Integration of Principal Component Analysis and Very Low Frequency Geophysical Method for Modeling Recharge Area Distribution

The Case Study of Krawak Spring, Tuban

Moh. Singgih Purwanto
Universitas Brawijaya, Malang, Indonesia | Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
singgih16@student.ub.ac.id

Adi Susilo
Universitas Brawijaya, Malang, Indonesia
adisusilo@ub.ac.id (corresponding author)

Agus Naba
Universitas Brawijaya, Malang, Indonesia
anaba@ub.ac.id

Ayi Syaeful Bahri
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
syaeful_b@geofisika.its.ac.id

Siti Navisa
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
sitinavisabcd@gmail.com

Received: 8 February 2025 | Revised: 11 March 2025 | Accepted: 19 March 2025

 DOI: https://doi.org/10.48084/etasr.10508

ABSTRACT

Recharge areas play an important role in the subsurface hydrological cycle, especially in karst regions such as Tuban Regency, which has unique hydrological and topographical conditions with steep morphology. This research was conducted around the Krawak spring, as this spring is very important for the needs of the surrounding community, so it must be maintained to ensure that the land function does not change, especially the recharge area, thus preserving the sustainability of the Krawak spring. Considering the difficult terrain for direct field surveys, this research utilizes a machine learning method based on Principal Component Analysis (PCA) with Very Low Frequency (VLF) data and Landsat-8 satellite imagery as supporting data. VLF data are used as training data for the machine learning model, whereas Landsat-8 imagery serves as the main data source, which is processed to produce five classification parameters, namely Normalized Difference Vegetation Index (NDVI), land cover, elevation, slope, and soil type. The results of this classification are then weighted and scored to produce a map of the potential recharge area distribution. The results of the PCA analysis show consistency with conventional scoring methods as well as the results from VLF data, making this method effective for mapping recharge areas in regions with challenging topography. This research provides an efficient and accurate alternative for modeling recharge areas in karst environments, where with direct field surveys are limited.

Keywords-PCA; recharge area; machine learning; VLF; Landsat-8

I. INTRODUCTION

Groundwater is the flow of rainwater beneath the earth's surface that occurs due to geological structures, differences in soil moisture potential, and the earth's gravity [1]. The groundwater cycle begins in the groundwater recharge area, which is an area where groundwater flows vertically downward or follows the slope of the aquifer [2]. Recharge areas play a

crucial role in maintaining the sustainability of the global water cycle. East Java Province is among the regions prone to water crises, with 23 areas, including Tuban, falling into the critical drought category [3]. In Tuban Regency, there are 9 sub-districts with drought potential, including Singgahan and Montong, which have many limestone or karst hills. Therefore, the analysis and determination of recharge areas in Tuban, particularly in Singgahan and Montong sub-districts, is very important for maintaining groundwater sustainability.

Based on its geology shown in Figure 1, the research area falls within the Rembang Zone with a morphology of hills and folded valleys, commonly known as "the hilly district of Rembang" [4]. This zone is bordered by the northern part of the Java Sea and the Randublatung Depression Zone to the south, with the Kujung Fault as the boundary. Meanwhile, the western part is bordered by the Semarang-Rembang Zone and directly connected to the northern plains of Madura Island in the eastern part, interspersed with fluvial plains. The Rembang Zone has a width of approximately 50 km, with the highest peak being Gading (515 m above sea level), predominantly composed of carbonate rocks [5]. Then, for the stratigraphic arrangement of the Rembang Zone, the basement rocks are mainly covered by metamorphic rocks of Cretaceous age, such as schist, phyllite, and diorite igneous rocks. The overlying Ngimbang Formation consists mainly of Oligocene age carbonate deposits of the Kujung and Prupuh Formations. The formations above the Kujung Formation, namely the Tuban Formation, Tawun Formation, and Ngrayong Formation (middle Miocene to late Miocene), consist of quartz sandstone with intercalations of bioclastic limestone, sandy limestone, and claystone deposited in shallow marine environments. For the Bulu Formation (late Miocene), it is composed of bioclastic limestone, sandy limestone, and intercalated limestone sandstone deposited in a foreslope environment. Meanwhile, the Wonocolo Formation (late Miocene) consists of sandy shale and sandy limestone lithologies and is deposited in an open shelf environment. The Ledok Formation (early Pliocene) consists of sandy limestone, sandy shale, and limestone sandstone. The Mundu Formation (early Pliocene) consists of types of marl, sandy limestone, and limestone sandstone, and its depositional environment is in an open shelf environment. Meanwhile, the Paciran Formation (Pleistocene) consists of reef limestone, bioclastic limestone, and chalk limestone [4].

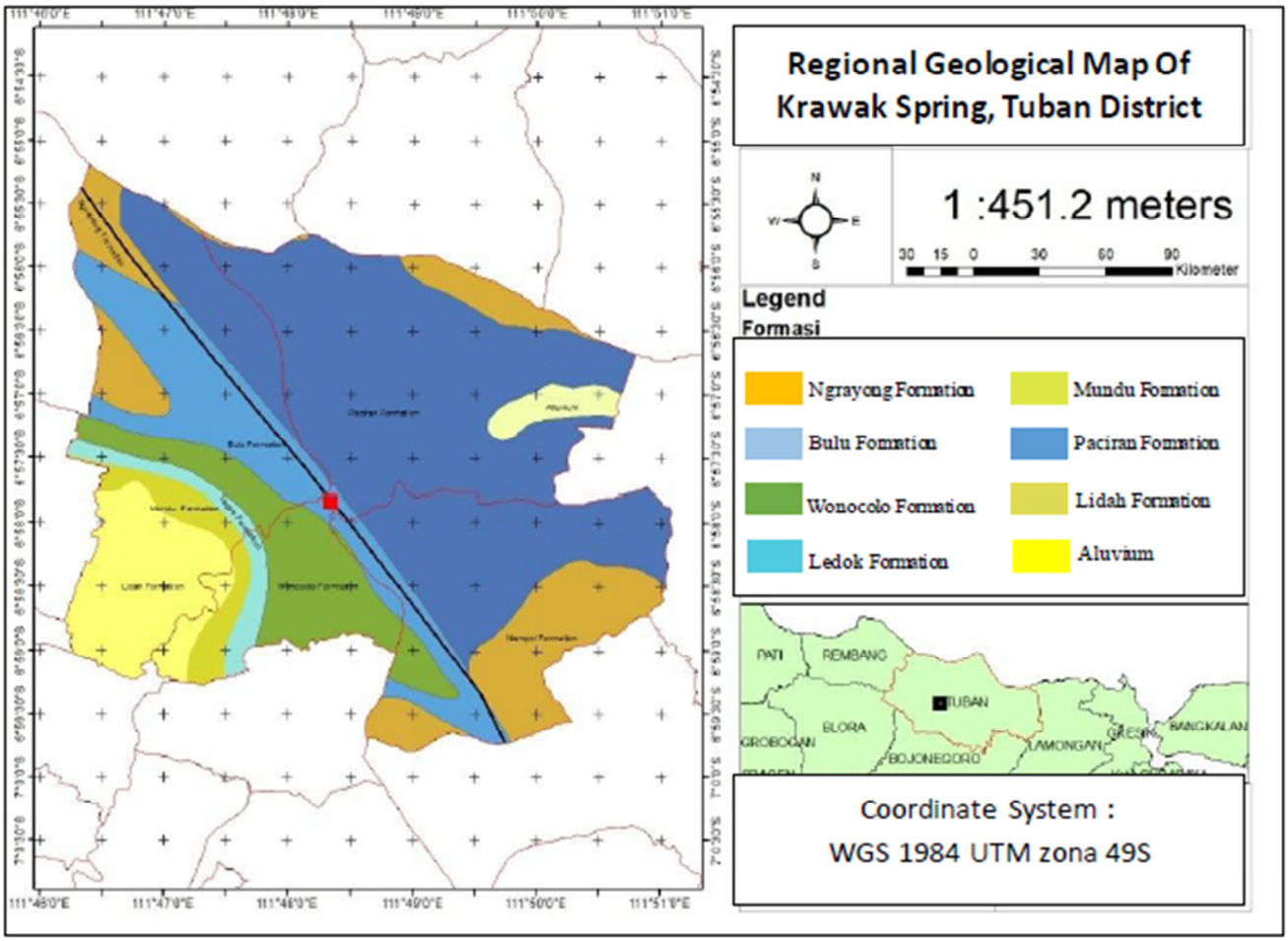


Fig. 1. Geological map around the Krawak spring source.

The landscape or morphology of this region is karst, which has a unique hydrogeological condition resulting from a combination of high rock solubility and secondary porosity in the form of intensive faults and well-developed fractures. Areas with karst landscapes are characterized by surface morphology (exokarst) such as single hills, dolinas, springs, and the disappearance of surface rivers that later become underground rivers, as well as endocast morphology, indicated by the presence of cave systems and underground river flows due to the process of karstification [6]. Based on the type of karst topography in the research area, an effective geophysical method is needed to map underground rivers. One of the selected methods is the Very Low Frequency Electromagnetic (VLF-EM) method, because it can effectively detect underground river flows in karst areas through differences in conductivity caused by cracks filled with water [7]. Previous research in Tuban and Pacitan used the VLF-EM method to map underground rivers [8, 9]. To accurately determine the recharge area, a scoring method is used to assess the relevant parameters [10]. Mapping is carried out by utilizing Geographic Information System (GIS) and remote sensing data, as done by Magesh et al. [11], who used various parameters such as lithology, slope, and land use. Various AI methods are also used to determine the quality and potential of groundwater. Che Nordin et al. [12] compared 4 AI methods, Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Evolutionary Algorithms (EA), and Support Vector Machines (SVM), in modeling the biggest challenges in determining groundwater quality. Pourghasemi et al. [13] compared the SVM, Multivariate Adaptive Regression Spline (MARS), and Random Forest (RF) methods in determining recharge areas using parameters such as Digital Elevation Model (DEM) and rainfall. Martinsen et al. [14] used the RF method to predict recharge areas using satellite data. Also, Abdelaziz et al. [15] used the Hierarchical Clustering Analysis (HCA) and Principal Component Analysis (PCA) methods to classify groundwater quality data and identify the most critical parameters. Based on this study, there is an opportunity to design a more effective system for identifying recharge areas, using Very Low Frequency (VLF) data as training data.

Through comprehensive analysis and examination of the issue, it was determined that there is a necessity to develop a system for automatic identification of recharge areas utilizing machine learning, with VLF results serving as the training data. A method is needed to minimize images based on various parameters, including Normalized Difference Vegetation Index (NDVI), DEM, land cover, and land slope, utilizing PCA to enhance the identification process of recharge areas. This study focuses on developing an automated system that can identify recharge areas and generate PCA-reduced images based on various relevant parameters. The Landsat-8 imagery was obtained from USGS Earth Explorer in August 2023. Several spectral bands were selected to optimize the identification of groundwater recharge areas, including band 2 (blue), band 3 (green), band 4 (red), band 5 (near-infrared), band 6 (shortwave infrared 1), and band 8 (panchromatic). The VLF-EM field measurements were conducted to provide supporting geophysical insights. The field survey consisted of multiple acquisition lines of varying lengths, each measured at three different frequencies per day to ensure comprehensive subsurface characterization. Measurement stations were spaced 3 m apart for most of the survey lines, except for one shorter line where measurements were taken at 1 m intervals to capture finer subsurface variations. The number of measurement points for each line varied, ranging from 77 to 105 points, depending on the total line length. Data acquisition followed specific orientations, with some lines oriented in a south-north direction, whereas others were measured in an east-west or west-east direction using inphase and quadrature components, tilt, and total field values, which were recorded particularly on selected survey lines to enhance data accuracy. Consequently, the findings of this study include:

1. The creation of recharge area mapping models utilizing VLF geophysical data through a machine learning approach which enhances identification accuracy and can act as a reference for water resource management in comparable regional contexts.
2. Reference for studies on image reduction utilizing the PCA method and the role of machine learning in locating recharge areas to assist in drought mitigation strategies.
3. Citation for government policies aimed at preserving recharge areas, such as the Krawak spring, to guarantee sustainable water resources and bolster regional spatial planning.

II. RELATED WORKS

This research is related to several previous studies, and the literature review shows that the PCA method has been applied before, but in a different context and with a different database. The previous research used as a reference in this study is presented in Table I. Based on the literature review, a research gap was identified regarding the use of VLF-EM geophysical data in recharge area modeling using a machine learning approach. Although the VLF method has long been used for shallow geophysical mapping such as subsurface structure identification, conductivity mapping, and groundwater zone detection, its use as a primary data source for recharge area modeling is still very limited. These limitations are particularly evident in the minimal integration between the interpretation results of VLF data and machine learning algorithms, which have great potential to improve the accuracy of recharge area predictions compared to conventional GIS-based methods. Furthermore, the application of PCA techniques has not yet been further developed, particularly in the field of modeling and mapping recharge areas. The research is increasingly relevant because no specific studies have been found that model and map recharge areas based on VLF data using the PCA algorithm in the Krawak spring area, Tuban, East Java, which has unique hydrological conditions due to its morphology. Therefore, the development of a predictive model that integrates VLF geophysical data and machine learning in this area can be an opportunity to address the research gap.

TABLE I. RELATED WORKS TO THIS RESEARCH

Reference	Topic	Method	Subject
[8]	Underground river mapping	VLF-EM	Conductive zone analyzed as river continuity
[16]	Geological state	Geological study	Geological setting of Tuban
[17]	Near surface geophysics	VLF-EM	Subsurface resistivity anomaly
[18]	Recharge area determination	AHP, overlay	Recharge area mapping using geomorphology, geology, soils, slope, land use, and data lineage
[19]	Classification and mapping	Remote sensing	Apply remote sensing technologies in delineating sugarcane plantations and identifying its growth
[20]	Recharge area	GIS, scoring, overlay, AHP	Recharge area mapping using scoring method
[21]	Monitoring water quality	Remote sensing	Estimating water quality using PH, DO, turbidity, and conductivity parameters
[22]	Image classification	Racah polynomial, Discrete Moment	Improvement of image classification accuracy using discrete moment variant
[23]	Geophysical data dimension	PCA	Better interpretation, and minimization of information loss
[24]	Clustering	PCA	Seismic facies clustering for seismic mapping
[25]	Image analysis	Wavelet transformation	Hidden feature detection of landslide image from remote sensing data
[26]	Object prediction in UAV images	CNN, tranformer	Object detection using a combination of CNN and transformer architecture
[27]	Data reduction	PCA	Enhanced-PCA based dimensionality reduction

III. METHODOLOGY

In this research, several approaches are used, which are explained in detail in the following section.

A. Classification Standard

The classification criteria of this research are based on the scoring methodology. The scoring approach involves assigning scores or values to each parameter to assess its level of competence, and this assessment is based on established criteria [28]. Each criterion has a distinct weight or score and classification that is used to determine the recharge area. The land cover classification is shown in Table II with the corresponding scores ranging from 1 to 5. Table III presents the classification criterion for the NDVI, which is based on its impact and is divided into five classes. For elevation, the classification criterion is outlined in Table IV, based on the elevation of the area, suggesting that regions at higher elevations are more likely to function as groundwater recharge zones. Table V presents the classification of the slope or inclination of the land and Table VI presents the classification of the soil type.

TABLE II. LAND COVER CLASSIFICATION [29]

Land Cover	Class	Score
Dense forest	Very high	5
Production forest	High	4
Meadow, field, moor	Medium	3
Settlement	Low	2
Water body	Very low	1

TABLE III. NDVI CLASSIFICATION [30]

NDVI	Score
Very high	5
High	4
Average	3
Low	2
Very low	1

TABLE IV. TOPOGRAPHY CLASSIFICATION [31]

Topography	Score
Very high	5
High	4
Average	3
Low	2
Very low	1

TABLE V. SLOPE CLASSIFICATION [29, 32]

Slope	Class	Score
<8%	Very high	5
8-15%	High	4
15-25%	Medium	3
25-40%	Low	2
>40%	Very low	1

TABLE VI. SOIL TYPE CLASSIFICATION [29, 33]

Soil Type	Class	Score
Regosol	Very high	5
Aluvial	High	4
Kambisol, gleisol	Medium	3
Litosol, mediteran	Low	2
Grummusol	Very low	1

B. Research Location

The research site illustrated in Figure 2 is located in Krawak, Singgahan, Krajan, Guwoterus, Montong, Tuban Regency. The exact coordinates of the Krawak spring source are -6.96334, 111.804103 and it is located within the protected Krawak forest area in Guwoterus village, Montong district, Tuban Regency, East Java. The surrounding area benefits from it as a source of clean water, irrigation for rice fields, and daily living needs [34]. This spring is located in the Rembang zone of the Bulu Formation, which is characterized by a lithology that includes bioclastic limestone and sandy limestone with interspersed limestone sandstone [28]. In the area of the Krawak spring, multiple water sources emerge from below the surface, and there is a cave located approximately 200 m north of the spring. The present sources are shown in detail in Figures 3, 4, and 5.

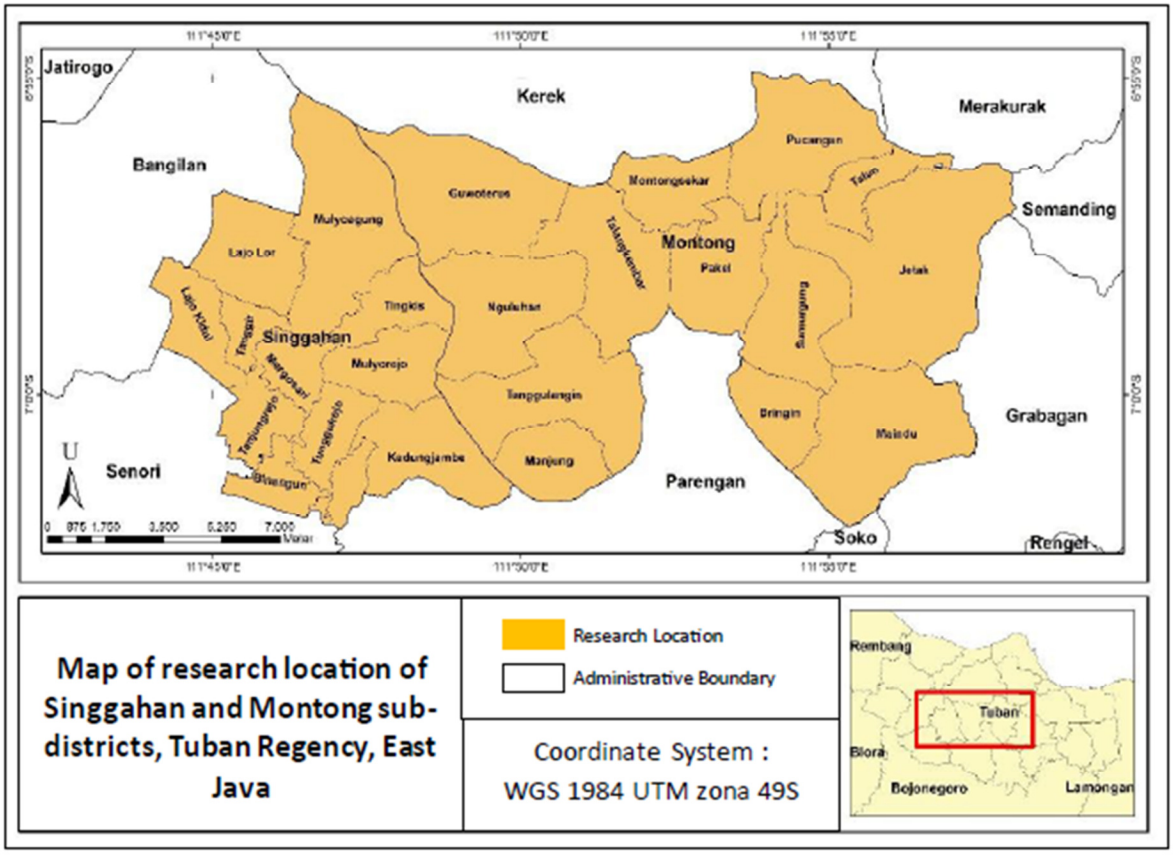


Fig. 2. Research site map.

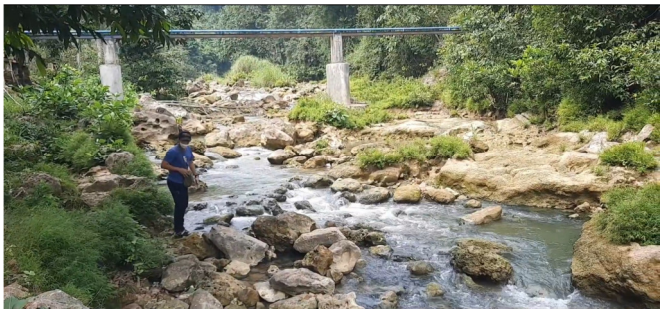


Fig. 3. Krawak river from the Krawak spring.

Fig. 4. Underground spring.

Fig. 5. Cave located in the northern region of the Krawak spring.

C. Research Procedure

This study consists of a collection of laboratory and field investigations. The stages of laboratory research include the collection of Landsat-8 OLI/TIRS data, identification of the Area of Interest (AoI), as well as preprocessing and processing steps. The field investigations include VLF-EM measurements to determine the orientation of the subsurface water continuity.

1) Data Collection

The data collection process utilized two distinct methods: online and offline approaches. The online approach involves gathering data from platforms that provide resources, including Google and Landsat-8 satellite imagery available at https://earthexplorer.usgs.gov. In contrast, the offline approach

involves conducting surveys and performing direct measurements in the field. Field measurements were performed using the VLF-EM method, which served as training data, as shown in Figure 6.

Fig. 6. Site of VLF survey, © Airbus, CNES / Airbus, Maxar Technologies.

2) *Examination and Structuring*

Both qualitative and quantitative analyses were performed. The VLF-EM data underwent filtering through Noise-Assisted Multivariate Empirical Mode Decomposition (NA-MEMD) to remove noise from the dataset. Quantitative analysis was conducted using Inv2DVLF software, whereas qualitative analysis was performed using MATLAB software. The results of the process include 2D and 3D cross-sections, utilized with Surfer 11 and SketchUp software, respectively.

The system design is implemented through an analysis of the VLF-EM data, taking into account various parameters. The parameters considered include NDVI, land cover, DEM, land slope, and soil type. All these parameters are subjected to image classification to generate score labels, and then reduction is performed using the PCA method. The development of the system utilizes Python alongside GIS software.

D. Data Processing Procedure

In this study, VLF data are used as training data to identify the recharge area through the PCA method. The subsequent phase in the identification of the recharge area involves the development of a system, which uses an image captured by the Landsat-8 satellite as an input. The subsequent stage of image processing by the system involves the analysis of 5 parameters: NDVI, land cover, elevation or DEM, land slope, and soil type. Each parameter is categorized according to the pixel value of the image. The parameters for NDVI, DEM, and land slope rely on pixel color, whereas land cover segmentation utilizes both pixel color and pixel pattern, with the satellite image featuring a pixel resolution size of 30×30 m. The classified image, with each pixel assigned a value according to the established class, enters a further stage where the value of each pixel is subjected to a weighting. The weighting plays a crucial role in defining the recharge area. The outcome of this weighting is processed by the PCA method, anticipating that the PCA results will produce a more practical, accurate, and efficient representation in identifying the recharge area, as shown more clearly in Figure 7.

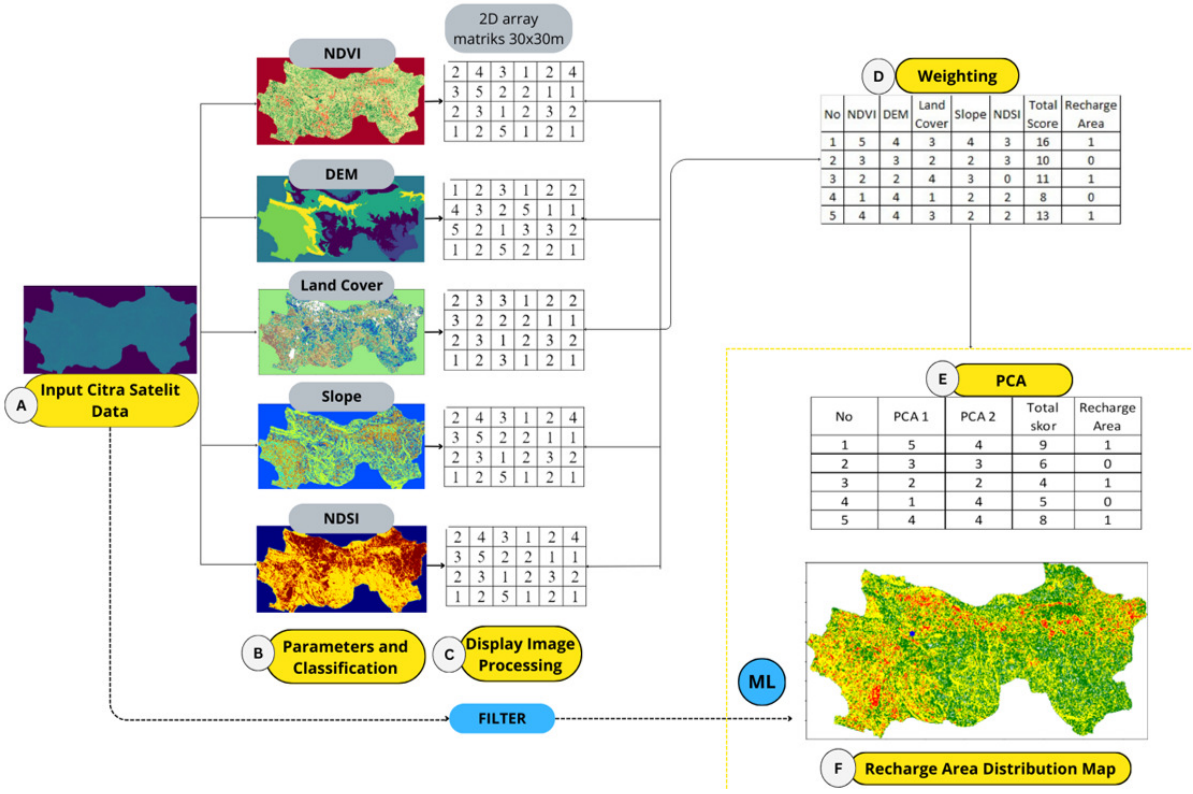


Fig. 7. Data processing flowchart with PCA.

IV. RESULTS AND DISCUSSION

A. Area of Interest

AoI is derived from raster data that incorporate the spectral image of Landsat-8 band 1, along with a shapefile utilized to define the boundary for extracting the raw raster data. The procedure typically involves reading the image and shapefile, verifying and aligning the CRS, then cropping the image according to the specified geometry, and finally generating a visual map from the subsequent steps. The region highlighted in bright blue in Figure 8 represents the focal point of this study, whereas the areas beyond those limits fall outside the parameters of this investigation.

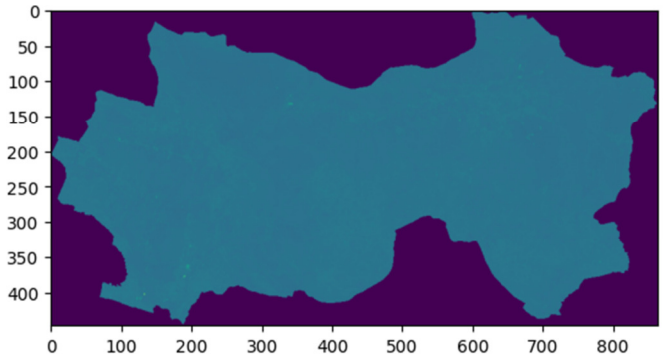


Fig. 8. Area of interest.

B. Land Cover Classification Results

The processing of land cover parameters involves the utilization of Landsat-8 imagery, specifically band 5 (near infrared), band 4 (red), and band 3 (green). These bands are derived from clipping for the purpose of clustering based on spectral data. The data from the three bands are organized into a 3D array characterized by height, width, and band dimensions, leading to a spectral combination from each band that encapsulates color information. The 3D data are converted into a 2D array, where each row corresponds to a pixel and each column indicates the value of each band. Subsequently, K-means clustering is employed to categorize the spectral data into multiple clusters based on the similarity of the spectral values. In this study, the K-means algorithm is used to categorize the pixels into 6 clusters based on their spectral values. The results are presented as labels and restructured in 2D to align with the original image dimensions, as shown in Figure 9.

The next step involves evaluating the land cover according to its clusters and delivering detailed descriptions of the various land cover types. Cluster 1 receives a score of 0, indicating that it is unclassified. Cluster 2 is assigned a score of 1, which characterizes it as surface water. Cluster 3 is assigned a score of 2, described as settlement. Cluster 4 has a score of 3, identified as a farm. Cluster 5 is given a score of 4, identified as plantation, whereas cluster 6 is given a score of 5, identified as forest. The result of this procedure is an image map where each pixel is assigned a land cover score according to the cluster, reflecting the recognized land cover.

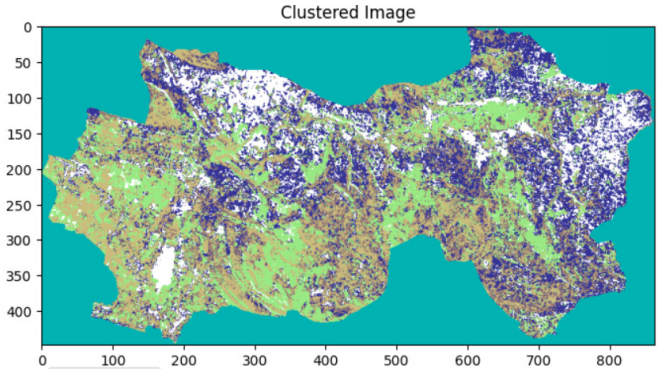


Fig. 9. Clustered land cover image.

The land cover parameter in this study accounts for 20% of the total parameters utilized. This parameter is a key consideration in identifying potential recharge areas in a region. Land that has been converted into settlements or productive areas, such as rice fields, gardens, or farms, will influence the dynamics during the rainy season. In such cases, rainwater is likely to become surface runoff rather than seeping into areas devoid of vegetation or woody plants. This condition leads to a reduction in recharge areas or spring sources. According to the analysis of Landsat-8 bands 3, 4, and 5 presented in Figure 10, the land cover classes can be categorized into 5 primary types: surface water, represented in light blue, which is less prominent than other land covers; residential areas shown in green, which notably dominate the Siggahan and Montong regions; agricultural land, represented in light brown and plantations, represented in dark brown, which are adjacent and exhibit a similar distribution; and finally, forests, represented in white on the map, which are located in the southwestern part of the AoI as well as in the northwestern and southeastern regions.

C. Normalized Difference Vegetation Index Classification Result

The NDVI parameter utilizes Landsat-8 band 5 (near-infrared) and band 4 (red) imagery in its processing, following the standard equation [35]:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \quad (1)$$

The final process displays the vegetation value of a specific area, with values spanning from -1 to +1. Higher values indicate a greater amount of vegetation present. The clustering process involves stacking 2 bands into a 3D format, then converting it back to 2D and applying the K-means algorithm for clustering, which yields 6 clusters as illustrated in Figure 11. Analysis of the 6 existing clusters revealed the following vegetation levels and corresponding scores: cluster 1 is categorized as "Unclassified" with a score of 0, cluster 2 is classified as "Very Low" with a score of 1, cluster 3 is designated as "Low" with a score of 2, cluster 4 is identified as "Average" with a score of 3, cluster 5 is marked as "High" with a score of 4, and cluster 6 is recognized as "Very High" with a score of 5.

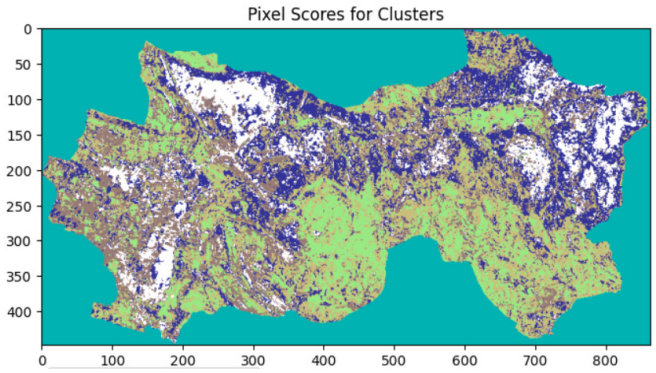


Fig. 10. Land cover Area classification.

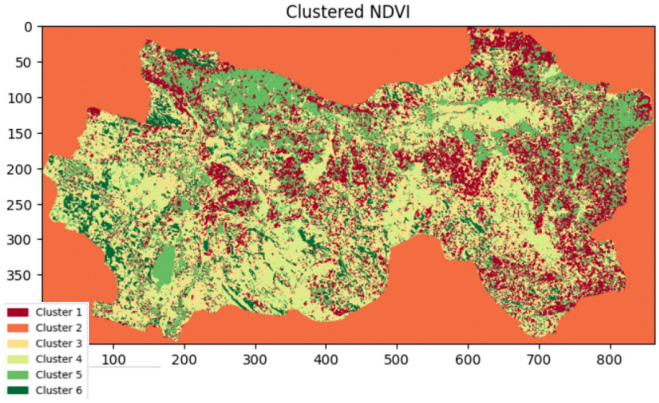


Fig. 11. Clustered NDVI image.

In the NDVI parameter, areas of very low vegetation are represented in orange, dispersed among regions of agricultural and plantation land cover. Low vegetation is represented in yellow, whereas average vegetation is represented in light green, both predominantly found in areas of residential land cover. High and very high are shown in green and represent areas of forest land cover, which may include dense forest. Regions exhibiting elevated NDVI values may serve as potential recharge zones beneath them, as the presence of trees allows for root systems that can effectively retain water underground.

Figure 12 presents the NDVI values, providing a comprehensive overview of the vegetation density levels in the area. High NDVI scores (green) indicate regions with denser vegetation, which plays a crucial role in maintaining hydrologic balance. This is due to the fact that dense vegetation typically slows surface runoff and improves the soil's ability to absorb water. This map corroborates the findings from the recharge map, indicating that regions with significant recharge often coincide with areas of dense vegetation, particularly forests and certain plantation areas. In contrast, regions exhibiting low NDVI scores, highlighted in red, typically feature minimal vegetation and often align with low recharge zones, including residential neighborhoods and other developed areas.

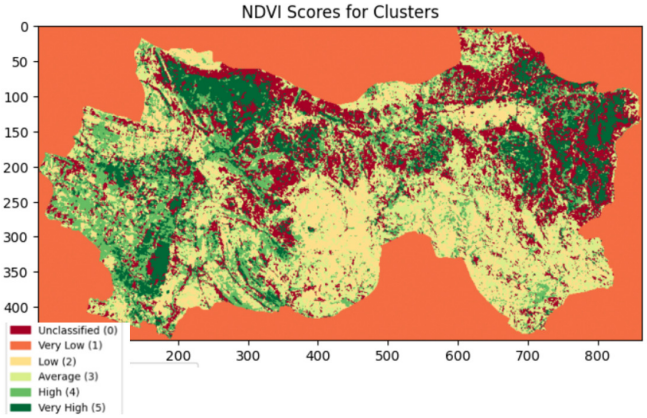


Fig. 12. NDVI area classification.

D. Elevation Classification Results

This elevation data processing utilizes a DEM, which serves as a digital representation of the area's elevation. The initial step involves examining the DEM data, which provides elevation information structured as a 2D array that illustrates the geographic height of the region. A visualization is then performed to illustrate the variations in elevation, with lighter hues representing higher elevations and darker hues representing lower elevations. To perform the clustering analysis, the DEM data, initially structured as a 2D array, are transformed into a 1D array. In this 1D array, each element corresponds to the elevation value at a specific point. The K-means algorithm is then used for clustering.

At this stage, the K-means algorithm analyzes the elevation data and classifies the DEM pixels into clusters according to the similarity of their elevation values. From the 4 clusters, 6 elevation classifications were derived: cluster 1: "Unclassified" with a score of 0 and a mean elevation of 141.9 m; cluster 2: "Very Low" with a score of 1 and a mean elevation of 245.899 m; cluster 3: "Low" with a score of 2 and a mean elevation of 0.0 m; cluster 4: "Average" with a score of 3 and a mean elevation of 48.10 m; cluster 5: "High" with a score of 4 and a mean elevation of 309.63 m; cluster 6: "Very High" with a score of 5 and a mean elevation of 193.9 m. The results are presented as a 2D array, reflecting the true dimensions of the DEM data, as shown in Figures 13 and 14. Regions characterized by elevated terrain are more likely to serve as rainwater recharge zones, as precipitation tends to gather and generate surface runoff that naturally flows toward lower elevations under the influence of gravity.

E. Slope classification results

Using the available DEM data, we conducted calculations to assess elevation changes in both the horizontal (X) and vertical (Y) directions. The slope is then determined by applying the fundamental gradient formula as outlined in (2).

$$\text{Slope} = \arctan(\sqrt{x^2 + y^2} \quad (2)$$

The results provide the slope value in radians, which is then converted to degrees. Once the slope values are obtained, the next stage involves clustering using the K-means algorithm. In this study, 6 clusters and 6 classifications were used: cluster 1:

"Outside" with a score of 0 and a mean slope value of 68.8 degrees; cluster 2: "Very Low" with a score of 1 and a mean slope value of 0.0 degrees; cluster 3: "Low" with a score of 2 and a mean slope value of 30.6 degrees; cluster 4: "Average" with a score of 3 and a mean slope value of 79.2 degrees; cluster 5: "High" with a score of 4 and a mean slope value of 47 degrees; cluster 6: "Very High" with a score of 5 and a mean slope value of 58.4 degrees, as illustrated in Figure 15.

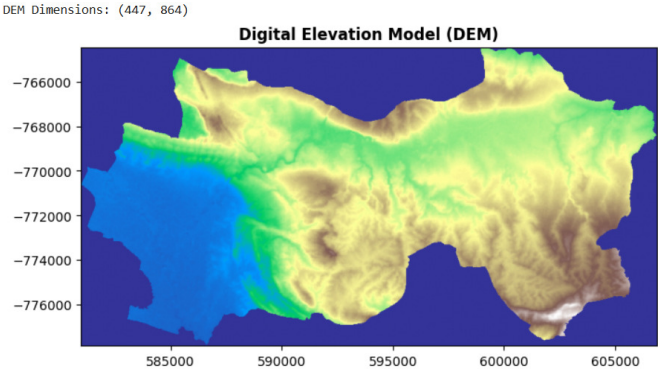


Fig. 13. Clustered elevation image.

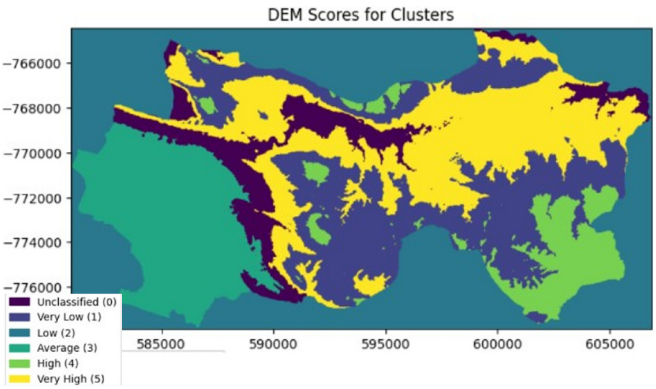


Fig. 14. Topographic area classification.

Very low classifications are shown in orange, low classifications are shown in dark blue, and average classifications are shown in red. These classifications are typically found in regions characterized by residential, plantation, and field land cover, which are predominant in the study area. One notable classification is high slope, which is frequently observed at both the northern and southern boundaries of the study area, in addition to the southern boundary of the study area. These steep regions continue to serve as productive land. Another classification refers to the region characterized by the steepest slopes or very high slopes, predominantly located in the western part of the study area, and is associated with forest cover. Regions with high infiltration potential are typically located in areas with flat topography and permeable soils. In contrast, regions with low potential are often found on steep terrains or consist of soils that are poorly conducive to water infiltration. Slopes maps illustrate the categorization of slopes according to their degree of steepness. Areas with low to very low slopes exhibit significant recharge capacity, as water can be absorbed more readily. Conversely, areas with steep slopes tend to experience increased surface runoff, leading to a decreased recharge potential.

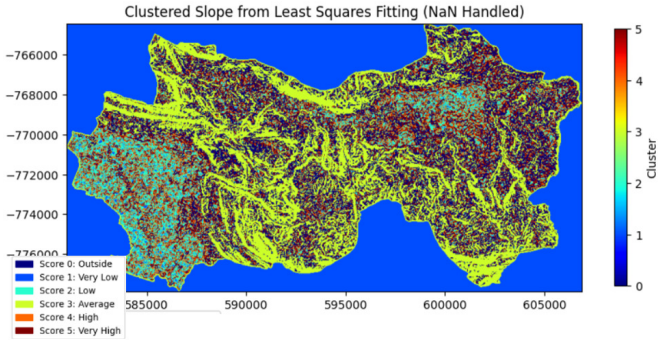


Fig. 15. Slope area classification.

F. Soil Type Classification Results

The parameter for this soil type was derived from the Normalized Difference Soil Index (NDSI), utilizing input data from Landsat-8 satellite imagery bands 5 and 6, and subsequently calculated using (3).

$$\text{NDSI} = \frac{\text{Band 5} - \text{Band 6}}{\text{Band 5} + \text{Band 6}} \quad (3)$$

The data are stacked and converted to a 3D array, then converted back to a 2D array for K-means clustering, as shown in Figure 16. Six clusters were identified: cluster 1: "Kambisol" with a score of 3; cluster 2: "Outside" with a score of 0; cluster 3: "Litosol" with a score of 2; cluster 4: "High" with a score of 2; cluster 5: "Mediteran" with a score of 2; cluster 6: "Gleisol" with a score of 3.

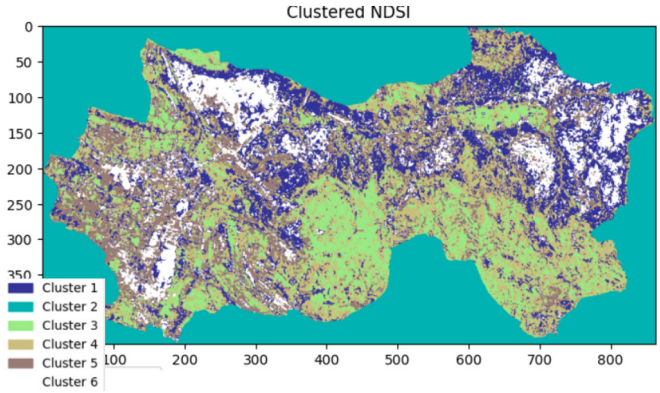


Fig. 16. Clustered NDSI image.

The results are shown in the map in Figure 17. The study revealed that the "Kambisol" and "Gleisol" soil types, characterized by a dark brown color, are predominantly found in the northern region and areas with forest cover. In contrast, the southern region of the study is primarily characterized by "Litosol", "High", and "Mediterranean" soil types, which are indicated by a yellow color. The "Kambisol" and "Gleisol" soils exhibit excellent permeability, indicating their significant potential as subsurface recharge areas. "Kambisol" soils exhibit excellent water absorption properties and are often found in

regions with significant recharge potential. Conversely, "Litosol" and "Mediterranean" soils, which are characterized by their lower permeability, are often found in regions with steep slopes and limited water infiltration capabilities. "Gleisol", characterized by poor drainage, can impede the infiltration process even in relatively flat areas.

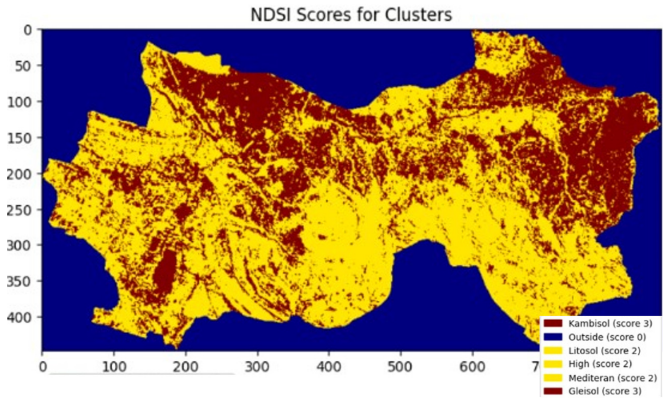


Fig. 17. Soil type area classification.

G. Principal Component Analysis

Following the identification of 5 key parameters as input variables in the PCA, namely land cover, NDVI, soil type, elevation, and slope, the first action required is to gather and organize the training data to be used in the PCA analysis. At this point, several critical aspects must be considered to ensure the precision and significance of the analysis, such as data normalization, variance ratio, eigenvalue, and loading factor.

1) Variance Ratio and Eigenvalue

The explained variance, or variance ratio, indicates the proportion of the total variance in the dataset accounted for by each principal component. A higher value of explained variance for a component indicates a greater contribution to explaining the data distribution, as shown in (4).

$$S^2 = \frac{\Sigma_{i=1}^{n}(Xi - \bar{X})^2}{(n-1)} \quad (4)$$

Variance quantifies the extent to which the data deviate from the mean. In PCA, variance measures the extent to which the principal component can account for the variation in the original data. A greater variance in a principal component indicates that the component contains a greater amount of information derived from the original data.

2) Loading Factor

The loading factor shows the relationship between the initial variables and the principal components in PCA. The loading factor is used to clarify the degree to which a variable influences each principal component. A higher value of the loading factor indicates a more significant contribution of that variable to a particular principal component.

H. Training Data

The training of data is derived from the pixel coordinates of the VLF data. At this stage, a radius of 500 m is defined, yielding training data consisting of 30×30 pixels, each pixel representing an area of 30 m, as shown in Figure 18.

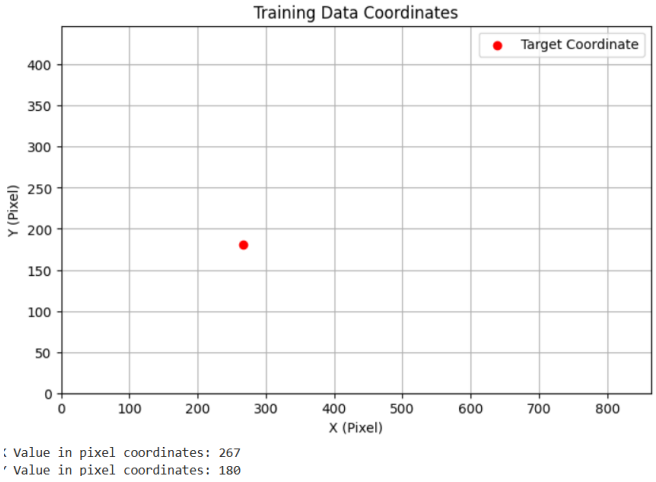
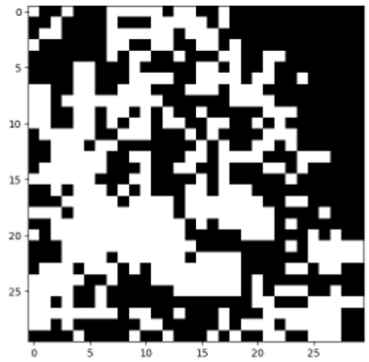


Fig. 18. Coordinates of the VLF training dataset.

The data are normalized using StandardScaler to be compatible with PCA standards and then the eigenvalue calculation is performed with n components set to 5. The results obtained from the analysis of the PCA training dataset, as shown in Figure 19, indicate the presence of two primary components with eigenvalues greater than 1, which account for the majority of the variance in the dataset. The first component exhibits an eigenvalue of 4.11, accounting for 35.24% of the overall variance, whereas the second component shows an eigenvalue of 2.69, contributing 25.39% of the total variance. Together, these two components account for 61.18% of the overall variance, suggesting that a significant portion of the information in the dataset can be captured by these two primary dimensions. Next, the loading factor reflects the proportionate influence of each variable on each principal component. The first component reveals variables with the highest loading factor values, highlighting their significant influence on the overall structure of the data. This suggests that the variable plays a more prominent role in the data distribution pattern revealed by the PCA, as shown in Figure 20. Loading factor analysis indicates that in the first principal component (PCA1), the land cover and NDVI variables make significant positive contributions, whereas the slope variable makes the largest negative contribution. The first dimension in the PCA analysis primarily reflects the variability in land cover and vegetation levels, whereas slope serves as a factor that counteracts this variability. In the second principal component (PCA2), the DEM variable has a significant positive contribution, whereas the land cover variable has the largest negative contribution. The analysis indicates that the second dimension is closely associated with elevation variation, with alterations in land surface elevation serving as the primary factor in elucidating the data patterns, whereas variations in land cover play a contrasting role.



```
Training Data Dimensions: 30 x 30
Varians data per variabel: [1.73778222e-01 8.90550006e-31 7.69249877e-02 2.11328889e-02
 2.10305556e-02]
Jumlah komponen berdasarkan eigenvalue > 1: 2
Eigenvalue untuk setiap komponen:
Komponen 1: 1.4115648909178056
Komponen 2: 0.6298738552681254
Komponen 3: 0.9246256215530685
Komponen 4: 1.0383850204701313
Komponen 5: 0.0
Explained Variance Ratio: [0.35249912 0.25930781 0.23089956 0.1572935  0.        ]
Cumulative Explained Variance: [0.35249912 0.61180694 0.8427065  1.         1.        ]
Loading Factors:
Komponen 1: [ 4.10078966e-01  1.31984816e-16  6.15466363e-02 -7.62154748e-01
  8.11623241e-01]
Komponen 2: [-5.19871306e-01  7.07081467e-17  8.42314210e-01 -2.40681235e-01
 -2.71415250e-02]
Komponen 3: [ 7.34091871e-01 -1.06756221e-16  5.31400317e-01  2.89878772e-01
 -1.38991898e-01]
Komponen 4: [-1.54241821e-01  1.32168597e-16  7.37891128e-02  5.27554576e-01
  5.67736428e-01]
Komponen 5: [ 0.  0.  0. -0. -0.]
Threshold Dinamis: -5.031562215137337e-17
```

Fig. 19. Training dataset visualization.

The explained variance graph in Figure 21 shows that the two primary components (PCA1 and PCA2) together account for 61.18% of the total variance in the data, with individual contributions of 35.24% and 25.39%, respectively. This suggests that both components can effectively capture a substantial amount of the information present in the dataset, thereby serving as strong representatives of its underlying data structure. Figure 22 illustrates the values of PCA 1 and PCA 2 categorized into zones according to the first quartile (Q1) and the third quartile (Q3). The classification of these zones into "Low Zone" (green), "Average Zone" (yellow), and "High Zone" (red) facilitates the identification of data distribution patterns. Each data point is annotated with coordinates to indicate its relative position within the PCA space. The representation of the data distribution within the PCA space illustrates the grouping of data into three primary areas: zone 1 (green), zone 2 (yellow), and zone 3 (red). Zone 1 (green) indicates a low recharge area, zone 2 (yellow) indicates a moderate recharge area, and zone 3 (red) indicates a high recharge area.

Figure 23 shows the two primary components of the PCA results through a scatter plot that represents the data distribution based on the two principal dimensions, specifically the first and second components (PC1 and PC2). The scatter plot is generated by plotting each pixel of the processed data in a two-dimensional space, with PC1 as the X-axis and PC2 as the Y-axis. The scatter plot illustrates individual pixels from the map and facilitates observation of the data distribution patterns derived from the key results of the PCA. The observed distribution pattern suggests that PCA successfully identifies and distinguishes data according to recharge area characteristics that are predicted to be influenced by environmental parameters such as land cover, vegetation, slope, and elevation.

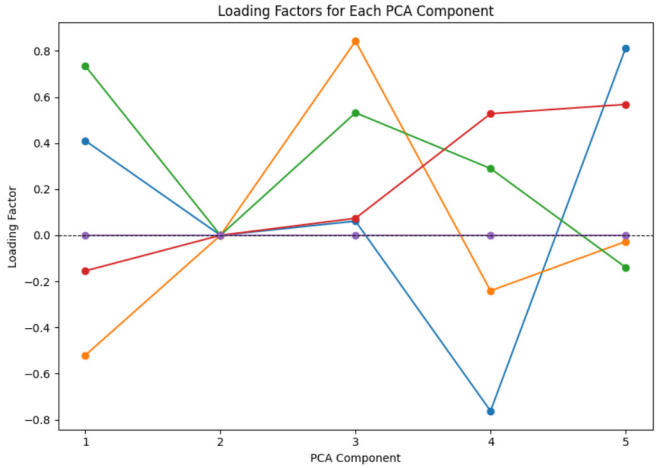


Fig. 20. Loading factors for the PCA components of the training dataset.

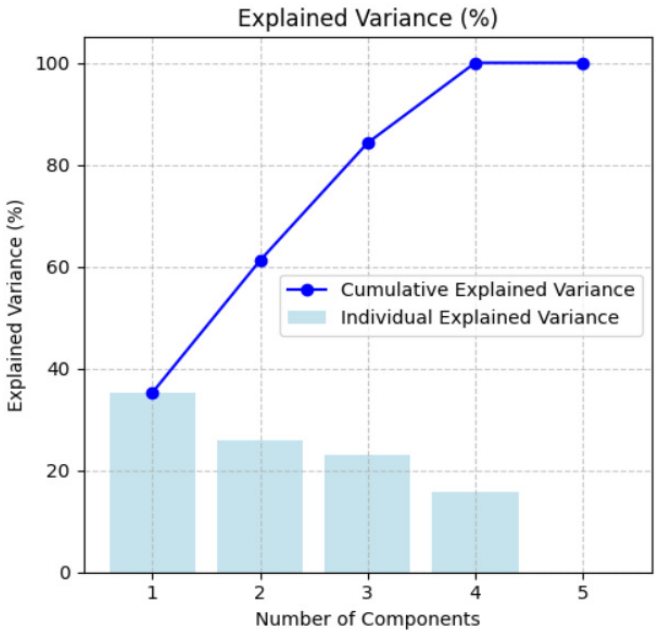


Fig. 21. Explained variance for the training dataset.

I. Testing Data

The testing dataset was obtained from the entire Singgahan and Montong regions, resulting in a testing dataset of 447×864 pixels, with each pixel measuring 30 m in size. The results are illustrated in Figure 24.

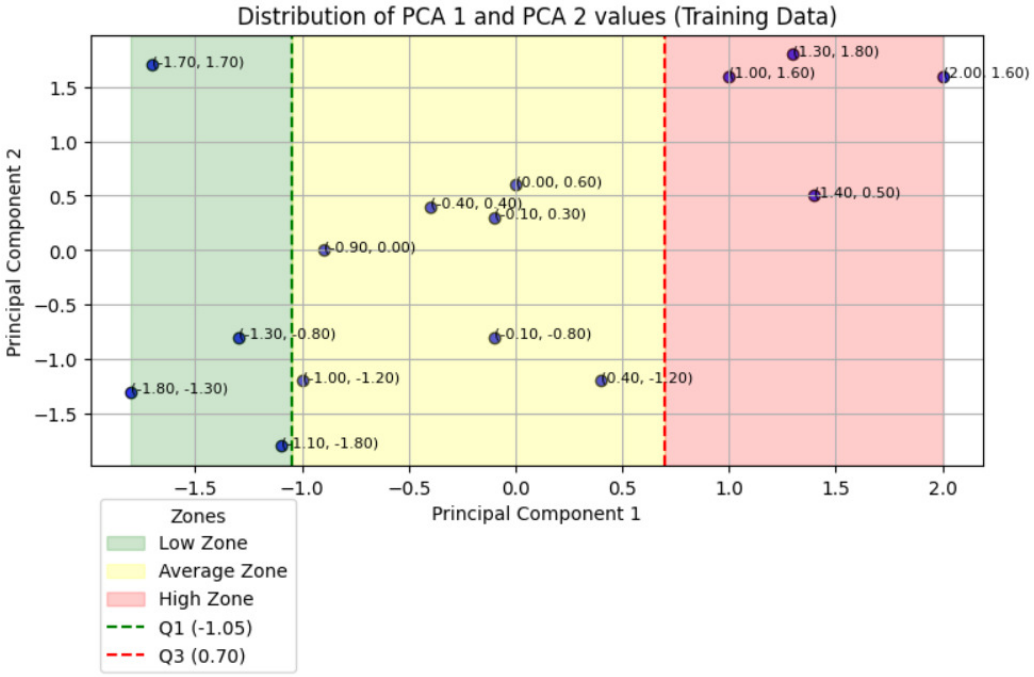


Fig. 22. PCA values distribution of the training dataset.

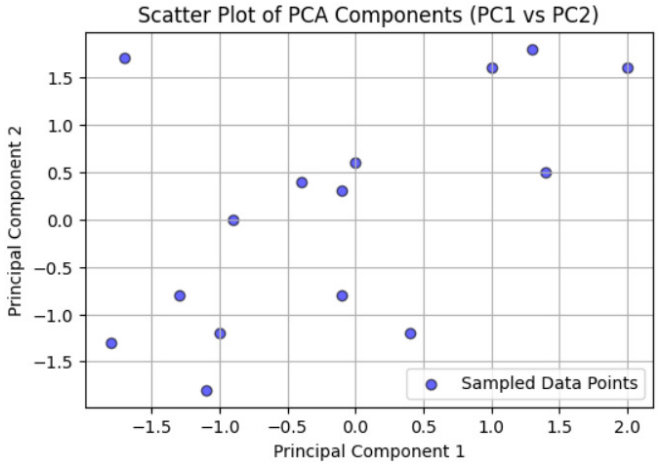


Fig. 23. Scatter plot of the PCA components (PC1 & PC2) of the training dataset.

The dataset was then normalized using StandardScaler to meet the standards required for PCA. After normalization, eigenvalues calculations were performed using a total of 5 components, yielding the following results: component 1: 2.3327; component 2: 0.3597; component 3: 0.4744; component 4: 0.9845; component 5: 0.8487. The PCA results indicate that there is a single principal component with an eigenvalue greater than 1 that accounts for most of the variance in the dataset. The first component has an eigenvalue of 2.3327, accounting for 46.65% of the overall variance. In total, these 5 main components represent 100% of the overall variance; however, it is primarily the first component that plays an important role in representing the information in the dataset. The loading factor results indicate that the first component contributes the largest amounts and has the highest weight on certain variables.

Next, the feature extraction was carried out by determining a dynamic threshold of -0.1721, which enabled the creation of a binary mask. In turn, a performance evaluation was conducted by calculating the Intersection over Union (IoU) score, which gave a value of 0.3306, reflecting the degree of agreement between the predictions and the actual results. The loading factor analysis shown in Figure 25 indicates that in the first principal component (PCA1), the land cover and DEM variables have significant positive contributions, whereas the NDVI variable has the most significant negative contribution. The first dimension in the PCA analysis represents changes in land cover and elevation, with vegetation levels used as an independent factor to these variations. In the second principal component (PCA2), the slope variable has a significant positive contribution, whereas the NDSI variable has the most crucial negative contribution. This suggests that the second dimension is closely associated with the terrain's slope, where variations in slope emerge as the primary factor in explaining the data pattern, with NDSI influencing it in the opposite direction.

The explained variance graph shown in Figure 26 illustrates that the two principal components (PCA1 and PCA2) collectively account for the majority of the variance in the dataset. PCA1 has the most significant contribution at 67%, whereas PCA2 follows with a contribution of 15%. The combined total of these two values is 82%. This means that these two primary components effectively illustrate the structure of the data.

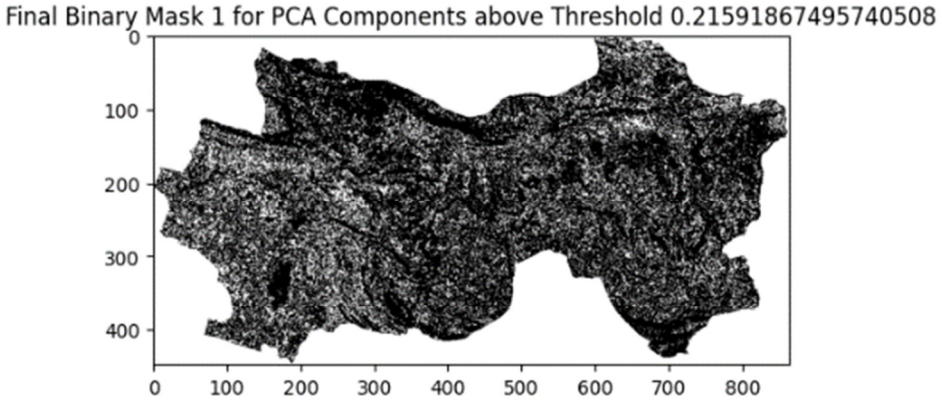


Fig. 24. Dataset testing area.

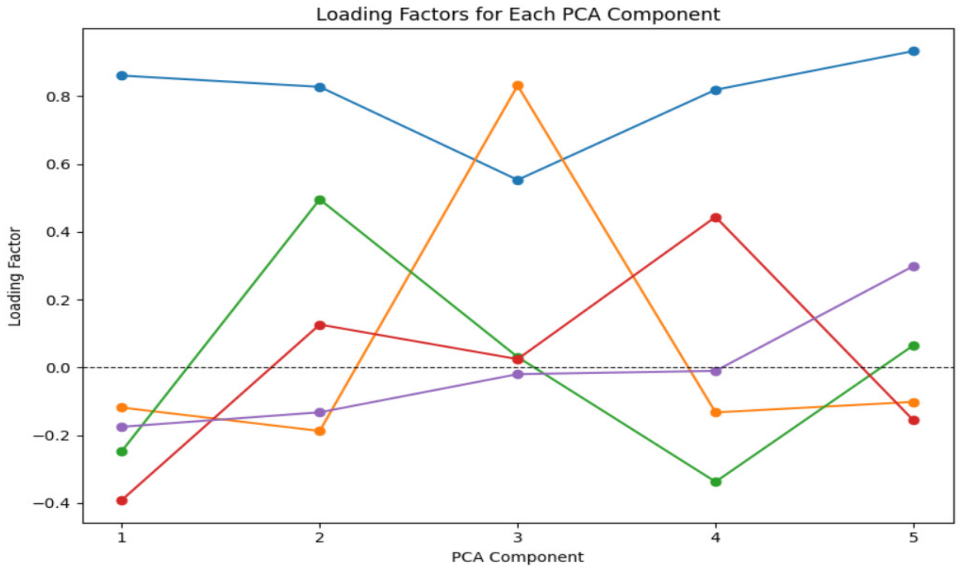


Fig. 25. Loading factor of the testing dataset.

Figures 27 and 28 present the two primary components of the PCA results through a scatter plot that represents the data distribution based on the two principal dimensions.

The input data used in this process are derived from the previous PCA results where spatial data surrounding various variables such as land cover, DEM, slope, NDVI, and NDSI were subjected to a weighting process. The following procedure is to perform PCA analysis on the data, which is designed to simplify complexity while preserving the maximum variance. The PCA process generates several components that represent linear combinations of the initial variables. Among the generated components, two primary components, specifically PC1 and PC2, have been selected for further analysis. The first principal component corresponds to the dimension exhibiting the highest variance, whereas the second principal component corresponds to the dimension with the next highest variance.

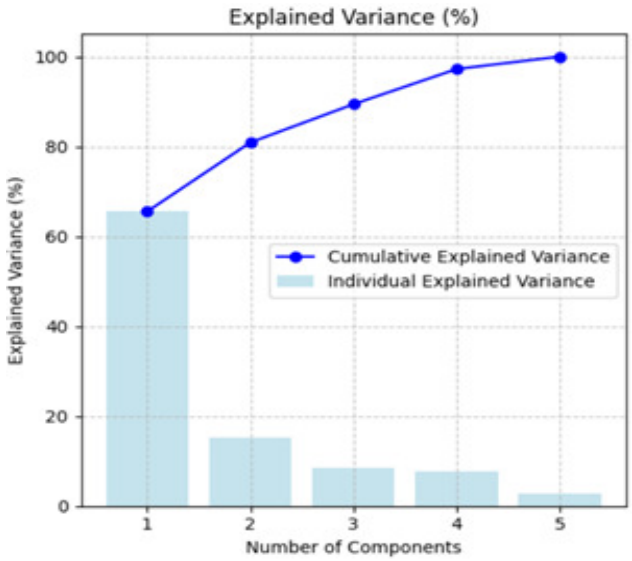


Fig. 26. Explained variance of the testing dataset.

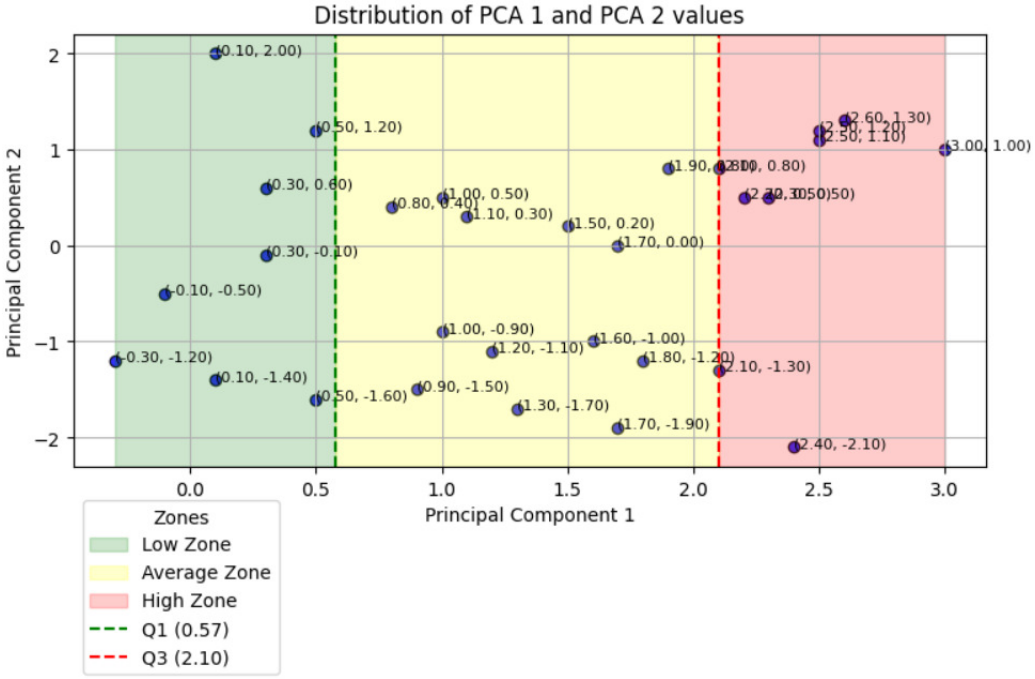


Fig. 27. PCA values distribution of the testing dataset.

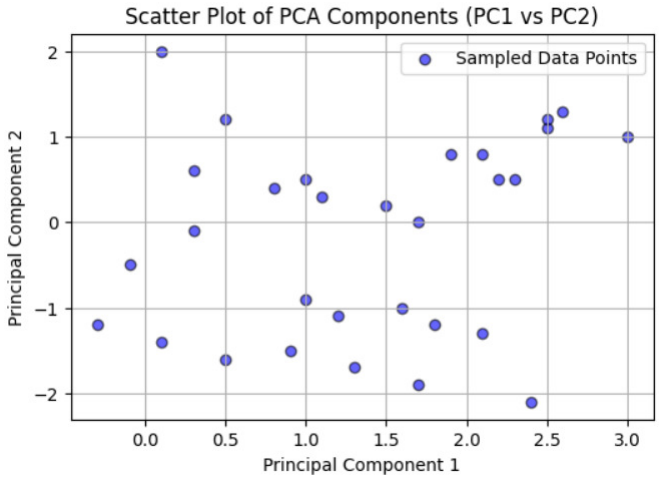


Fig. 28. Scatter plot of the PCA components (PC1 & PC2) of the testing dataset.

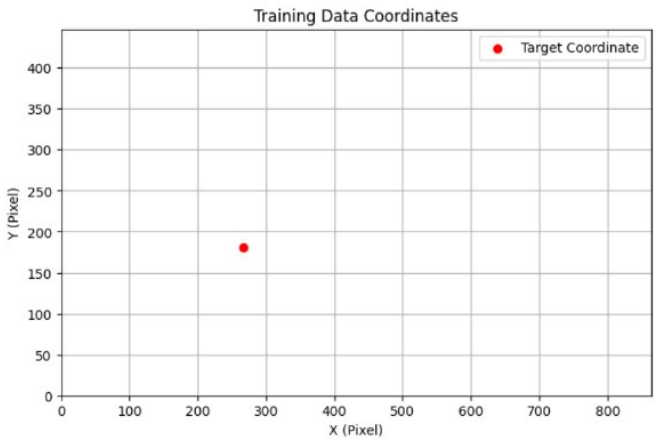


Fig. 29. Krawak spring area location.

J. Distribution of Recharge Area based on PCA Results

The recharge area distribution derived from the PCA results in the Mata Air Krawak area is illustrated in pixel coordinates in Figure 29. The resulting locations, plotted for the Singgahan and Montong areas, are shown in Figure 30, which illustrates an apparent correlation between high recharge areas, represented in red on the recharge map, and regions characterized by dense, healthy vegetation, shown in green.

This alignment is further supported by land use data indicating forests and plantations, as well as the presence of high topography suggesting the presence of Kambisol or Gleisol soil types in these areas. The region exhibiting significant recharge was identified to cover an area of 5.26 km². Low recharge areas (green) appear to be associated with residential and agricultural areas (scores 2 and 3 on the land classification map) characterized by low to medium vegetation levels. These areas are also located in regions with Mediterranean and Latosol soil types and are classified as having very low or extremely low topography.

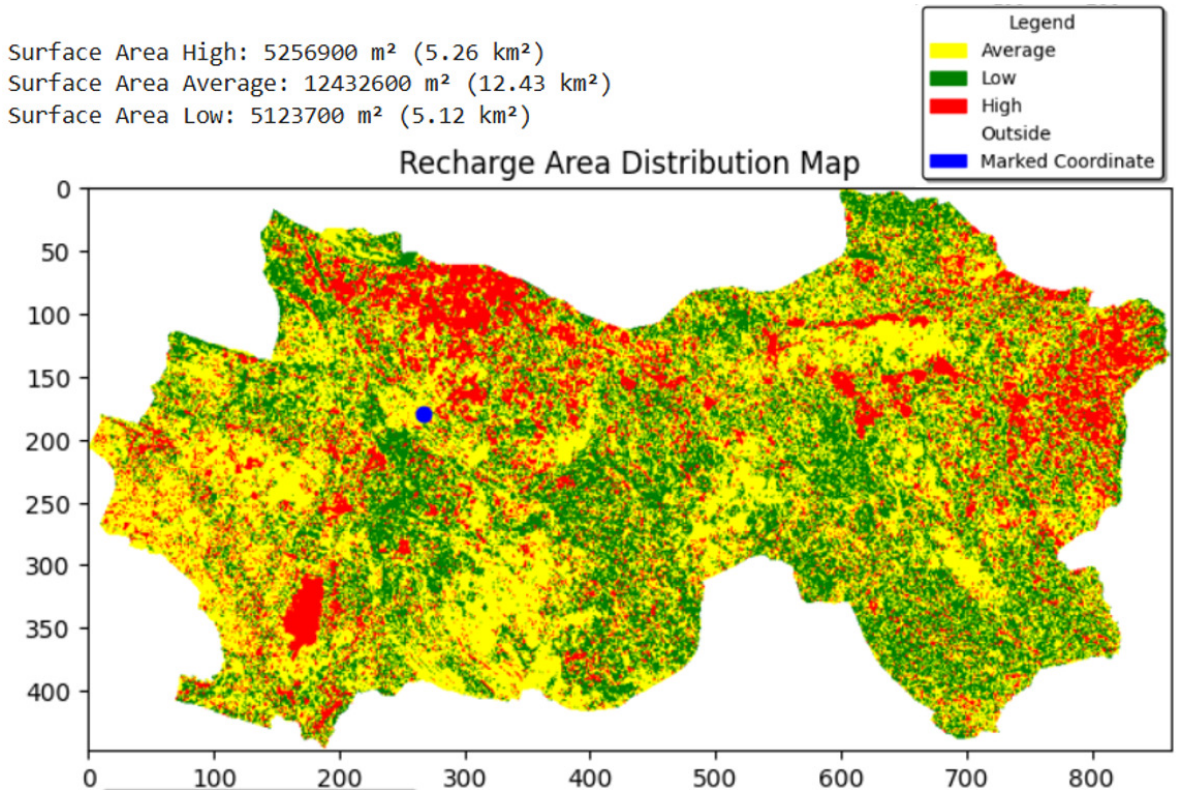


Fig. 30. Recharge area distribution map.

The region characterized by low recharge potential covers an area of 5.12 km². The areas still under consideration, specifically in Singgahan and Montong, exhibit a moderate recharge potential and cover a total area of 12.43 km²

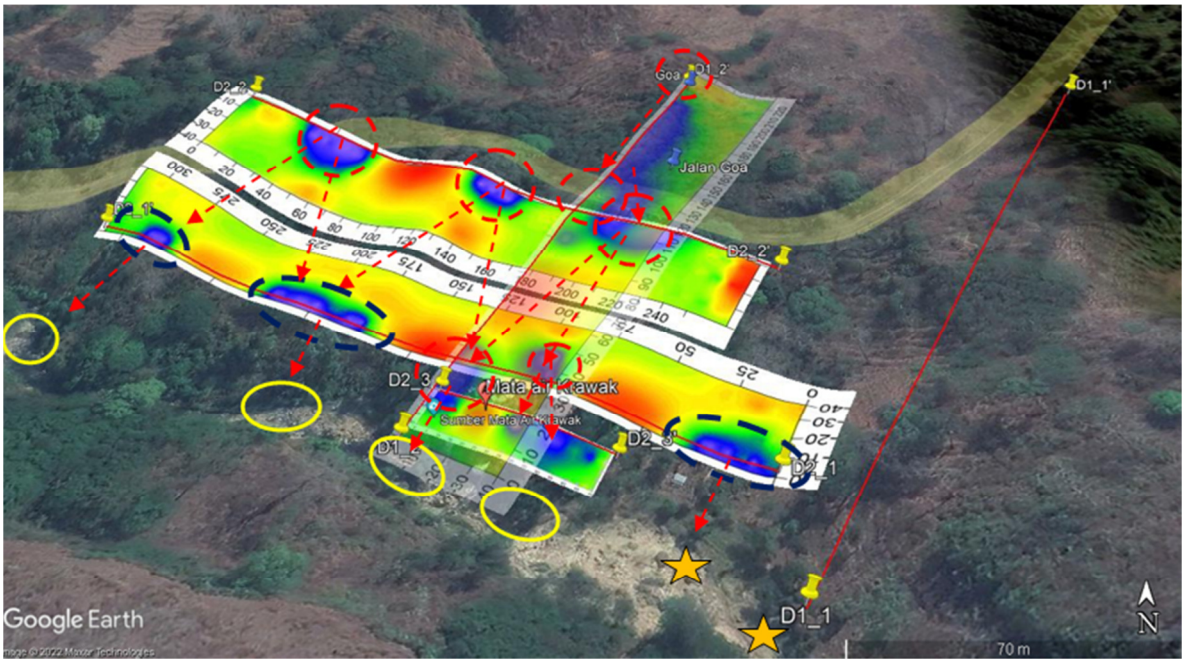
K. Discussion

Prior to this, a study was conducted to mapp recharge areas utilizing the VLF geophysical method and a scoring method incorporating a 4-component approach including rainfall, soil type, slope, and land cover in the areas of Singgahan and Montong [36, 37]. As shown in Figure 31, this research indicates that the flow direction of the Krawak spring below the surface has a northeast-southwest pattern. Information has been obtained indicating that the flow depth varies between 1 and 20 m. The result serves as an important guideline for exploring the distribution of potential recharge areas around the Krawak spring, particularly within the sub-districts of Singgahan and Montong. Regions with higher conductivity in the VLF results could suggest that these locations have the potential to be used as channels for groundwater infiltration. The formation of an aquifer system can be controlled by these conditions, which can be detected using VLF methods. The analysis of the water flow identified by VLF indicates an interaction with regions characterized by gentler slopes and properly permeable soil types in high recharge zones that promote the movement of water into the aquifer system. The VLF results serve as a training dataset due to the local scale of analysis of this method, which allows for a more detailed mapping of an area [36].

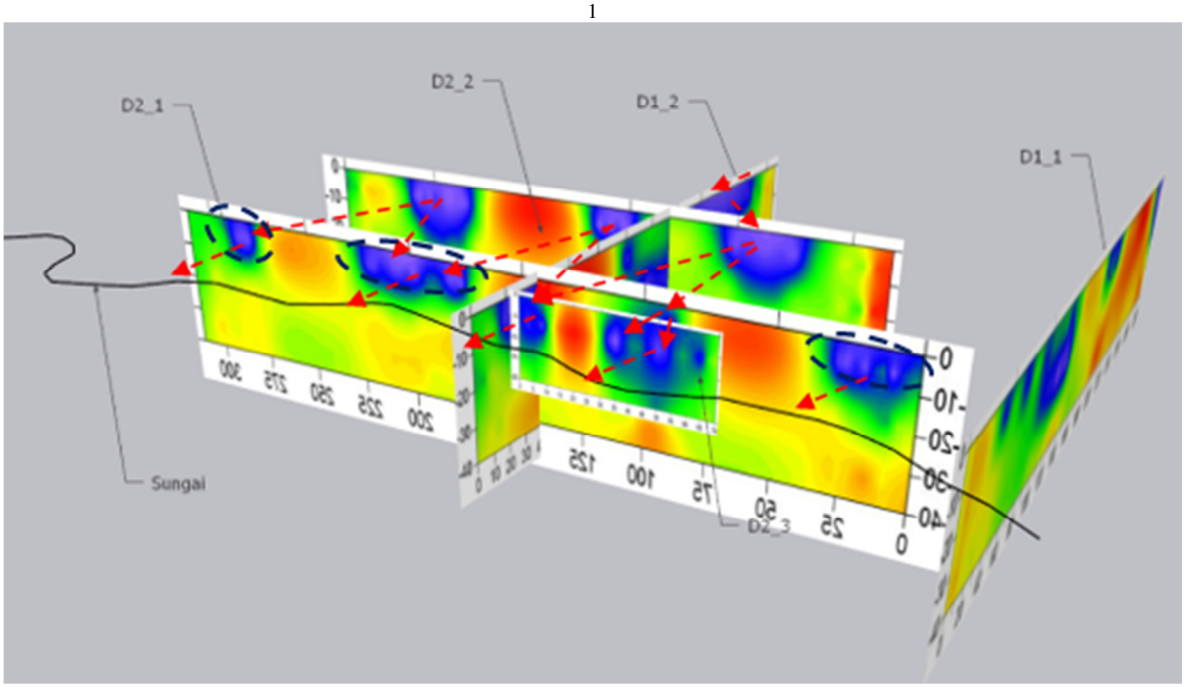
Figure 32 shows the results obtained through the manual scoring approach, whereas Figure 33 shows the distribution of recharge areas as analyzed by PCA. The distribution map uses red for high potential areas, yellow for moderate potential areas, and green for areas with low potential to develop into recharge zones. The distribution of potential recharge areas using the manual scoring method shows an obvious difference when compared to PCA. The differences are due to the different amounts of data and the underlying principles of the methods. The scoring method uses a deterministic method with only four parameters: rainfall, soil type, land cover, and slope. Meanwhile, the PCA method is an advanced process that can analyze interactions between data in a more complicated manner, using complex input parameters such as VLF field data, NDVI, NDSI, land cover, slope, and elevation/topography. However, the final result also depends on the quality of the training dataset used. The significant difference between the results of Figures 32 and 33 lies in the distribution of high recharge area potential. In Figure 32, this potential is predominantly concentrated in the southern region of the study area, whereas Figure 33 reveals a more balanced distribution of high recharge potential across the northern part of the study area. The results of the manual scoring method primarily indicate an average potential, whereas the PCA results reveal a more balanced distribution of low potential regions across the study area.

The blue point in Figure 33 indicates the site selected for VLF data acquisition to be used as training data. The results shown in Figure 31 indicate that the area of groundwater continuity is oriented in a northeast-southwest direction. The

results are consistent with the recharge area analysis performed by PCA as shown in Figure 33, which indicates that the high potential areas (highlighted in red) are predominantly located to the northeast and southwest relative to the VLF field data acquisition location. This is in contrast to the manual scoring results shown in Figure 32, which only reflect the concentration of high potential in the southwest region of the VLF field data collection site.



(a)



(b)

Fig. 31. Results of recharge area analysis based on data processing: (a) horizontal view, © Airbus, Maxar Technologies, (b) 2.5D visualization.

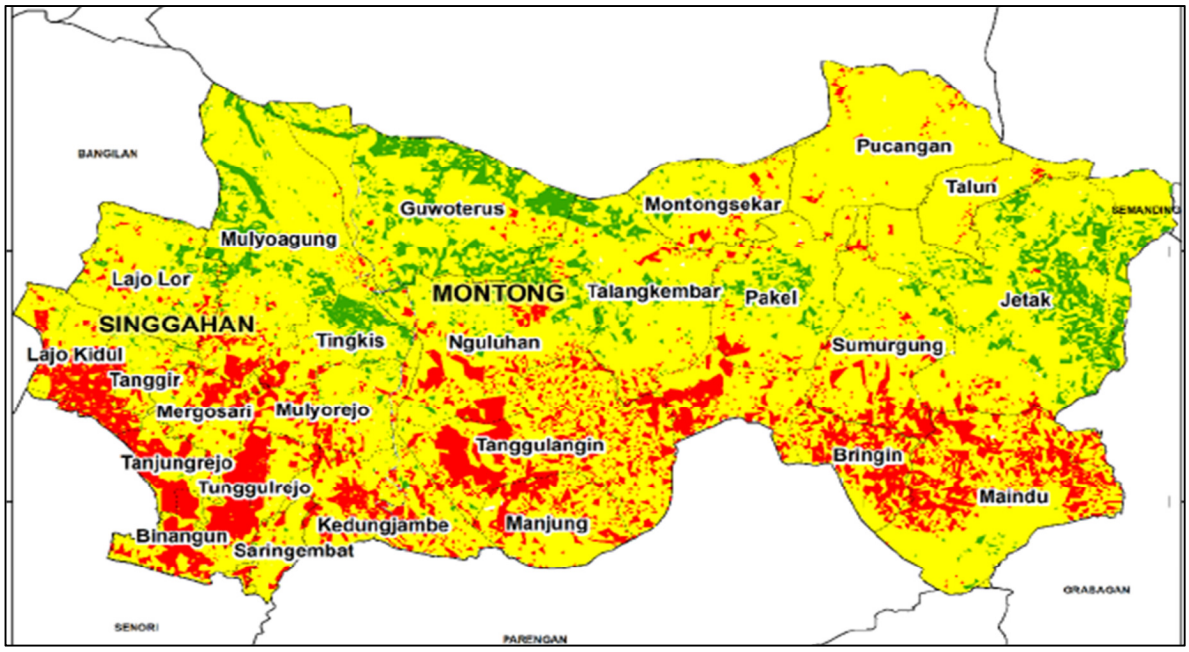


Fig. 32. Results of recharge area analysis based on the scoring method.

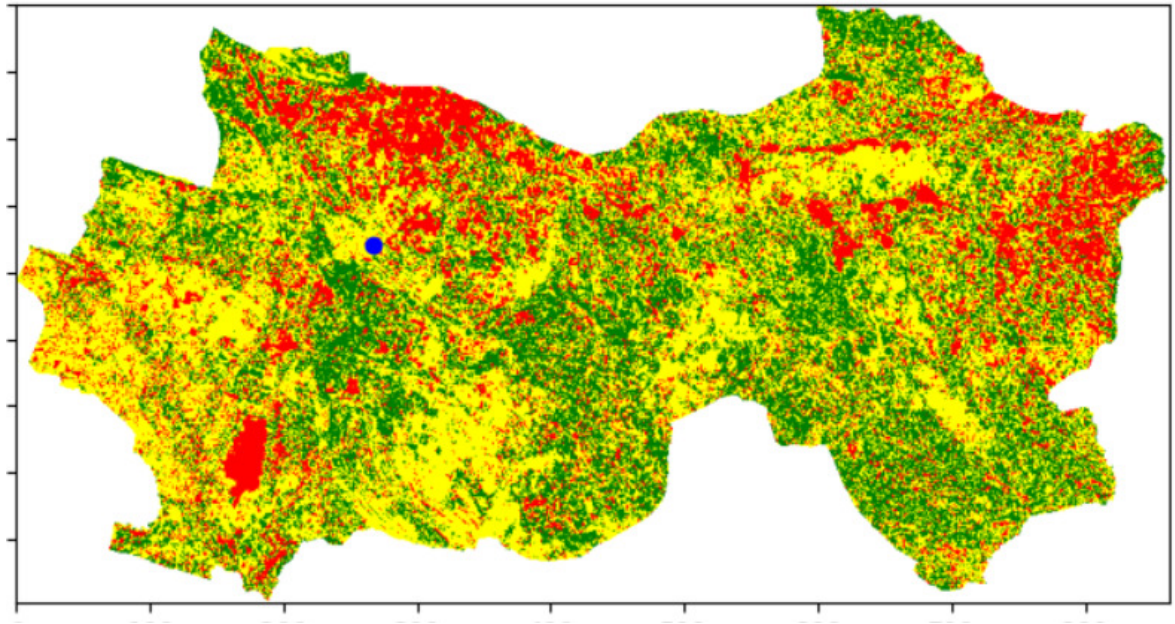


Fig. 33. Results of recharge area analysis based on PCA.

V. CONCLUSION

The analysis of groundwater recharge areas using the Principal Component Analysis (PCA) method reveals a distribution that closely matches the hydrologic conditions derived from Very Low Frequency (VLF) field data. The persistent correlation between the PCA results and the groundwater flow patterns identified by the VLF method underscores the effectiveness of this technique in defining recharge zones. This study uses VLF data as training data to identify recharge areas using PCA. The subsequent phase involves the development of a system that processes Landsat-8 satellite imagery as input. The image processing stage focuses on the analysis of five key parameters: Normalized Difference Vegetation Index (NDVI), land cover, Digital Elevation Model (DEM), land slope, and soil type. Each parameter is categorized based on pixel values, with NDVI, DEM, and slope relying on pixel color, whereas land cover segmentation incorporates both pixel color and pixel pattern analysis. Compared to conventional manual scoring methods, PCA proves to be more effective in capturing multivariate correlations between parameters, leading to improved accuracy

through the utilization of high eigenvalues. The evaluation of the testing dataset indicates that the primary contributing factors in this method include land cover, topography, and slope gradient, which together account for 82% of the total parameter influence in recharge area delineation. The validation procedure using the Intersection Over Union (IoU) score yielded a result of 0.3306, indicating an acceptable correlation between predicted recharge areas and actual field conditions. This highlights the reliability of the PCA method in providing a data-driven assessment of recharge zones. The novelty of the research lies in the combination of PCA with VLF geophysical validation, providing a more objective and data-driven approach than manual scoring methods. PCA's ability to concurrently evaluate many datasets, and its superior representation in identifying recharge areas, make it a suitable approach for future studies using unsupervised machine learning in hydrogeological investigations.

REFERENCES

[1] C. Asdak, *Hidrologi dan pengelolaan daerah aliran sungai*. Yogyakarta, Indonesia: Gadjah Mada University Press, 2002.

[2] M. Riastika, "Pengelolaan Air Tanah Berbasis Konservasi di Recharge Area Boyolali (Studi Kasus Recharge Area Cepogo, Boyolali, Jawa Tengah)," *Jurnal Ilmu Lingkungan*, vol. 9, no. 2, pp. 86–97, Oct. 2012, https://doi.org/10.14710/jil.9.2.86-97.

[3] S. Heri, "8 Kecamatan di Tuban Dilanda Kekeringan." Situs Resmi Pemerintah Kabupaten Tuban. https://tubankab.go.id/entry/8-kecamatan-di-tuban-dilanda-kekeringan.

[4] E. Maulana, "Analysis of Land Capability in Alluvial Plain and Volcanic Slope of Rembang District using Landforms Approach," in *2nd International Conference of Indonesian Society for Remote Sensing Remote Sensing for a Better Governance*, Yogyakarta, Indonesia, 2016, pp. 252–259.

[5] R. W. Van Bemmelen, *The Geology of Indonesia, vol. I.A. General Geology*. The Hague, Netherlands: Martinus Nyhoff, 1949.

[6] A. Muhartanto, D. S. Hidartan Djohor, S. Djohor, and N. Mukti, *Kawasan Karst Gunung Sewu & Potensinya*. Jakarta, Indonesia: FTKE Universitas Trisakti, 2007.

[7] N. Coppo, P.-A. Schnegg, M. Défago, and GSCB, "Mapping a shallow large cave using a high-resolution Very Low Frequency Electromagnetic Method," in *8th Conference on Limestone Hydrogeology*, Neuchâtel, Switzerland, 2006, pp. 71–74.

[8] M. S. Putri, "Aplikasi Filter NA-MEMD pada Data VLF-EM untuk Mengidentifikasi Kemenerusan Sungai Bawah Permukaan (Studi Kasus Desa Sekar, Pacitan)," B.S. thesis, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2020.

[9] F. S. Wardhana, "Identifikasi Kemenerusan Sistem Sungai Bawah Permukaan Kawasan Karst Dersono Pacitan dengan Metode VLF-EM," B.S. thesis, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2019.

[10] R. M. Sampurno and A. Thoriq, "Klasifikasi Tutupan Lahan Menggunakan Citra Landsat 8 Operational Land Imager (Oli) di Kabupaten Sumedang," *Teknotan: Jurnal Industri Teknologi Pertanian*, vol. 10, no. 2, pp. 61–70, Nov. 2016.

[11] N. S. Magesh, N. Chandrasekar, and J. P. Soundranayagam, "Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing, GIS and MIF techniques," *Geoscience Frontiers*, vol. 3, no. 2, pp. 189–196, Mar. 2012, https://doi.org/10.1016/j.gsf.2011.10.007.

[12] N. F. Che Nordin, N. S. Mohd, S. Koting, Z. Ismail, M. Sherif, and A. El-Shafie, "Groundwater quality forecasting modelling using artificial intelligence: A review," *Groundwater for Sustainable Development*, vol. 14, Aug. 2021, Art. no. 100643, https://doi.org/10.1016/j.gsd.2021.100643.

[13] H. R. Pourghasemi, N. Sadhasivam, S. Yousefi, S. Tavangar, H. Ghaffari Nazarlou, and M. Santosh, "Using machine learning algorithms to map the groundwater recharge potential zones," *Journal of Environmental Management*, vol. 265, Jul. 2020, Art. no. 110525, https://doi.org/10.1016/j.jenvman.2020.110525.

[14] G. Martinsen *et al.*, "Developing a pan-European high-resolution groundwater recharge map – Combining satellite data and national survey data using machine learning," *Science of The Total Environment*, vol. 822, May 2022, Art. no. 153464, https://doi.org/10.1016/j.scitotenv.2022.153464.

[15] S. Abdelaziz, M. I. Gad, and A. H. M. H. El Tahan, "Groundwater quality index based on PCA: Wadi El-Natrun, Egypt," *Journal of African Earth Sciences*, vol. 172, Dec. 2020, Art. no. 103964, https://doi.org/10.1016/j.jafrearsci.2020.103964.

[16] J. B. F. Moreira, S. H. Yuwanto, and E. Mahardjo, "Pemetaan Geologi dan Penentuan Lingkungan Pengendapan Batugamping Berdasarkan Analisis Petrografis di Kecamatan Semanding dan Sekitarnya Kabupaten Tuban Provinsi Jawa Timur," *Prosiding Seminar Teknologi Kebumian dan Kelautan*, vol. 1, no. 1, pp. 154–163, Sep. 2019, https://doi.org/10.31284/j.semitan.2019.834.

[17] R. Faizal, S. Sismanto, R. Handayani, and A. Asta, "Pendugaan Aliran Sungai Bawah Tanah Dalam Pemenuhan Kebutuhan Air Masyarakat Desa Hargosari Gunungkidul Berdasarkan Data VLF-EM Terkoreksi Topografi," *Borneo Engineering : Jurnal Teknik Sipil*, vol. 1, no. 2, pp. 44–53, Dec. 2017, https://doi.org/10.35334/be.v1i2.601.

[18] T. W. Githinji, E. W. Dindi, Z. N. Kuria, and D. O. Olago, "Application of analytical hierarchy process and integrated fuzzy-analytical hierarchy process for mapping potential groundwater recharge zone using GIS in the arid areas of Ewaso Ng'iro – Lagh Dera Basin, Kenya," *HydroResearch*, vol. 5, pp. 22–34, Jan. 2022, https://doi.org/10.1016/j.hydres.2021.11.001.

[19] M. K. Villareal and A. F. Tongco, "Remote Sensing Techniques for Classification and Mapping of Sugarcane Growth," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 6041–6046, Aug. 2020, https://doi.org/10.48084/etasr.3694.

[20] M. Makonyo and M. M. Msabi, "Identification of groundwater potential recharge zones using GIS-based multi-criteria decision analysis: A case study of semi-arid midlands Manyara fractured aquifer, North-Eastern Tanzania," *Remote Sensing Applications: Society and Environment*, vol. 23, Aug. 2021, Art. no. 100544, https://doi.org/10.1016/j.rsase.2021.100544.

[21] M. V. Japitana and M. E. C. Burce, "A Satellite-based Remote Sensing Technique for Surface Water Quality Estimation," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 3965–3970, Apr. 2019, https://doi.org/10.48084/etasr.2664.

[22] I. Batioua, R. Benouini, K. Zenkouar, and A. Zahi, "Image classification using separable invariants moments based on Racah polynomials," *Procedia Computer Science*, vol. 127, pp. 320–327, 2018, https://doi.org/10.1016/j.procs.2018.01.128.

[23] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Feature extraction for hyperspectral image classification," in *2017 IEEE Region 10 Humanitarian Technology Conference*, Dhaka, Bangladesh, 2017, pp. 379–382, https://doi.org/10.1109/R10-HTC.2017.8288979.

[24] E. B. Troccoli, A. G. Cerqueira, J. B. Lemos, and M. Holz, "K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis," *Journal of Applied Geophysics*, vol. 198, Mar. 2022, Art. no. 104555, https://doi.org/10.1016/j.jappgeo.2022.104555.

[25] Y. Liu and L. Wu, "Geological Disaster Recognition on Optical Remote Sensing Images Using Deep Learning," *Procedia Computer Science*, vol. 91, pp. 566–575, 2016, https://doi.org/10.1016/j.procs.2016.07.144.

[26] W. F. Hendria, Q. T. Phan, F. Adzaka, and C. Jeong, "Combining transformer and CNN for object detection in UAV imagery," *ICT Express*, vol. 9, no. 2, pp. 258–263, Apr. 2023, https://doi.org/10.1016/j.icte.2021.12.006.

[27] P. More and P. Mishra, "Enhanced-PCA based Dimensionality Reduction and Feature Selection for Real-Time Network Threat Detection," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6270–6275, Oct. 2020, https://doi.org/10.48084/etasr.3801.

[28] J. A. Telaumbanua, C. Prasetyadi, and A. Subandrio, "Geologi dan Studi Lingkungan Pengendapan Formasi Ngrayong Daerah Mulyoagung dan Sekitarnya, Kecamatan Singgahan, Kabupaten Tuban, Provinsi Jawa Timur," *Jurnal Ilmiah Geologi Pangea*, vol. 3, no. 1, pp. 39–49, Jun. 2016.

[29] M. S. Purwanto *et al.*, "Penentuan Recharge Area Pada Kabupaten Tanah Datar Menggunakan Citra Landsat 8 dan Sistem Informasi Geografis (SIG)," *Jurnal Geosaintek*, vol. 8, no. 3, pp. 242–249, Dec. 2022, https://doi.org/10.12962/j25023659.v8i3.14615.

[30] E. Parizi, S. M. Hosseini, B. Ataie-Ashtiani, and C. T. Simmons, "Normalized difference vegetation index as the dominant predicting factor of groundwater recharge in phreatic aquifers: case studies across Iran," *Scientific Reports*, vol. 10, no. 1, Oct. 2020, Art. no. 17473, https://doi.org/10.1038/s41598-020-74561-4.

[31] T. Murtono, A. M. Imran, and M. A. Thaha, "Zonasi Imbuhan Air Tanah Pada Daerah Aliran Sungai Lahumbuti Provinsi Sulawesi Tenggara," *Geosains*, vol. 9, no. 2, pp. 89–98, 2013.

[32] A. N. Kholis and M. I. Rendra, "Potensi Resapan Air Tanah di Kabupaten Bojonegoro Dengan Pendekatan GIS," *RADIAL: JuRnal PerADaban SaIns RekAyasan dan TeknoLogi*, vol. 10, no. 2, pp. 222–233, Dec. 2022.

[33] M. Khoiri, L. M. Jaelani, and A. Widodo, "Landslides Hazard Mapping Using Remote Sensing Data in Ponorogo Regency, East Java," *Internet Journal of Society for Social Management Systems*, vol. 11, no. 2, pp. 100–109, Jul. 2018.

[34] S. Amanah, "Pengaruh Kerusakan Hutan Lindung Krawak Terhadap Produktivitas Pertanian di Kecamatan Singgahan Kabupaten Tuban," *Swara Bhumi*, vol. 2, no. 1, pp. 126–134, May 2014.

[35] T. Lillesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*, 4th ed. Hoboken, NJ, USA: Wiley, 1999.

[36] M. S. Purwanto, A. Susilo, A. S. Bahri, A. Naba, U. I. Sari, and A. T. W. Almais, "Mapping Underground River Flows in karst Areas with the VLF-EM Method (Case Study of the Krawak Region, Singgahan Tuban)," *IOP Conference Series: Earth and Environmental Science*, vol. 1307, no. 1, Feb. 2024, Art. no. 012005, https://doi.org/10.1088/1755-1315/1307/1/012005.

[37] S. Purwanto *et al.*, "Analysis and Mapping of the Distribution of Groundwater Recharge Areas Using the Scoring Method (Case Study: Singgahan and Montong District, Tuban).," *IOP Conference Series: Earth and Environmental Science*, vol. 1418, no. 1, Dec. 2024, Art. no. 012057, https://doi.org/10.1088/1755-1315/1418/1/012057.

AUTHORS PROFILES

Moh. Singgih Purwanto earned his bachelor's degree from the Department of Mathematics at the Sepuluh Nopember Institute of Technology in Surabaya, Indonesia, in 2005. He pursued further education and earned a master's degree in Civil Engineering, specializing in Remote Sensing Systems, from the Sepuluh Nopember Institute of Technology in 2009. At present, he is a Ph.D. student in the Department of Physics, concentrating on Geophysics. He additionally holds a position as a lecturer in the Department of Geophysical Engineering at the Sepuluh Nopember Institute of Technology in Surabaya. His areas of focus encompass Geographic Information Systems, Databases, and disaster management.

Adi Susilo earned his bachelor's degree from the Department of Geophysics at Gadjah Mada University in Yogyakarta, Indonesia, in 1989. He earned his master's degree from the Department of Geophysics at Gadjah Mada University in 1997, followed by a Ph.D. from James Cook University in Australia in 2004. At present, he holds the position of Professor in the area of Disaster and Natural Resource Exploration within the Department of Physics at Brawijaya University. His areas of focus encompass disasters and the exploration of natural resources, geosciences, groundwater studies, mangrove ecosystems, disaster management strategies, geophysical analysis, and geological investigations.

Agus Naba earned his bachelor's degree from the Department of Physics at Brawijaya University in Malang, Indonesia, in 1995. He pursued further education and earned a master's degree from the Bandung Institute of Technology in 2000, followed by a Doctorate (Dr.Eng.) from the University of Tsukuba in 2007. At present, he holds the positions of Lecturer and Professor in Intelligent Science within the Department of Physics at Brawijaya University. His areas of focus encompass intelligent systems and signal processing.

Ayi Syaeful Bahri earned a bachelor's degree from the Department of Science at Padjadjaran University in Indonesia in 1994. He pursued further education and earned a Master's degree from the Bandung Institute of Technology in 2003, followed by a Doctorate from the same institution in 2010. His area of expertise lies in applied geophysics, with a focus on utilizing geophysical methods including VLF, magnetic techniques, and the exploration of water resources, as well as groundwater management.

Siti Navisa Currently studying for earning a bachelor's degree in Geophysical Engineering from the Sepuluh Nopember Institute of Technology in Surabaya. The current activities are related to seismology, hydrology, and utilizing satellite data for primary studies.