image mosaic for analysis. Similar filtering was applied to Sentinel-2 and Landsat 9 datasets based on date and location. However, Sentinel-2 required additional cloud masking to ensure a clear image. This was achieved using the QA60 band, which contains cloud and cirrus flags. Pixels flagged as cloudy, or cirrus were masked out. For Landsat 9, a simple cloud score algorithm was applied, combining brightness, temperature,

and the Normalized Difference Snow Index (NDSI) to identify and mask cloudy pixels. The final cloud-free images for all datasets are illustrated in Figure 2.

Machine Learning Dataset & Algorithm

The pre-processed imagery was then used to generate training and test datasets. A 70:30 ratio was adopted for splitting data into training and test sets, a commonly used approach in machine learning. Sampling was performed using stratified random sampling, ensuring that class representation in the samples reflects the distribution in the imagery. Consequently, the number of samples per class was not uniform, as classes like water bodies occupy smaller areas compared to urban or agricultural land. For instance, water samples were fewer since the region contains only rivers and lakes, with no coastal water within the main study area. The classification scheme for the Special Region of Yogyakarta and Yogyakarta City included five classes: barren land, agriculture, forest, urban, and water. For the Cilacap mangrove study, the classes were simplified to three categories: mangrove, water, and others. The distribution of training and test samples for the Yogyakarta classification is shown in Table 1.

Table 1. Training and test dataset distribution

Class	Training Sample	Test Sample	Ratio Train-Test
Barren Land	50	20	71:29
Farm	100	45	69:31
Forest	75	30	71:29
Urban	100	45	69:31
Water	25	10	71:29
Total	**350**	**150**	**70:30**

In total, 500 samples were used for the Yogyakarta study—350 for training and 150 for testing. Sampling locations were spatially distributed to avoid clustering in one area. The pixel values of each dataset (Embedding, Sentinel-2, and Landsat 9) were extracted for these points using the sampling regions function in GEE, with separate processes for training and testing sets.

Four machine learning algorithms were applied for classification: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Trees (CART). All classifiers were implemented in GEE. For the Random Forest classifier, two configurations were tested: 100 trees and 1,000 trees, to evaluate the effect of tree count on performance. For input features, all image bands were used. This means the embedding dataset utilized all 64 embedding dimensions, while Sentinel-2 included original bands plus two indices: Normalized Difference Vegetation Index (NDVI) and Bare Soil Index (BSI) to enhance discrimination between vegetation and bare land. Landsat 9 was used with its original bands, serving as a baseline multispectral configuration.

This design enables an explicit comparison across datasets with different feature dimensionalities, allowing the analysis to examine whether higher-dimensional embedding representations provide additional discriminative information for land cover classification compared to conventional multispectral imagery. Additional derived indices were not included for Landsat 9 to avoid introducing features that could bias the comparison.

Additionally, the study compared pixel-based classification with Object-Based Image Analysis (OBIA) following the approach in (Tassi & Vizzari, 2020). OBIA was implemented in GEE using the Simple Non-Iterative Clustering (SNIC) segmentation algorithm (Radhakrishna & Susstrunk, 2011), followed by RF classification. Due to its higher memory and computational requirements, OBIA was applied only to Yogyakarta City and not the entire province. The OBIA results were then compared with pixel-based classification using the embedding dataset for the same extent. For the mangrove classification in Cilacap, RF with 100 trees was employed. To validate mangrove classification accuracy, the Global Mangrove Dataset available in GEE (Giri et al., 2011) was used as a reference.

Evaluation

Classification accuracy was assessed using a confusion matrix, which compares predicted classes against reference (ground truth) data. Rather than serving as ground truth, Sentinel-2 imagery was used as a reference benchmark due to its comparable spatial resolution and widespread use in land cover mapping studies. This approach enables a resolution-consistent, apple-to-apple comparison between the AEF embedding dataset and conventional multispectral imagery.

It is important to note that Sentinel-2–based classifications are themselves subject to uncertainties inherent in remotely sensed data and therefore do not represent absolute ground truth. Consequently, the evaluation focuses on relative classification agreement and comparative performance, rather than definitive land cover accuracy. From the confusion matrix, user's accuracy, producer's accuracy, omission error, and commission error were computed for each class. Overall classification performance was evaluated using Overall Accuracy (OA) and the Kappa coefficient, calculated from the confusion matrix using the following general formula:

$$Accuracy \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

The TP represents the number of pixels correctly classified as belonging to a given class, while TN denote pixels correctly identified as not belonging to that class. Conversely, FP corresponds to pixels incorrectly assigned to a class despite belonging to another category (commission error), and FN are pixels that truly belong to a class but were misclassified as another (omission error).

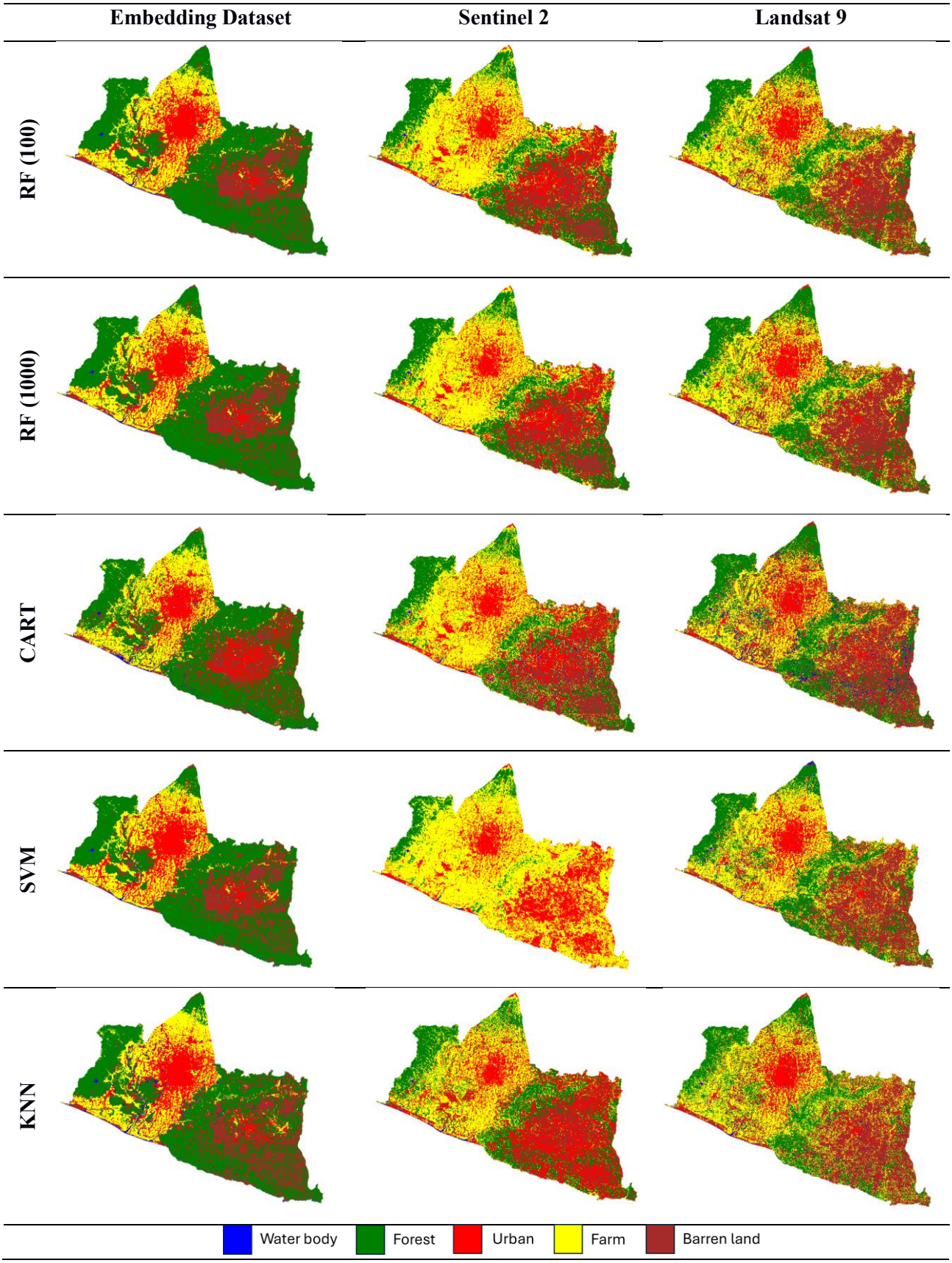
Results and Discussion

Embedding Dataset Classification Result

The classification results for the embedding dataset, Sentinel-2, and Landsat 9 using various algorithms are presented in Table 2. All classifications were performed using the same training data described in Table 1. The overall classification patterns vary significantly across datasets. For the embedding dataset, the spatial distribution appears consistently across different algorithms. In general, the eastern part of the Special Region of Yogyakarta is dominated by forest, while urban and agricultural areas are concentrated in the central region and extend toward the south. Differences between algorithms for the embedding dataset are relatively minor, with only localized variations in class interpretation. Despite these small discrepancies, the classification patterns remain largely uniform across algorithms for the embedding dataset.

In contrast, classifications based on Sentinel-2 and Landsat 9 exhibit greater inconsistencies when compared to the embedding dataset. For Sentinel-2, the eastern region is predominantly classified as a mix of urban and barren land, with minimal forest coverage—an incorrect result, as this region is mountainous and heavily forested. Additionally, the central area, extending south and west, is primarily classified as agricultural land, with urban areas concentrated around Yogyakarta City. While some of these results align with actual land cover, others are notably inaccurate. Particularly, the SVM classification for Sentinel-2 exhibits substantial noise, with most regions classified as agricultural land, suggesting that the algorithm failed to capture class-specific spectral characteristics effectively. For Landsat 9, most algorithms produce similar patterns; however, the eastern region is largely classified as barren land, again showing low forest presence. The central region exhibits a high degree of noise, characterized by salt-and-pepper effects between urban and agricultural classes.

Table 2. Comparison of classification results for the Embedding Dataset, Sentinel-2, and Landsat 9 across algorithms (Source: GEE catalogue, processed on GEE by the authors).



The embedding dataset consistently produces cleaner and less noisy classification outputs compared to Sentinel-2 and Landsat 9. The visual appearance of its results is more structured, resembling object-based classification rather than traditional pixel-based methods. In contrast, Sentinel-2 and Landsat 9 classifications exhibit pronounced noise across all algorithms. Furthermore, the two optical datasets differ in their interpretation of the same areas—for instance, the eastern region is classified as urban by Sentinel-2 but as barren land by Landsat 9. Similarly, the delineation of urban areas in Yogyakarta and the distribution of agricultural land appears questionable, particularly in the SVM-based Sentinel-2 classification, where agricultural land dominates almost all areas.

The superior performance of the embedding dataset can be attributed to its input structure: 64 embedding vectors per pixel, derived from multi-source, multi-temporal observations, condensed into a rich, compact feature set. This dimensionality reduction provides a comprehensive representation of surface characteristics, improving classification accuracy. In comparison, Sentinel-2 uses 14 bands (including NDVI and BSI indices added to enhance vegetation and bare-soil discrimination), while Landsat 9 uses 11 spectral bands. Thus, the embedding dataset offers 4.5 times more features than Sentinel-2 and nearly six times more than Landsat 9. Regarding RF classification with different tree counts, the resulting maps were visually similar for all datasets. This suggests that, under the current study conditions—considering the study area size, class distribution, and sample size—the number of trees did not significantly influence classification outcomes. However, these observations are based on visual interpretation. To confirm these findings, quantitative accuracy assessment is presented in the next section.

Performance Analysis

The overall accuracy was computed using the test dataset distributed across the study area, as illustrated in Figure 3. Similar to the training dataset, the test samples were selected strategically based on class area or pixel count to ensure representative coverage. The results show that the embedding dataset consistently outperformed Sentinel-2 and Landsat 9 across all classification algorithms. For all algorithms, the embedding dataset achieved an overall accuracy exceeding 90%. The lowest-performing algorithm, CART, reached 90%, while the highest, SVM, achieved 96%. These results demonstrate that the embedding dataset delivers consistently high accuracy regardless of the classifier used. In contrast, Sentinel-2 exhibited overall accuracy values ranging from 77.3% (SVM) to 87.3% (RF with 1,000 trees), indicating a variability of approximately 10% across classifiers. Similarly, Landsat 9 results varied from 79.3% (CART) to 88.7% (SVM). Interestingly, both the embedding dataset and Landsat 9 achieved their highest accuracy with SVM, while Sentinel-2 obtained its lowest accuracy with the same algorithm.

The performance trends in Figure 3 align with the visual interpretation in Table 2. The embedding dataset results, which appear more object-based in structure, correspond to their superior overall accuracy across all classifiers. On average, the five classification algorithms applied to the embedding dataset yielded an overall accuracy of 94%, validating the consistency and robustness of its classification results. In comparison, Landsat 9 and Sentinel-2 achieved average accuracies of 83.1% and 82.5%, respectively—approximately 11–12% lower than the embedding dataset. The anomaly observed in Sentinel-2 SVM classification, which appeared visually inconsistent in Table 2, is confirmed by its low accuracy score in Figure 3.

Figure 3. Performance comparison chart

Further analysis was conducted to assess the effect of tree quantity in RF classification. For the embedding dataset, varying the number of trees from 100 to 1,000 had no impact, as both configurations achieved 94.7% accuracy. For Sentinel-2, increasing the number of trees provided only a marginal improvement of 0.6%, from 86.7% to 87.3%. Conversely, Landsat 9 experienced a slight decrease of 0.7%, from 82.7% to 82%. These negligible differences indicate that the number of trees does not significantly influence classification performance in this study, a finding consistent with the visual similarity between RF100 and RF1000 results shown in Table 2.

To evaluate the performance of each algorithm for each class, Table 3 is provided to compare each dataset used. In Table 3, Producer Accuracy (PA) and Consumer Accuracy (CA) are presented for each algorithm and each class. The evaluation and performance are assessed at the class level to measure class consistency and to identify whether there is any bias tendency toward a specific class. From the results in Table 3, the embedding dataset shows strong class consistency with low class bias, where most classes achieve high values for both PA and CA, indicating very stable classification performance. However, the barren land class exhibits low PA but high CA across all algorithms, indicating that this class is under-detected.

Table 3. Producer Accuracy (PA) and Consumer Accuracy (CA) comparison for each class in each dataset and each algorithm

Dataset	Class	RF100		RF1000		CART		SVM		KNN	
		PA	CA	PA	CA	PA	CA	PA	CA	PA	CA
Embedding Dataset	Barren Land	0.650	0.929	0.650	0.929	0.650	0.929	0.750	0.938	0.600	1.000
	Farm	1.000	0.938	1.000	0.957	1.000	0.882	1.000	0.957	1.000	0.978
	Forest	1.000	0.968	1.000	0.968	0.833	0.893	1.000	0.968	1.000	0.938
	Urban	0.978	0.936	0.978	0.917	0.933	0.933	0.978	0.957	1.000	0.900
	Water	1.000	1.000	1.000	1.000	1.000	0.833	1.000	1.000	1.000	1.000
Sentinel 2	Barren Land	0.550	0.688	0.550	0.786	0.700	0.700	0.000	0.000	0.500	0.667
	Farm	0.933	0.913	0.911	0.891	0.800	0.837	0.867	0.780	0.778	0.875
	Forest	0.967	0.906	0.967	0.879	0.900	0.794	0.867	0.813	0.900	0.794
	Urban	0.844	0.826	0.889	0.851	0.778	0.833	0.911	0.707	0.844	0.792
	Water	1.000	1.000	1.000	1.000	1.000	0.909	1.000	1.000	1.000	0.769

Landsat 9	Barren Land	0.150	0.429	0.050	0.200	0.550	0.407	0.500	0.714	0.400	0.471
	Farm	0.889	0.755	0.911	0.759	0.867	0.796	0.933	0.808	0.867	0.848
	Forest	1.000	0.968	1.000	0.968	0.933	0.933	0.967	1.000	1.000	1.000
	Urban	0.911	0.837	0.911	0.820	0.711	0.914	0.933	0.955	0.844	0.809
	Water	1.000	1.000	1.000	1.000	0.900	1.000	1.000	0.909	0.900	0.900

For Sentinel-2, both PA and CA values across all algorithms and classes are lower than those of the embedding dataset. Several classes exhibit overprediction tendencies, confusion between forest and farm classes, and strong omission errors. Sentinel-2 also shows strong class bias in the SVM algorithm. In Sentinel-2, the barren land class remains highlighted as a class with strong bias and high instability. A similar trend is also observed in Landsat 9, where overprediction is present and confusion occurs between the farm and forest classes. Landsat 9 exhibits strong omission bias for several classes, particularly the barren land class. Based on Table 3, the embedding dataset demonstrates the most consistent and balanced classification behavior with minimal bias compared to the two other reference datasets.

Comparison with OBIA Analysis

The classification results obtained using the embedding dataset across all classifiers (Table 2) exhibit a visual resemblance to object-based classification outputs. To examine this further, we compared the embedding dataset results directly with an OBIA classification performed in GEE. The comparison is illustrated in Figure 4. For the OBIA workflow, we used Sentinel-2 imagery and the RF100 algorithm to classify Yogyakarta City into three classes: urban, vegetation, and barren land. The pixel-based classification result from Sentinel-2 revealed a common issue—salt-and-pepper noise, with numerous small, misclassified pixels scattered across the area. To address this, the OBIA approach employed the SNIC segmentation algorithm prior to classification, resulting in more accurate and homogeneous class representation.

Remarkably, the embedding dataset produced classification results with an object-based appearance without requiring any segmentation process. As shown in Table 2 and supported by the high accuracy scores in Figure 3, the embedding dataset delivers visually coherent and spatially consistent results. In Figure 4, the difference is clear: the same RF classifier yields a significantly cleaner output when applied to the embedding dataset compared to Sentinel-2. This indicates that the embedding dataset inherently supports object-like classification, offering an advantage over traditional pixel-based approaches. Although some small-scale noise is still present in the embedding dataset classification, it is considerably reduced compared to pixel-based classification, making it much closer in appearance to OBIA outputs. Furthermore, this behavior remains consistent across all classifiers, reinforcing the robustness of the embedding dataset in producing object-like classifications without additional segmentation steps.

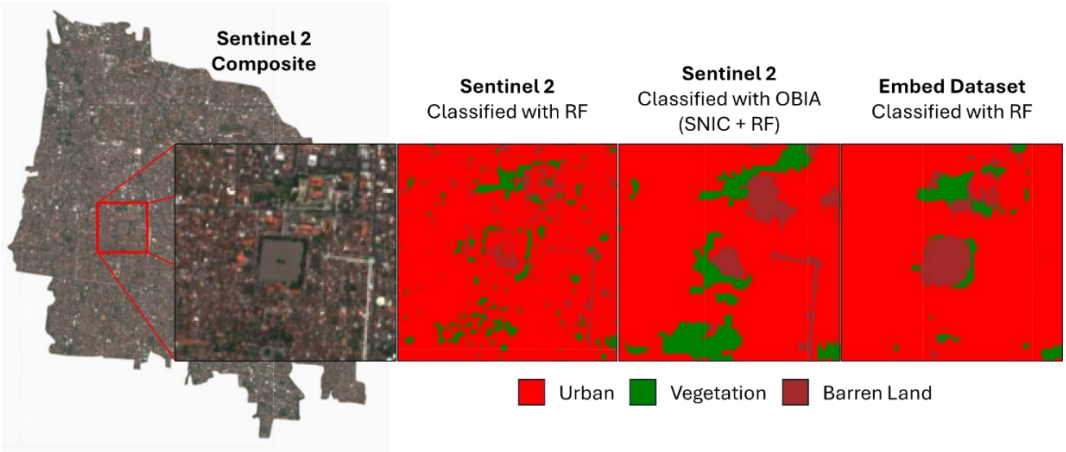


Figure 4. Comparison of embedding dataset classification result with OBIA classification (Source: GEE catalogue, processed on GEE by the authors).

To validate this observation quantitatively, we compared the overall accuracy of OBIA (Sentinel-2 + SNIC segmentation + RF) and the embedding dataset (RF). The OBIA approach achieved an accuracy of 86.7%, while the embedding dataset reached 90%, consistent with earlier findings. This 3.3% improvement underscores the superior performance of the embedding dataset despite not relying on image segmentation. The advantage likely stems from its 64-dimensional embedding vectors, which integrate multi-temporal information, compared to OBIA's reliance on 15 features (12 Sentinel-2 bands, 2 spectral indices—NDVI and BSI, and 1 segmentation layer). To further compare the two methods, we calculated the area of each class in hectares, as shown in Figure 5. Both approaches exhibited similar patterns: urban occupies the largest area, followed by vegetation and then barren land. Using the embedding dataset, the estimated areas are approximately 3,055 ha for urban, 213 ha for vegetation, and 86 ha for barren land. According to the National Statistics Agency of Yogyakarta, the urban area is approximately 3,200 ha, which aligns closely with the embedding dataset estimate compared to other methods. This consistency reinforces the reliability of the embedding dataset for both classification accuracy and spatial representation.

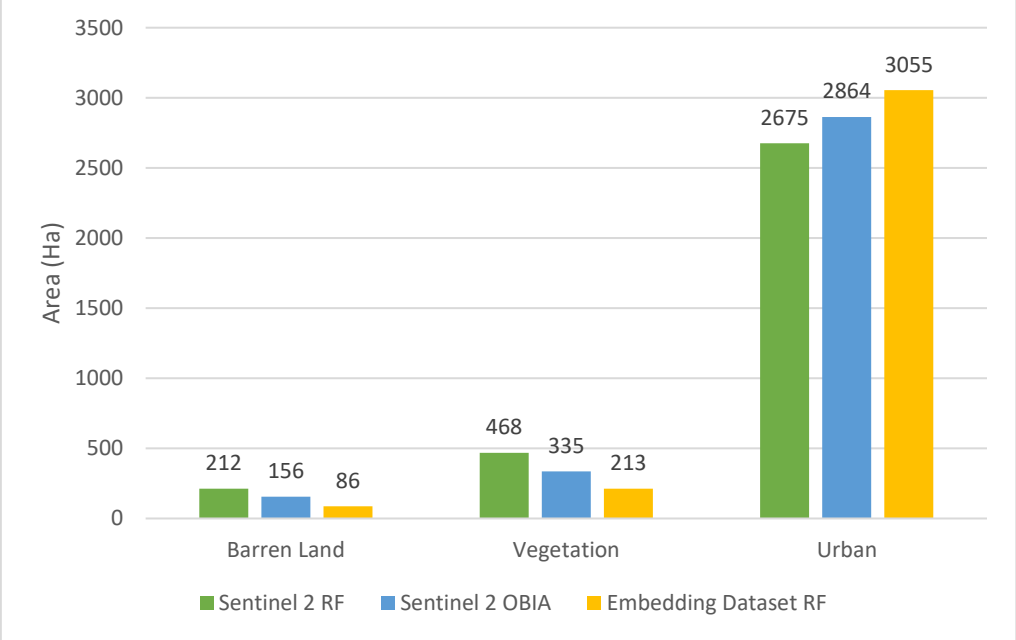


Figure 5. Comparison of class area between OBIA classification and embedding dataset

Case Study: Mangrove Classification

To further analyze and evaluate the performance of the embedding dataset, we conducted an experiment on mangrove area classification in Cilacap. This region contains a significant mangrove ecosystem, making it a suitable real-world case for testing land cover classification using the embedding dataset. We applied a train-test sampling approach within the study area and utilized the RF algorithm with 100 trees to classify imagery from the embedding dataset, Sentinel-2, and Landsat 9. The classification results are presented in Figure 6. To validate the outcomes, we compared the classifications with the Global Mangrove Dataset. The results demonstrate a similar spatial pattern across all datasets. Most areas on the southern islands are classified as mangroves, while the eastern peninsula is consistently identified as an urban area. In contrast, the northern region is dominated by farms and forests, which are classified as "others."

Despite these similarities, some variations were observed in the Landsat 9 and Sentinel-2 classifications. For Sentinel-2, the water bodies in the northern part were misclassified, likely due to the dark tone of the surface resembling water. A similar issue occurred in Landsat 9. Additionally, Landsat 9 exhibited several misclassifications in the southern area, where sea regions were incorrectly labeled as other classes. This issue is likely caused by cloud cover, which affected the algorithm's ability to correctly classify those pixels. Based on these classifications, we computed the overall accuracy to assess the model performance. The embedding dataset achieved the highest accuracy at 83.4%, outperforming the other datasets. However, its advantage was

less pronounced compared to earlier results, as Landsat 9 and Sentinel-2 achieved accuracies of 80% and 75%, respectively. Although the differences among datasets were relatively small (approximately 3–8%), the embedding dataset consistently achieved the highest accuracy, reinforcing its robustness for land cover classification tasks, particularly in mangrove mapping.

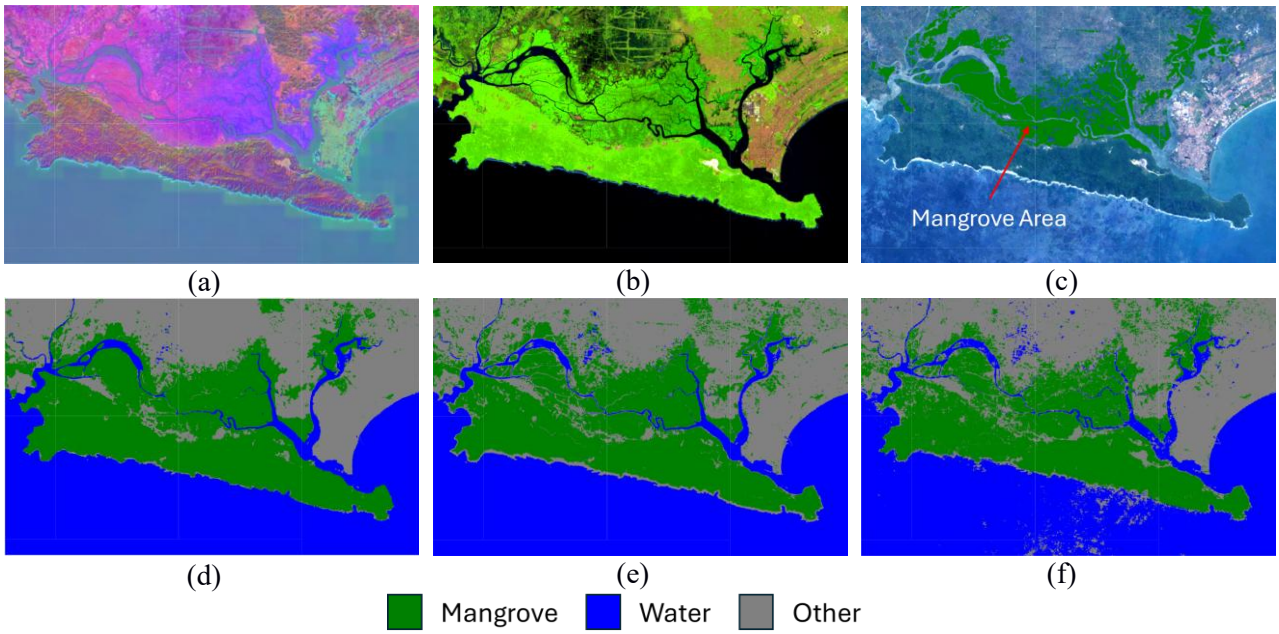


Figure 6. Study case of Mangrove Classification: (a) Embedding Dataset; (b) Sentinel 2 False Color Visualization; (c) Global Mangrove Dataset; (d) Embedding Data RF classification result; (e) Sentinel 2 RF classification result; (f) Landsat 9 RF classification result (Source: GEE catalogue, processed on GEE by the authors).

Discussion

The results demonstrate that the embedding dataset consistently outperforms other datasets across all classification algorithms, while also exhibiting object-based visual characteristics with minimal salt-and-pepper noise. This combination of high accuracy and visually coherent classification makes the embedding dataset a highly advantageous alternative to traditional imagery. Specifically, it achieves accuracy scores above 90% for all tested algorithms, with an average of 94%, which is significantly higher than Sentinel-2 (82.5%) and Landsat 9 (83.1%). Similar performance trends were observed in the mangrove classification case study. Moreover, the embedding dataset produces classification outputs comparable to OBIA, despite having the same spatial resolution as Sentinel-2 (10 m). This suggests that the embedding dataset mitigates common limitations of pixel-based methods without the additional computational complexity required by OBIA.

The superior performance of the embedding dataset can be attributed to its 64-dimensional embedding vectors, which capture rich and diverse information, including spectral, spatial, and temporal features. These embeddings effectively compress multi-sensor time-series datasets into a highly informative feature space, offering more robust classification parameters than conventional band-based inputs. Furthermore, the embedding dataset was designed for low-shot learning, meaning it can achieve strong performance with relatively small training datasets (Brown et al., 2025). This characteristic reduces the need for extensive sampling efforts while maintaining accuracy. Compared to the raw pixel values used in traditional imagery, the embedding vectors provide a more comprehensive representation of Earth surface characteristics, which logically aligns with the significant accuracy improvements observed in this study.

The classification results for Sentinel-2 in this study are consistent with those reported in previous research. For example, Afininnas et al. (2024) classified the Special Region of Yogyakarta into four land cover classes—water bodies, vegetation, barren land, and urban—using CART, RF, and SVM. Their reported overall accuracies were 81.26%, 83.02%, and 80.50%, respectively, resulting in an average accuracy of 81.59%. In

comparison, this study achieved an average of 81.78% for Sentinel-2 using the same classifiers, confirming the reliability of our findings and the inherent limitations of Sentinel-2 for land cover classification, which appear to plateau around 81%. Similarly, Hastoro & Yudinugroho (2023), who applied an unsupervised Wishart H-Alpha angle method on Sentinel-1 SAR data for the same region, reported an accuracy of 80.72%, further reinforcing the consistency of these results across different sensors and methodologies. These comparisons highlight the limitations of conventional datasets and emphasize the potential of the embedding dataset as a viable alternative that delivers higher accuracy with fewer training samples.

The practical implications of these findings are significant. The embedding dataset offers a simpler and more efficient approach to land cover classification compared to OBIA, delivering similar or superior visual and quantitative results without requiring computationally expensive segmentation processes. Additionally, it eliminates the need for preprocessing tasks such as cloud masking, which are necessary for Sentinel-2 and Landsat 9. Its integration within GEE further simplifies the workflow, requiring only basic filtering and mosaicking steps to prepare the dataset for analysis. This ease of use, combined with its superior accuracy and scalability to large areas (e.g., provincial-level mapping), positions the embedding dataset as a promising tool for operational land cover classification tasks.

Despite these advantages, this study focused exclusively on supervised classification. Another limitation of this study is related to the computational constraints of GEE. The embedding dataset is provided exclusively through the GEE platform and consists of a high number of bands, which increases computational complexity and memory requirements compared to conventional multispectral imagery. Although classification at local or provincial scales can be executed within acceptable processing times, analyses over larger spatial extents may result in longer execution times or memory-related task failures due to GEE's platform limitations. As a result, the scalability of embedding-based analyses within GEE remains constrained. Future research should explore the embedding dataset's potential in other applications, such as unsupervised classification, clustering, regression, and similarity search. Additionally, the integration of deep learning approaches with embedding datasets could further enhance classification performance and open new avenues for advanced geospatial analytics.

Conclusions

This study demonstrates the strong potential and performance of the GEE embedding dataset in supervised land cover classification tasks. The results indicate that this dataset consistently outperforms traditional satellite imagery, including Landsat 9 and Sentinel-2, across all tested classification algorithms. Experiments conducted at a large spatial scale (provincial level) using five algorithms—RF with 100 and 1,000 trees, SVM, KNN, and CART—show that the embedding dataset delivers highly accurate and reliable classifications. The embedding dataset achieved an overall accuracy exceeding 90%, with an average accuracy of 94%, substantially higher than Sentinel-2 (82.5%) and Landsat 9 (83.1%). Interestingly, the classification output from the embedding dataset exhibits characteristics similar to object-based classification, effectively reducing the "salt-and-pepper" effect commonly observed in pixel-based approaches. Furthermore, its performance is comparable to, and in some cases surpasses, that of OBIA. The use of 64 embedding vectors appears to capture essential spectral and spatial information, enabling high accuracy and visually coherent classification outputs. These findings highlight the embedding dataset as a promising alternative to conventional remote sensing imagery for land cover mapping.

Future research should explore the integration of embedding datasets into a broader range of applications, including clustering, regression, and similarity-based searches. Additionally, investigating the relative importance and interpretability of individual embedding vectors offers a valuable direction for maximizing the utility of this approach across various remote sensing and geospatial tasks. Another suggestion for future research is the use of in-situ sampling, particularly for assessing the accuracy and performance of the model.

Acknowledgment

The authors gratefully acknowledge Google and Google Earth Engine (GEE) for providing the remote sensing imagery catalog and the powerful cloud-based platform for image processing. We also acknowledge the USGS

and NASA for providing the Landsat imagery collection, and the ESA for the Sentinel imagery collection. The authors would like to thank the Google AlphaEarth Foundation for developing a powerful model and embedding dataset. Finally, we extend our gratitude to the Department of Geodetic Engineering, Faculty of Engineering, Universitas Gadjah Mada, for providing the facilities that supported this research.

References

Adorno, B. V., Körting, T. S., & Amaral, S. (2023). Contribution of time-series data cubes to classify urban vegetation types by remote sensing. *Urban Forestry and Urban Greening*, *79*. https://doi.org/10.1016/j.ufug.2022.127817

Adrian, J., Sagan, V., & Maimaitijiang, M. (2021). Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*(December 2020), 215–235. https://doi.org/10.1016/j.isprsjprs.2021.02.018

Afininnas, F., Nawang Wulandari, Y., Fioren, A., Golo, J., & Kurniawan, R. (2024). Analisis Perbandingan Metode Klasifikasi Pada Pemetaan Tutupan Lahan di Provinsi DI Yogyakarta Tahun 2023. *Seminar Nasional Sains Data*, *2024*.

Ahmad, S. K., Hossain, F., Eldardiry, H., & Pavelsky, T. M. (2020). A Fusion Approach for Water Area Classification Using Visible, near Infrared and Synthetic Aperture Radar for South Asian Conditions. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(4), 2471–2480. https://doi.org/10.1109/TGRS.2019.2950705

Askari, F., Fateh, A., & Mohammadi, M. R. (2025). Enhancing Few-Shot Image Classification through Learnable Multi-Scale Embedding and Attention Mechanisms. *ArXiv*. https://doi.org/10.1016/j.neunet.2025.107339

Ba Alawi, A., & Bozkurt, F. (2025). Performance Analysis of Embedding Methods for Deep Learning-Based Turkish Sentiment Analysis Models. *Arabian Journal for Science and Engineering*, *50*(10), 7299–7321. https://doi.org/10.1007/s13369-024-09360-4

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P. (2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *ArXiv*. http://arxiv.org/abs/2507.22291

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big Data for Remote Sensing: Challenges and Opportunities. *Proceedings of the IEEE*, *104*(11), 2207–2219. https://doi.org/10.1109/JPROC.2016.2598228

Dritsas, E., & Trigka, M. (2025). Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. *Remote Sensing*, *17*(3). https://doi.org/10.3390/rs17030550

Ez-zahouani, B., Teodoro, A., El Kharki, A., El Kharki, O., & Rhachi, H. (2025). Evaluating the potential of Landsat 8/9 and Sentinel 2 data and different spectral and spatial indices for segment extraction in large watersheds for OBIA approach in remote sensing: A case study of the Sebou watershed. *Remote Sensing Applications: Society and Environment*, *38*. https://doi.org/10.1016/j.rsase.2025.101575

Fattore, C., Abate, N., Faridani, F., Masini, N., & Lasaponara, R. (2021). Google earth engine as multi-sensor open-source tool for supporting the preservation of archaeological areas: The case study of flood and fire mapping in metaponto, italy. *Sensors*, *21*(5), 1–27. https://doi.org/10.3390/s21051791

Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., Masek, J., & Duke, N. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, *20*(1), 154–159. https://doi.org/10.1111/j.1466-8238.2010.00584.x

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Harris, N., Butani, A., & Hashmy, S. (2024). Enhancing Embedding Performance through Large Language Model-based Text Enrichment and Rewriting. *ArXiv*. http://arxiv.org/abs/2404.12283

Hastoro, D. A., & Yudinugroho, M. (2023). Analisis Klasifikasi Tutupan Lahan dengan Citra Sentinel 1A Menggunakan Metode Dekomposisi Polarimetrik di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Ilmiah Geomatika*, *3*(2), 58. https://doi.org/10.31315/imagi.v3i2.10778

Hird, J. N., DeLancey, E. R., McDermid, G. J., & Kariyeva, J. (2017). Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sensing*, *9*(12). https://doi.org/10.3390/rs9121315

Kang, D., Kwon, H., Min, J., & Cho, M. (2021). Relational Embedding for Few-Shot Classification. *ArXiv*. http://cvlab.postech.ac.kr/research/RENet

Kumar, L., & Mutanga, O. (2018). Google Earth Engine applications since inception: Usage, trends, and potential. *Remote Sensing*, *10*(10). https://doi.org/10.3390/rs10101509

Li, H., Cui, J., Zhang, X., Han, Y., & Cao, L. (2022). Dimensionality Reduction and Classification of Hyperspectral Remote Sensing Image Feature Extraction. *Remote Sensing*, *14*(18). https://doi.org/10.3390/rs14184579

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, *51*, 47–60. https://doi.org/10.1016/j.future.2014.10.029

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. (2019). The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the Google Earth Engine cloud computing platform. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010043

Padarian, J., Minasny, B., & McBratney, A. B. (2015). Using Google's cloud-based platform for digital soil mapping. *Computers and Geosciences*, *83*, 80–88. https://doi.org/10.1016/j.cageo.2015.06.023

Parente, L., Taquary, E., Silva, A. P., Souza, C., & Ferreira, L. (2019). Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, *11*(23). https://doi.org/10.3390/rs11232881

Radhakrishna, A., & Susstrunk, S. (2011). Superpixels and Polygons using Simple Non-Iterative Clustering. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4651–4660. https://doi.org/10.1109/CVPR.2017.520

Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science*, *163*, 292–299. https://doi.org/10.1016/j.procs.2019.12.111

Shlens, J. (2003). A Tutorial on Principal Component Analysis. *ArXiv*.

Tassi, A., & Vizzari, M. (2020). Object-Oriented LULC Classification in Google Earth Learning Algorithms. *Remote Sensing*, *12*, 1–17. doi:10.3390/rs12223776

Velastegui-Montoya, A., Montalván-Burbano, N., Carrión-Mero, P., Rivera-Torres, H., Sadeck, L., & Adami, M. (2023). Google Earth Engine: A Global Analysis and Future Trends. *Remote Sensing*, *15*(14). https://doi.org/10.3390/rs15143675

Yang, C., & Suh, C. P. C. (2023). Applying machine learning classifiers to Sentinel-2 imagery for early identification of cotton fields to advance boll weevil eradication. *Computers and Electronics in Agriculture*, *213*. https://doi.org/10.1016/j.compag.2023.108268