IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 29 March 2023, accepted 4 May 2023, date of publication 12 May 2023, date of current version 25 July 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3275852*

RESEARCH ARTICLE

# Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters

**AGUNG TEGUH WIBOWO ALMAIS[1,2], ADI SUSILO[1], AGUS NABA[1], MOECHAMMAD SAROSA[3], CAHYO CRYSDIAN[2], IMAM TAZI[2], MOKHAMAD AMIN HARIYADI[2], MUHAMMAD AZIZ MUSLIM[4], PUSPA MILADIN NURAIDA SAFITRI ABDUL BASID[2], YUNIFA MIFTACHUL ARIF[2], MOHAMMAD SINGGIH PURWANTO[5], DIYAN PARWATININGTYAS[6], SUPRIYONO[2], AND HENDRO WICAKSONO[7]**

[1]Department of Physics, Universitas Brawijaya Malang, Malang 65145, Indonesia
[2]Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang 65144, Indonesia
[3]Department of Electrical Engineering, State Polytechnic of Malang, Malang 65141, Indonesia
[4]Department of Electrical Engineering, Universitas Brawijaya, Malang 65145, Indonesia
[5]Department of Geophysical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[6]Department of Physics, Universitas Pertahanan Republik Indonesia, Bogor 10430, Indonesia
[7]School of Business, Social and Decision Science, Constructor University, 28759 Bremen, Germany

Corresponding author: Adi Susilo (adisusilo@ub.ac.id)

This work was supported by the International Collaborative Development Research in 2022, under Contract 1293.

**ABSTRACT** Post-disaster sector damage data is data that has features or criteria in each case the level of damage to the post-natural disaster sector data. These criteria data are building conditions, building structures, building physicals, building functions, and other supporting conditions. Data on the level of damage to the post-natural disaster sector used in this study amounted to 216 data, each of which has 5 criteria for damage to the post-natural disaster sector. Then PCA is used to look for labels in each data. The results of these labels will be used to cluster data based on the value scale of the results of data normalization in the PCA process. In the data normalization process at PCA, the data is divided into 2 components, namely PC1 and PC2. Each component has a variance ratio and eigenvalue generated in the PCA process. For PC1 it has a variance ratio of 85.17% and an eigenvalue of 4.28%, while PC2 has a variance ratio of 9.36% and an eigenvalue of 0.47%. The results of data normalization are then made into a 2-dimensional graph to see the data visualization of the results of each main component (PC). The result is that there is 3 data cluster using a value scale based on the PCA results chart. The coordinate value (n) of each cluster is cluster 1 ($n<0$), cluster 2 ($0 \leq n < 2$), and cluster 3 ($n \geq 2$). To test these 3 groups of data, it is necessary to conduct trials by comparing the original target data, there are two experiments, namely testing the PC1 results based on the original target data, and the PC2 results based on the original target data. The result is that there are 2 updates, the first is that the distribution of PC1 data is very good when comparing the distribution of data with PC2 in grouping data, because the eigenvalue of PC1 is greater than that of PC2. While second, the results of testing the PC1 data with the original target data produce good data grouping, because the original target data which has a value of 1 (slightly damaged) occupies the coordinates of group 1 ($n < 0$), the original target data which has a value of 2 (moderately damaged) occupies group 2 coordinates ($0 \leq n < 2$), and for the original target data the value 3 (heavily damaged) occupies group 3 coordinates ($n \geq 2$). Therefore, it can be concluded that PCA, which so far has been used by many studies as feature reduction, this study uses PCA for labeling unsupervised data so that it has appropriate data labels for further processing.

**INDEX TERMS** Post-disaster sector damage, PCA, labels, cluster, unsupervised.

## I. INTRODUCTION

Natural disasters are a big problem for countries with tropical climates, such as Indonesia [1], [2], [3]. This country

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

is in an area flanked by oceans, has many volcanoes, and large rivers that cause potential natural disasters [1], [4]. The State of Indonesia has an agency responsible for managing information on natural disasters in Indonesia, namely Badan Nasional Penanggulangan Bencana (BNPB) [5]. BNPB during 2021 has recorded 3,092 natural disasters [6] including landslides [7], floods [8], earthquakes [3], [9], tsunamis [10], and volcanic eruptions [11]. The natural disaster had a significant impact on damage, besides that the disaster also caused damage to all sectors including infrastructure [12], settlement or housing [13], economy or property [14], and local sectors such as religion, education, and social [15].

Determining the level of damage to the sector after a natural disaster involves surveyors such as volunteers, village government, and local community members who have been assigned by the local government [16], to an area affected by a natural disaster. The surveyors must have sufficient provisions in determining the level of damage to the natural disaster sector, for example, the criteria for determining sector damage, in their research [17] there are several criteria for determining the level of damage to the post-natural disaster sector including the condition of the building, the state of the building structure, the physical condition of the building, the function of the building, and other supporting conditions. Almais et al. apply existing criteria in research [17] and then combined them with the Decision Support System (DSS) method to determine the level of damage to the sector after natural disasters [18]. Results of post-disaster sector damage levels in research [16], [17], [18], [19] refer to the level of damage that usually occurs in the field, namely light damage, moderate damage, and heavy damage.

Scientifically there has been no discussion of the reference to the type of level of damage to the post-natural disaster sector, whereas according to government regulations it already exists but is not in numerical form. So it is necessary to research how to label data based on the distribution of data that usually occurs in the field so that the label for the level of damage to the sector used is appropriate and by the conditions of the sector in the field. Because the determination of the label for the level of damage to the sector after a natural disaster is related to the distribution of aid. To determine the numerical value of the label for the level of damage to the post-natural disaster sector, namely light damage, moderate damage, and heavy damage requires a study that uses data division that can produce numerical values. One technique for dividing data to find the numerical value of each label can use the Principal Component Analysis (PCA) technique.

Data clustering technique to determine labels from the numerical value results of the PCA technique which uses 216 data on damage to the post-natural disaster sector that has gone through an analysis process and has a value on each criterion. Then PCA processes the 216 data to produce numerical values based on the results of the PCA graphs, the results of these numerical values are clustered to determine the type label for the level of damage to the sector after natural disasters. In geophysics [20] using PCA is to classify seismic facies by automatically labeling and sizing them so that seismic maps can be grouped and presented neatly and nicely. According to *Uddin et al*. PCA is a technique for reducing the dimensions of a data set, increasing interpretability, and minimizing the loss of information in data [21]. In addition, using PCA can reduce image complexity and execution time [22]. PCA in addition to grouping a set of data into a normal data group, in research [23], [24], [25] uses PCA to fault rate of an object.

In our research, we apply PCA to label a level of damage to the post-natural disaster sector using the Python programming language and use post-natural disaster sector damage data that has already gone through an analysis process. Python processes post-disaster sector damage data by applying the PCA process to produce a numeric value that has a meaning that can produce a labeling of data.

This research offers several contributions as follows:

1. Make labels on unsupervised data sets using PCA.
2. Using PCA can make a range or distance between certain values that produce a range of values in a case or object.
3. Label the results of the PCA to scientifically determine the type of damage to the sector after a natural disaster.

The contents of this article explain the following: background on using PCA to create labels and the use of PCA in several studies in ''Related Work''. Then ''Method and Data Preparation'' explains the steps of PCA and data acquisition. The ''results and discussion'', explains the PCA experimental process using the Python programming language to create labels for the level of damage to the sector after natural disasters and validation of the results of labeling data. Finally, the ''conclusion'' summarizes the results of the experiment and opportunities for future research.

## II. RELATED WORK

There are several conceptual references to using PCA and implementing clustering to pre-process data, as shown in TABLE 1. According to Cao et al. PCA is a good data preprocessing technique for mapping high-dimensional data to low-dimensional spaces while maintaining the characteristics of the data. [26]. Apart from being able to map data, according to Ahsan et al. PCA can monitor mixed characteristics (attributes and variables) in a product with a non-linear relationship that results in different values for the attributes and variables in the product [27], [28]. According to Liang et al. PCA is an unsupervised learning method, practically if there are many variables and correlations between variables and a lot of overlapping information then PCA is one way to reduce redundant information and increase the accuracy of calculating a data model [29]. In addition, PCA can map and divide data sets, according to Ueda, PCA can be used to estimate the particle size and shape characteristics of 3D images using 2D image input. [30]. In their research, Ferreira et al. use two methods to label data on grass types in agriculture, these methods Deep Representations and Image Clusters (JULE)

**TABLE 1.** Related work for PCA and clustering.

| References | Topic | Method | Subject |
|---|---|---|---|
| [20] | Clustering | K-Means Clustering and PCA | Grouping high-dimensional data into a low-dimensional space |
| [26] | Mapping data | PCA and SVM | Mapping high-dimensional data into a low-dimensional space. |
| [27], [28] | Separate data | KPCA (Kernel Principal Component Analysis) | Separating the differences in the value of attributes and variables on the product. |
| [29] | Reduce information | PCA and SVM | Eliminate overlapping and redundant information in the data. |
| [30] | Estimating data | PCA | Estimating the size and shape characteristics of the particles. |
| [31] | Label data | JULE and DeepCluster | Labeling of data types of grass in agriculture. |
| [32] | Sharing data | Clustering | Distribute the questions to respondents. |
| [33] | Extract data | Probabilistic PCA | Extract information on abundant, non-linear, and dynamic data. |
| [34] | Characterizing information | GWPCA (Geographically Weighted Principal Component Analysis) | Characterizing heavy metals to determine the potential of soil. |
| [35] | Modeling data changes | PCA | Evaluate quantitatively the form of anatomical changes in an object. |
| Ours | Clustering and Labeling data | PCA and Clustering | Clustering and labeling data on the level of damage to the post-natural disaster sector. |

and Deep Clustering for Unsupervised Learning of Visual Features (DeepCluster) [31].

Hafida et al. explained that the clustering technique is a technique for sharing data that can be used to share questions for respondents, totaling 364 school students around areas prone to natural disasters, the eruption of Mount Merapi in Indonesia [32]. According to Zhang et al. PCA has the disadvantage that it takes a long time to extract data that is abundant, non-linear, and dynamic. So PCA requires an additional probabilistic function to solve the problem [33]. Aidoo et al. explained that to find out the characteristics of heavy metals to be able to find out potential information on the soil to modify PCA to become GWPCA (Geographically Weighted Principal Component Analysis) [34]. Argota-Perez et al. also explained that PCA can model changes in data and evaluate data quantitatively [35].

Previous studies have implemented PCA by using a framework for data reduction, extracting information or data, dividing data, separating data, grouping data, and mapping data. However, labeling a complex post-natural disaster sector damage data set needs to be done so that the data can be used in the next process and can be used as a reference in further research. These two things need to be done because the label for the type of level of damage to the sector after a natural disaster needs a scientific reference. Bachriwindi et al.explained to determine the level of damage to the sector after a natural disaster using labels 1 (slightly damaged), 2 (moderately damaged), and 3 (severely damaged) [17]. However, this research [17] references the level of damage to the post-disaster sector using references from the 2021 BNPB rules [4], scientifically there has been no research that discusses determining the level of damage to the sector after natural disasters. Therefore, this research seeks to create a data label for the type of damage level in the post-natural disaster sector using PCA and clustering. On the other hand, we are also trying to make the results of PCA and clustering labels on the type of level of damage to the sector after natural disasters that can be used as a reference by future researchers and the government.

## III. PROPOSED METHOD AND DATA PREPARATION

In our method, it explains two sub-chapters, namely the proposed method and data preparation. The first sub-chapter, namely the proposed method, explains the steps of PCA. While the second sub-chapter is data preparation which explains how to obtain data, variable data, and contents of data on damage to the post-natural disaster sector.

### A. PCA FOR DATA LABELS

PCA is a statistical method that is often used in high-dimensional data analysis, dimensionality reduction, noise filtering, and feature selection [20], [36].

According to *Deisenroth et al.* that PCA can project the original dataset to a new database with lower dimensions [37]. Projecting data into lower dimensions is a main

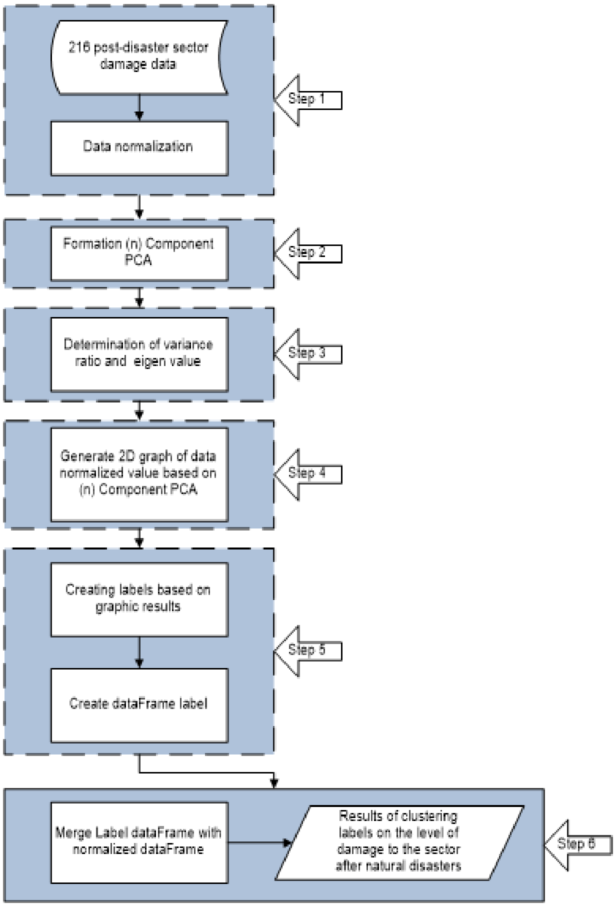


**FIGURE 1. Our main architecture of the whole paper.**

component (PC), to present as much information as possible beforehand [38].

Using PCA for clustering data on damage to the sector after natural disasters, then the clustering results produce data labels according to the value of the results from PCA. In Figure 1 it can be seen that after creating a 2D graph of data from normalized values based on (n) PCA components, then creating labels from the graph in Figure 1 and creating a dataFrame from the labels that have been successfully created. The PCA steps used are as follows:

Step 1: Setting up data.
Using data from surveyors in assessing the level of damage to the post-disaster sector that occurred in East Java Province, Indonesia. The data on damage to buildings or sectors after a natural disaster used is data that has been analyzed from the surveyor's data, which totals 216 data.

Step 2: Normalize the data.
The normalization process is one of the processes to make data standard so that it is by PCA standards. According to *Susilo et al*, the PCA data standard is that the data used must have the same degree or value and be balanced for each data [39]. In his research [40] there are 6 methods for standardizing data, namely Normalization (NR), Standardscale (SS), MinMax (MM), MaxAbs (MA), Robust Scale (RS), and Quantile Transformer (QT). This study uses the Standard scale (SS) because the type of data on damage to buildings after natural disasters uses a standard scale of 1/2/3. The equation of the Standard scale (SS) uses the following equation (1) [41]:

$$\chi_{standart} = \chi - mean(\chi) \Big/ standartdeviation(\chi) \quad (1)$$

The standard deviation uses the equation (2) as follows [41]:

$$\bar{X} = \sum\nolimits_{i=1}^{n} \chi_i \Big/ n \quad (2)$$

Symbol $\bar{X}$ (X bar) describes the average value of the set X.

Step 3: Determine the variance ratio and eigenvalue values.
The variance ratio is a measure of the spread of data. The equation for determining the variance ratio can use the following equation (3):

$$s^2 = \sum\nolimits_{i=1}^{n} (\chi_i - \bar{\chi})^2 \Big/ (n-1) \quad (3)$$

Equation (3) has the understanding that the distribution of data has a certain size that can determine the amount of data distribution.
While the eigenvalue is a value that occupies a place in the eigenvector in the form of a matrix [41].

Step 4: Determine the number of PC.
In his research [42] explained that determining the number of principal components in PCA is a very important process because it can affect the level of accuracy of a data set. Besides that, determining the right number of main components will get the most optimal eigenvalues and variance ratios [43].

Step 5: Creating the visualization.
Visualization is a very important thing in representing a result [44], especially on PCA results. By using 3-dimensional (3D) graphics to represent a range of values whose results describe the results in 2-dimensional (2D) images.

Step 6: Create a range of values.
According to Lambers et al. value range can be done by maximizing the highest value in a data set and minimizing the loss of information caused by data reduction [45]. In our research, to create a range of values using the results from the PC.

Step 7: Clustering data.
Clustering is a learning technique that divides data into several parts of unsupervised data into several homogeneous data groups [46], [47]. In our research for clustering data using PC coordinate point data based on a predetermined range of values. The results of clustering data require validation to know the level of truth.

Step 8: Result and Validation.

The result is a value that has gone through a certain process, whereas to know the level of truth of a result it is necessary to carry out a validation [48], [49]. Validation is a key role in determining whether the results obtained are by existing requirements [50]. Following are the steps for validating the results:

- Visualize 2D and 3D of PCA result data, namely PC1 and PC2 data.
- Visualize 2D of original target data results that have passed surveyor validation.
- Visualize 2D PC1 results with original target data.
- Visualize 2D PC2 results with original target data.
- Comparing the 2D visualization results from PC1 based on the original target data with PC2 based on the original target.

Based on the visualization trial steps, this study validates the results using a comparison of the PC data clustering data with the original target data which is correlated to get the smallest error value.

The steps above can be implemented in a complete computational procedure using the following algorithm:

### B. DATA PREPARATION

The data used were 216 data from the analysis of damage to the post-natural disaster sector in the study. To find out the details of the data used can be seen in TABLE 2.

The data in TABLE 2 comes from the results of the analysis of damage data after natural disasters in the province of East Java which does not yet have a label. The label is the final result of the level of damage to the sector after a natural disaster which describes the condition of a sector. TABLE 2 is the head() data or the first 5 data out of 216 data that will be processed using PCA. The data uses 216 cases that have 5 criteria, namely building conditions, building structures, building functions, and other support conditions. Each of the criteria in TABLE 3 has a value of 1/2/3, this value has an understanding that refers to research [17] which is as shown in TABLE 3.

## IV. RESULT AND DISCUSSION

The results and discussion explain the process of labeling data with PCA and testing the results with data from experts or surveyors to prove that PCA can be a wrong technique for labeling data.

### A. DATA STANDARDIZATION

Standardize the data contained in TABLE 2 so that the data has the same weight when forming a PC. The results from the PC produce a new value that contains the value from the normalization results that are in the PCA process. Normalization uses the StandardScaler function which produces values like Figure 2.

Figure 2 is the result of the normalization process for post-natural disaster sector damage data whose results will later be used for the process of determining the variance ratio and eigenvalue values. In the data standardization process, it produces a value that is different from the original value, in TABLE 2 it has a value of 1,2,3 for each data but the normalization results have a value that is smaller than the original value.

**Algorithm 1 Clustering Using PCA for Labeling Data**

**Input**: `Input data using functions` ***`pd.read_csv('filename.CSV)`***

**Process**: •`Data normalization using functions` ***`StandardScaler`***() and ***`fit_transform`***()
• `Generate variance ratio using functions` ***`explained_variance_ratio_`***()
• `Generate eigenvalues using functions` ***`explained_variance`***()
• `Generate PCA components using functions` ***`PCA`***()
• `Generate data frame resulting from PCA components using functions` ***`pd.DataFrame`***()
• `Visualization of normalization results using` ***`scatterplot`***()
• `Generate labels using branching (if else) and branching (for)`

**Output**: • `The results of combining the dataFrame results of the PCA components with the results of using labels` ***`concat`***()
•`The results of visualization of` • `data labeling results using PCA with training data validation results from surveyors using` ***`scatterplot`***()

**TABLE 2.** Data on damage to the sector after natural disasters.

| Case | Building Condition | Building Structure | Building Physical | Building Function | Other Supporting Conditions |
|---|---|---|---|---|---|
| **0** | 1 | 2 | 1 | 1 | 2 |
| **1** | 1 | 2 | 1 | 1 | 1 |
| **2** | 3 | 3 | 3 | 3 | 2 |
| **3** | 1 | 1 | 1 | 1 | 2 |
| **4** | 2 | 2 | 1 | 2 | 2 |

**TABLE 3.** Level of damage to buildings after natural disasters.

| Value | Description |
|---|---|
| 1 | Light Damage |
| 2 | Moderate Damage |
| 3 | Heavy Damage |

```
[[-0.77326675  0.31671066 -0.78219001 -0.59520603 -0.304095  ]
 [-0.77326675  0.31671066 -0.78219001 -0.59520603 -2.18079556]
 [ 1.56275587  1.51687736  2.93106365  1.70058866 -0.304095  ]
 ...
 [-0.77326675  0.31671066 -0.78219001 -0.59520603 -2.18079556]
 [-0.77326675 -0.88345604 -0.78219001 -0.59520603 -0.304095  ]
 [-0.77326675  0.31671066 -0.78219001 -0.59520603 -2.18079556]]
```

**FIGURE 2.** Results of standardization data.

### B. VALUE OF VARIANCE RATIO, EIGENVALUE, AND FACTOR LOADING VARIABLE

Determining the variance ratio and eigenvalue values can use equation (3) by applying an existing function to the clustering algorithm using PCA for labeling data by implementing it in the Python programming language. In this research, the eigenvalues and variance ratios for each feature are shown in TABLE 4 which explains the 5 features after being reduced using PCA and dividing them into 5 main components which produce the values of the explained variance ratio, the cumulative explained variance ratio, and the factor loading variable which affects each PC and varies for each feature.

Visually, the variance ratio value for each PC in TABLE 4 there is also in Figure 3, which illustrates the increase in the variance ratio value for each PC.

### C. NUMBER OF PC

To determine the number of PC, look at the eigenvalue and variance ratio. If you look at TABLE 4, the highest eigenvalue is found in main component 1 (PC1) with an eigenvalue of 85.16 and a variance ratio of 85.16%. However, the highest eigenvalue cannot be used as a reference in this study because the purpose is to label the data, not to determine the most dominant feature. So in addition to using PC1 which has the highest eigenvalue we also use principal component 2 (PC2) which has a lower eigenvalue than PC1 which is 9.35 and the variance ratio value of PC2 is 94.52%.

The use of PC2 as a comparison to PC1 is to determine the accuracy of the coordinates between each of these PCs and the original target data, whereby the results of these coordinates will produce a form of data labeling. Figure 4 shows the variance ratio values for PC1 and PC2. Figure 4 explains that the variance ratio values for PC1 and PC2 form a straight line, which means that the PC2 data has wider and more widely distributed data when compared to the data distribution for PC1.

**TABLE 4.** Eigen value and variance ratio.

| Number of Components | Eigenvalue | Variance ratio (%) | Factor Loading Variable |
|---|---|---|---|
| 1.0 | 85.16 | 85.16 | 1. Building Condition<br>2. Building Function<br>3. Building Structure<br>4. Building Physical<br>5. Other Supporting Conditions |
| 2.0 | 9.35 | 94.52 | 1. Other Supporting Conditions<br>2. Building Physical<br>3. Building Structure<br>4. Building Function<br>5. Building Condition |
| 3.0 | 4.56 | 99.09 | 1. Building Physical<br>2. Building Function<br>3. Building Structure<br>4. Other Supporting Conditions<br>5. Building Condition |
| 4.0 | 0.73 | 99.82 | 1. Building Function<br>2. Building Structure<br>3. Building Physical<br>4. Other Supporting Conditions<br>5. Building Condition |
| 5.0 | 0.17 | 100.0 | 1. Building Condition<br>2. Building Structure<br>3. Building Physical<br>4. Other Supporting Conditions<br>5. Building Function |

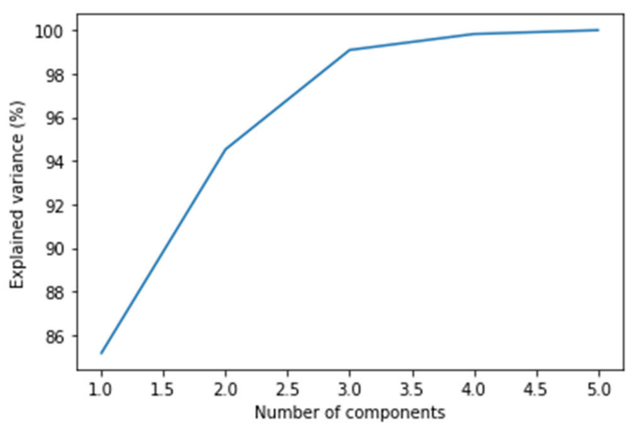


**FIGURE 3.** Variance ration chart.

### D. MAIN COMPONENT VISUALIZATION

Visualizing the main components is a process to find out how the data is distributed in each PC1 and PC2. The visualization can be seen in Figure 5.

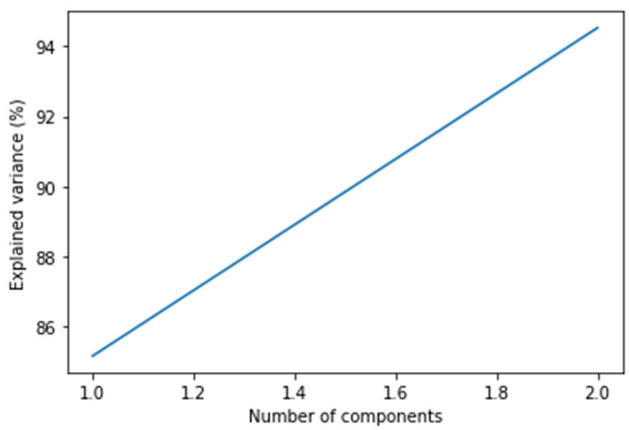


**FIGURE 4.** Graph of PC1 and PC2 variance ration value.

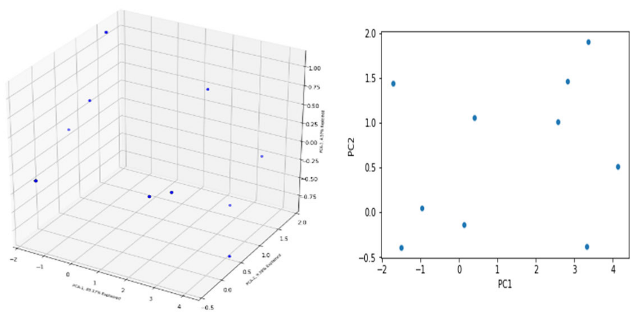

**FIGURE 5.** 3D and 2D visualization of data distribution on main components.

Based on Figure 5, explains that the distribution of data on PC1 and PC2 collects or clusters at the same points. This means that the points on the lines PC1 and PC2 occupy the same position, causing the distribution of 216 data to be invisible but only 10 data to be seen.

### *E. DATA CLUSTERING PROCESS*

Based on Figure 6, we get results that give rise to innovations, namely the results of the PC1 and PC2 data sharing which have gone through the PCA process to produce a data distribution that can form a data set that can be labeled based on the coordinate points of the data distribution. Figure 6 explains the distribution of the data generated by the coordinate points of PC1 to PC2.

For the value of the coordinate points of each data distribution generated by the coordinate points of PC1 to PC2, they are divided into 3 groups as shown in Figure 7.

Figure 7 illustrates that there are 10 points spread across 3 groups of different coordinate values. TABLE 5 explains the location of the coordinate points in each group of coordinate values.

Based on the results of the visualization in Figure 7 and the elaboration of the coordinate points in TABLE 5, it can be further detailed that the data is grouped into 3 parts, namely the green, orange, and red parts. The green part is data that has a coordinate value range less than 0, while the orange color is data that has a coordinate value range between 0 and 2, and the red color is data that has a coordinate value range greater than 2.

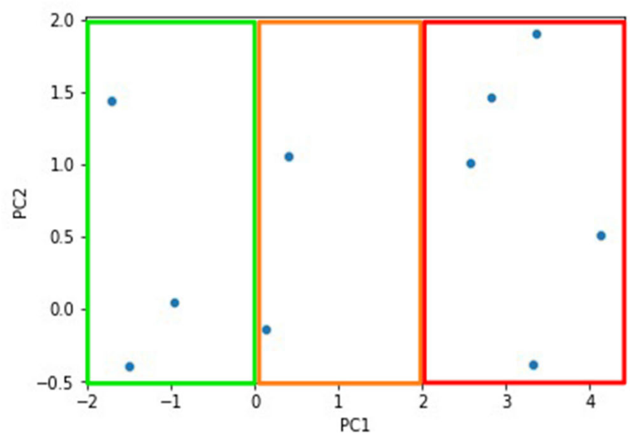


**FIGURE 6.** Distribution of PC1 and PC2 data.

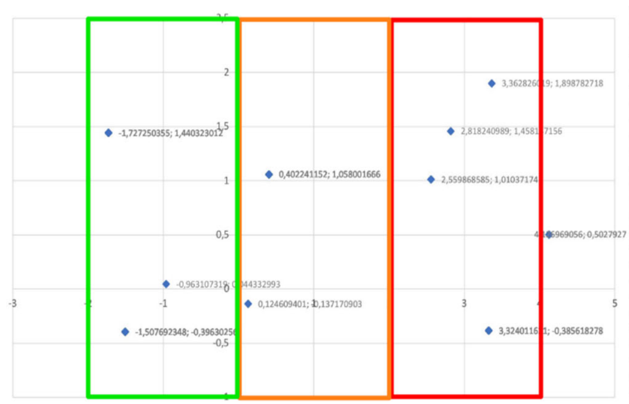

**FIGURE 7.** Coordinate Points of PC1 and PC2 data distribution.

**TABLE 5.** Point and coordinate value PC1 PC2.

| Color Clustering | Coordinate Point | | Coordinate Value Range (n) |
|---|---|---|---|
| | PC1 | PC2 | |
| Green | -1.507692348 | -0.396302569 | n < 0 |
| | -0.963107319 | -0.044332993 | |
| | -1.727250355 | 1.440323012 | |
| Orange | 0.124609401 | 0.137170903 | 0 ≤ n < 2 |
| | 0.402241152 | 1.058001666 | |
| Red | 3.324011621 | -0.385618278 | n ≥ 2 |
| | 4.126969056 | 0.5027927 | |
| | 2.559868585 | 1.010371741 | |
| **Color Clustering** | **Coordinate Point** | | **Coordinate Value Range (n)** |
| | **PC1** | **PC2** | |
| | 2.818240989 | 1.458147156 | |
| | 3.362826019 | 1.898782718 | |

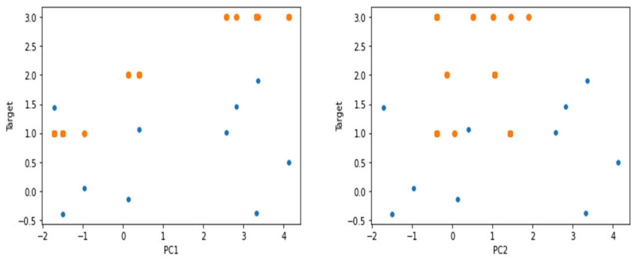

**FIGURE 8.** Results of Distribution of PC1 and PC2 Data based on original target data.

### F. VALIDATION OF DATA CLUSTERING RESULT

Testing the validation of data clustering results by comparing the results of the original target data from the 216 data used with the results of the distribution of data generated using PCA. The results of the distribution of PCA produce 2 main components, namely PC1 and PC2. For PC1 and PC2, when compared with the results of the original target data, the data distribution is shown in Figure 8.

The data distribution in Figure 8 explains the comparison of data distribution between PCI and the original data target and PC2 and the original data target, for the blue coordinates are the values of PC1 or PC2 while the orange coordinates are the values of the original data targets. Figure 8 explains that the PC1 results when compared with the original data target results produce a very good data distribution because when the target value is equal to 1 (Slight Damage) it is at coordinates between 0-(-2) whereas when the target value is equal to 2 (Moderately Damaged) is at coordinates 0-2 and when the target value is equal to 3 (Severely Damaged) it is at the coordinates above equal to 2.

Whereas PC2 data distribution when compared with the original target data cannot be as good as using PC1 because PC1 and PC2 have eigenvalues and variance ratios which cause the data distribution to also be different. Figure 9 shows a comparison of the results of the distribution of the original target data with the distribution of the results of PC1 data.

Figure 9 describes a comparison of the original target data with PCI which results in a data clustering into 3 parts, namely:

1) Cluster 1: The green color contains target data which has an original data-target value of 1.0.
2) Cluster 2: The orange color has target data which has an original data-target value of 2.0.
3) Cluster 3: The green color has target data which has an original data-target value of 3.0.

To explain the clustering data in Figure 9 it is in TABLE 6.

TABLE 6 explains that the name of the label for the level of damage is based on the results of the original data target, that is, if the original data-target value is 1.0 then it is slightly damaged, if it is 2.0 then it is moderately damaged and if it is 3.0 then it is heavily damaged. So when the value of the coordinates of PC1 (n) is in the range of values $n < 0$ then the label of PC1 data results is slightly damaged, whereas if

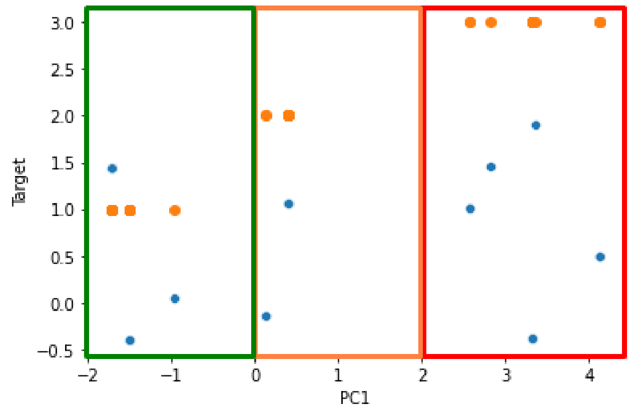

**FIGURE 9.** Results of PC1 Data clustering based on original data targets.

**TABLE 6.** PC1 comparison result data label based on original target data.

| Comparison Result | | Coordinate Value Range | Label |
|---|---|---|---|
| PC1 | Target Original Data | (n) | |
| -1.507692348 | 1.0 | $n < 0$ | Light Damage |
| -0.963107319 | 1.0 | | |
| -1.727250355 | 1.0 | | |
| 0.124609401 | 2.0 | $0 \leq n < 2$ | Moderate Damage |
| 0.402241152 | 2.0 | | |
| 3.324011621 | 3.0 | $n \geq 2$ | Heavily Damaged |
| 4.126969056 | 3.0 | | |
| 2.559868585 | 3.0 | | |
| 2.818240989 | 3.0 | | |
| 3.362826019 | 3.0 | | |

the coordinate value of PC1 (n) is a range of values $0 \leq n < 2$ then the label of the PC2 data results is moderately damaged, and for the coordinate values of PC1 (n) the range is the range value $n \geq 2$ then the label data results PC1 is badly damaged.

Based on the experiments that have been carried out, the results of labeling data using PCA are more efficient when compared to the research by *Troccoli et al.* which groups seismic facies by automatically labeling and measuring them, because PCA can use unsupervised data, while Troccoli et al. must use supervised data [20]. To measure the results of labeling data with PCA, you can look at TABLE 6 which explains the distribution of the data as follows:

1) When the original data-target value = 1.0, the data is grouped at the coordinate point (n) with a range of coordinate values $n < 0$, then get the data is labeled as slightly damaged. Because in the original target data a value of 1 = slightly damaged.
2) Whereas when the original data-target value = 2.0, the data is grouped at the coordinate point (n) with a range of coordinate values $0 \leq n < 2$, then get the data is labeled as moderately corrupted. Because the original target data for a value of 2 = moderately damaged.

3) When the original data-target value = 3.0, the data is grouped at the coordinate point (n) with a range of coordinate values $n \geq 2$, then get the data is labeled as heavily corrupted. Because the original target data is 3 = heavily damaged.

## V. CONCLUSION

Based on this research, it can be concluded that 2 things, according to researchers, are novelties, namely making labels of a data set into good data and easy for machine learning or deep learning processes to use PCA techniques. In previous studies, PCA is used to reduce the features of an image or data that has many features so it doesn't take long to process the data further. In addition, PCA has a unique value which is usually used to determine the number of main components, this value is the eigenvalue. The highest eigenvalue is the value attached to the best main component for determining data labels. This has been proven by this study, that PC1 has a higher eigenvalue than PC2, namely 85.16, while PC2 has an eigenvalue of 9.35. In addition, the value of the variance ratio of PC2 when compared to PC1 has increased by 9.36%, which means that the greater the variance ratio on each PC, the distribution of data on each PC becomes random and wide. Therefore, for labeling data on the level of damage to the post-disaster sector, PC1 data distribution is used, so that the data distribution is not random and wide. So that it can produce clustering data based on the distribution of the original target data with PC1 data. The result is that there are 3 clustering data, namely the original target data which has a value of 1 (slightly damaged) is grouped at the coordinate point of PC1 (n) with a range of values $n < 0$, the original data-target value which has a value of 2 (moderately damaged) clusters at the coordinate points of PC1 (n) with a range of values $0 \leq n < 2$ and the original data-target value of 3 (heavily damaged) is clustered at the coordinate point PC1 (n) with a range of values $n \geq 2$. So that going forward to label an unsupervised data set can use the Principal Component Analysis (PCA) technique.

## ACKNOWLEDGMENT

The authors would like to thank the research partners for their contributions which indirectly helped carry out this research, especially the University of Brawijaya, Malang, which has provided facilities to explore ideas and the process of forming this analysis. So that we can find out the labeling data updates using PCA to determine the level of damage to the sector after natural disasters and the Database Laboratory, Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, provides facilities for research to produce a publication.

## REFERENCES

[1] J. Widagdo, D. D. Putra, B. Syihabuddin, T. Juhana, E. Mulyana, and A. Munir, "Android-based disaster management application for after-disaster rapid mobile assessment," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IoTaIS)*, Jan. 2021, pp. 201–204, doi: 10.1109/IoTaIS50849.2021.9359695.

[2] T. Yudistira, J. Metaxian, M. Putriastuti, S. Widiyantoro, N. Rawlinson, F. Beauducel, Z. Zulfakriza, A. D. Nugraha, A. Laurin, A. A. Fahmi, and A. Budi-Santoso, "Imaging of a magma system beneath the Merapi Volcano complex, Indonesia, using ambient seismic noise tomography," *Geophys. J. Int.*, vol. 226, no. 1, pp. 511–523, Mar. 2021, doi: 10.1093/gji/ggab104.

[3] P. Supendi, A. D. Nugraha, S. Widiyantoro, J. D. Pesicek, C. H. Thurber, C. I. Abdullah, D. Daryono, S. H. Wiyono, H. A. Shiddiqi, and S. Rosalia, "Relocated aftershocks and background seismicity in eastern Indonesia shed light on the 2018 Lombok and Palu earthquake sequences," *Geophys. J. Int.*, vol. 221, no. 1, pp. 1845–1855, Dec. 2019, doi: 10.1093/gji/ggaa118.

[4] F. A. Maulana, *Fungsi Koordinasi Badan Penanggulangan Bencana Daerah (BPBD) Sidoarjo Dalam Penanggulangan Bencana Di Kabupaten Sidoarjo*. Surabaya, Indonesia: Universitas Pembangunan Nasional 'Veteran' Jatim, 2020.

[5] Y. Widyani and H. Laksmiwati, "Preliminary design of spatio-temporal disaster database in Indonesia to support emergency response," in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Aug. 2015, pp. 517–522, doi: 10.1109/ICEEI.2015.7352555.

[6] Badan Nasional Penanggulangan Bencana. (2021). *Catatan Refleksi Akhir Tahun Penanggulangan Bencana 2021*. 2021 BNPB mencatat 3.092,dan erupsi gunung api 1. diakses pada 1 Juni 2022 Pukul 21:40 WIB. [Online]. Available: https://www.bnpb.go.id/berita/catatan-refleksi-akhir-tahun-penanggulangan-bencana-2021#:~:text=Sepanjang

[7] A. H. Tantri and N. A. Rakhmawati, "Designing a natural disaster ontology for Indonesia," in *Proc. 12th Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Jul. 2019, pp. 130–134, doi: 10.1109/ICTS.2019.8850979.

[8] H. Riza, E. W. Santoso, I. G. Tejakusuma, and F. Prawiradisastra, "Advancing flood disaster mitigation in Indonesia using machine learning methods," in *Proc. Int. Conf. ICT Smart Soc. (ICISS)*, Nov. 2020, pp. 1–4, doi: 10.1109/ICISS50791.2020.9307561.

[9] A. Fariza, N. P. Abhimata, and J. A. N. Hasim, "Earthquake disaster risk map in east Java, Indonesia, using analytical hierarchy process—Natural break classification," in *Proc. Int. Conf. Knowl. Creation Intell. Comput. (KCIC)*, Nov. 2016, pp. 141–147, doi: 10.1109/KCIC.2016.7883638.

[10] R. Y. Setiawan, R. D. Susanto, A. Wirasatriya, I. Alifdini, A. D. Puryajati, L. Maslukah, and N. Nurdin, "Impacts of tropical cyclone Seroja on the phytoplankton chlorophyll—A and sea surface temperature in the Savu Sea, Indonesia," *IEEE Access*, vol. 9, pp. 152938–152944, 2021, doi: 10.1109/ACCESS.2021.3125605.

[11] P. Lundgren, T. Girona, and M. G. Bato, "Volcano monitoring with geodetic and thermal remote sensing time series," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Jul. 2021, pp. 977–979, doi: 10.1109/IGARSS47720.2021.9554167.

[12] S. Uhlemann, J. Chambers, P. Wilkinson, H. Maurer, A. Merritt, P. Meldrum, O. Kuras, D. Gunn, A. Smith, and T. Dijkstra, "Four-dimensional imaging of moisture dynamics during landslide reactivation," *J. Geophys. Res., Earth Surf.*, vol. 122, no. 1, pp. 398–418, Jan. 2017, doi: 10.1002/2016JF003983.

[13] K. D. Priyono, "Risk analysis of landslide impacts on settlements in Karanganyar, Central Java, Indonesia," *Int. J. GEOMATE*, vol. 19, no. 73, pp. 100–107, Sep. 2020, doi: 10.21660/2020.73.34128.

[14] P. Imani, G. Tian, S. Hadiloo, and A. A. El-Raouf, "Application of combined electrical resistivity tomography (ERT) and seismic refraction tomography (SRT) methods to investigate Xiaoshan district landslide site: Hangzhou, China," *J. Appl. Geophys.*, vol. 184, Jan. 2021, Art. no. 104236, doi: 10.1016/j.jappgeo.2020.104236.

[15] M. Kamiński, P. Zientara, and M. Krawczyk, "Electrical resistivity tomography and digital aerial photogrammetry in the research of the 'Bachledzki Hill' active landslide—In Podhale (Poland)," *Eng. Geol.*, vol. 285, May 2021, Art. no. 106004, doi: 10.1016/j.enggeo.2021.106004.

[16] A. T. W. Almais, M. Sarosa, and M. A. Muslim, "Implementation of multi experts multi criteria decision making for rehabilitation and reconstruction action after a disaster," *MATICS*, vol. 8, no. 1, p. 27, May 2016, doi: 10.18860/mat.v8i1.3480.

[17] A. Bachriwindi, E. K. Putra, U. M. Munawaroh, and A. T. W. Almais, "Implementation of Web-based weighted product use decision support system to determine the post-disaster damage and loss," *J. Phys., Conf. Ser.*, vol. 1413, no. 1, Nov. 2019, Art. no. 012019, doi: 10.1088/1742-6596/1413/1/012019.

[18] A. T. W. Almais, K. F. H. Holle, K. S. Kinasih, D. A. Wiranti, and S. Y. Yasin, ''Implementation fuzzy weighted product preparation post disaster reconstruction and rehabilitation action based dynamics decision support system,'' in *Proc. Int. Conf. Inf. Syst. Technol.*, 2019, pp. 272–277, doi: 10.5220/0009909002720277.

[19] A. H. Safitri, A. T. Wibowo Almais, A. Syauqi, and R. I. Melani, ''Pengujian optimization dan non-optimization query metode TOPSIS untuk menentukan tingkat kerusakan sektor bencana alam,'' *Jurnal ELTIKOM*, vol. 6, no. 1, pp. 89–99, Jan. 2022, doi: 10.31961/eltikom.v6i1.532.

[20] 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,'' *J. Appl. Geophys.*, vol. 198, Mar. 2022, Art. no. 104555, doi: 10.1016/j.jappgeo.2022.104555.

[21] M. P. Uddin, M. A. Mamun, and M. A. Hossain, ''Feature extraction for hyperspectral image classification,'' in *Proc. IEEE Region 10 Humanitarian Technol. Conf. (R10-HTC)*, Dec. 2017, pp. 379–382, doi: 10.1109/R10-HTC.2017.8288979.

[22] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, ''Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm,'' *Mach. Learn. With Appl.*, vol. 5, Sep. 2021, Art. no. 100044.

[23] Z. Geng, X. Duan, Y. Han, F. Liu, and W. Xu, ''Novel variation mode decomposition integrated adaptive sparse principal component analysis and it application in fault diagnosis,'' *ISA Trans.*, vol. 128, pp. 21–31, Sep. 2022, doi: 10.1016/j.isatra.2021.11.002.

[24] N. Taki, C. Delpha, D. Diallo, W. B. Hassen, and N. Ravot, ''Soft fault diagnosis in wiring networks using reflectometry and principal component analysis,'' *Measurement*, vol. 198, Jul. 2022, Art. no. 111378, doi: 10.1016/j.measurement.2022.111378.

[25] Y. Han, G. Song, F. Liu, Z. Geng, B. Ma, and W. Xu, ''Fault monitoring using novel adaptive kernel principal component analysis integrating grey relational analysis,'' *Process Saf. Environ. Protection*, vol. 157, pp. 397–410, Jan. 2022, doi: 10.1016/j.psep.2021.11.029.

[26] H. Cao, P. Sun, and L. Zhao, ''PCA-SVM method with sliding window for online fault diagnosis of a small pressurized water reactor,'' *Ann. Nucl. Energy*, vol. 171, Jun. 2022, Art. no. 109036, doi: 10.1016/j.anucene.2022.109036.

[27] M. Ahsan, M. Mashuri, and H. Khusna, ''Kernel principal component analysis (PCA) control chart for monitoring mixed non-linear variable and attribute quality characteristics,'' *Heliyon*, vol. 8, no. 6, Jun. 2022, Art. no. e09590, doi: 10.1016/j.heliyon.2022.e09590.

[28] P. Li, W. Zhang, C. Lu, R. Zhang, and X. Li, ''Robust kernel principal component analysis with optimal mean,'' *Neural Netw.*, vol. 152, pp. 347–352, Aug. 2022, doi: 10.1016/j.neunet.2022.05.005.

[29] N. Liang, Y. Tuo, Y. Deng, and T. He, ''PCA-based SVM classification for simulated ice floes in front of sluice gates,'' *Polar Sci.*, vol. 34, Dec. 2022, Art. no. 100839, doi: 10.1016/j.polar.2022.100839.

[30] T. Ueda, ''Estimation of three-dimensional particle size and shape characteristics using a modified 2D–3D conversion method employing spherical harmonic-based principal component analysis,'' *Powder Technol.*, vol. 404, May 2022, Art. no. 117461, doi: 10.1016/j.powtec.2022.117461.

[31] A. dos Santos Ferreira, D. M. Freitas, G. G. da Silva, H. Pistori, and M. T. Folhes, ''Unsupervised deep learning and semi-automatic data labeling in weed discrimination,'' *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104963, doi: 10.1016/j.compag.2019.104963.

[32] S. H. N. Hafida, N. K. M. Isa, M. H. Ibrahim, Jumadi, M. Toyib, and M. Musiyam, ''Data survey of students behavioral and psychological adaptations in disaster-prone areas of Mount Merapi in Indonesia,'' *Data Brief*, vol. 42, Jun. 2022, Art. no. 108229, doi: 10.1016/j.dib.2022.108229.

[33] J. Zhang, M. Chen, and X. Hong, ''Nonlinear process monitoring using a mixture of probabilistic PCA with clusterings,'' *Neurocomputing*, vol. 458, pp. 319–326, Oct. 2021, doi: 10.1016/j.neucom.2021.06.039.

[34] E. N. Aidoo, S. K. Appiah, G. E. Awashie, A. Boateng, and G. Darko, ''Geographically weighted principal component analysis for characterising the spatial heterogeneity and connectivity of soil heavy metals in Kumasi, Ghana,'' *Heliyon*, vol. 7, no. 9, Sep. 2021, Art. no. e08039, doi: 10.1016/j.heliyon.2021.e08039.

[35] R. Argota-Perez, J. Robbins, A. Green, M. V. Herk, S. Korreman, and E. Vásquez-Osorio, ''Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy,'' *Phys. Imag. Radiat. Oncol.*, vol. 22, pp. 13–19, Apr. 2022, doi: 10.1016/j.phro.2022.04.002.

[36] S. Gupta, ''A review of face detection and recognition using hybrid approaches,'' in *Proc. Int. J. Comput. Math. Sci.*, May 2017, pp. 78–85.

[37] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2020, doi: 10.1017/9781108679930.

[38] L.-C. Chang, J.-Y. Liou, and F.-J. Chang, ''Spatial-temporal flood inundation nowcasts by fusing machine learning methods and principal component analysis,'' *J. Hydrol.*, vol. 612, Sep. 2022, Art. no. 128086, doi: 10.1016/j.jhydrol.2022.128086.

[39] V. Susilo, R. R. Isnanto, and M. A. Riyadi, ''Herbal leaf pattern analisis using principal component analisis (PCA) and Canberra distance,'' in *Proc. 7th Int. Conf. Inf. Technol., Comput., Electr. Eng. (ICITACEE)*, Sep. 2020, pp. 100–104, doi: 10.1109/ICITACEE50144.2020.9239235.

[40] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, ''Effect of data scaling methods on machine learning algorithms and model performance,'' *Technologies*, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.

[41] L. I. Smith, ''A tutorial on principal components analysis,'' Dept. Comput. Sci., Univ. Otago, Dunedin, New Zealand, 2002. [Online]. Available: https://ourarchive.otago.ac.nz/bitstream/handle/10523/7534/OUCS-2002-12.pdf?sequence=1&isAllowed=y

[42] T. Gu, ''Detection of small floating targets on the sea surface based on multi-features and principal component analysis,'' *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 809–813, May 2020, doi: 10.1109/LGRS.2019.2935262.

[43] Iswan and I. Garniwa, ''Principal component analysis and cluster analysis for development of electrical system,'' in *Proc. 15th Int. Conf. Quality Res. (QiR), Int. Symp. Electr. Comput. Eng.*, Jul. 2017, pp. 439–443, doi: 10.1109/QIR.2017.8168526.

[44] Z. Kai, L. Daojing, C. Anjing, H. Dong, T. He, Y. Haifeng, D. Jianbo, L. Lei, Z. Yu, and Z. Running, ''Sparse flight spotlight mode 3-D imaging of spaceborne SAR based on sparse spectrum and principal component analysis,'' *J. Syst. Eng. Electron.*, vol. 32, no. 5, pp. 1143–1151, Oct. 2021, doi: 10.23919/JSEE.2021.000098.

[45] M. Lambers, H. Nies, and A. Kolb, ''Interactive dynamic range reduction for SAR images,'' *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 507–511, Jul. 2008, doi: 10.1109/LGRS.2008.922732.

[46] J. Wu, W. Zheng, J. Lai, and C. Y. Suen, ''Euler clustering on large-scale dataset,'' *IEEE Trans. Big Data*, vol. 4, no. 4, pp. 502–515, Dec. 2018, doi: 10.1109/TBDATA.2017.2742530.

[47] D. Huang, C. Wang, H. Peng, J. Lai, and C. Kwoh, ''Enhanced ensemble clustering via fast propagation of cluster-wise similarities,'' *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 1, pp. 508–520, Jan. 2021, doi: 10.1109/TSMC.2018.2876202.

[48] A. Thabet, N. Gasmi, G. B. H. Frej, and M. Boutayeb, ''Sliding mode control for Lipschitz nonlinear systems in reciprocal state space: Synthesis and experimental validation,'' *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 3, pp. 948–952, Mar. 2021, doi: 10.1109/TCSII.2020.3018016.

[49] F. Huang, J. Xia, C. Yin, X. Zhai, N. Xu, G. Yang, W. Bai, Y. Sun, Q. Du, M. Liao, X. Hu, P. Zhang, L. Duan, and Y. Liu, ''Assessment of FY-3E GNOS-II GNSS-R global wind product,'' *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7899–7912, 2022, doi: 10.1109/JSTARS.2022.3205331.

[50] F. Basile and L. Ferrara, ''Validation of industrial automation systems using a timed model of system requirements,'' *IEEE Trans. Control Syst. Technol.*, vol. 31, no. 1, pp. 130–143, Jan. 2023, doi: 10.1109/TCST.2022.3173890.

**AGUNG TEGUH WIBOWO ALMAIS** received the bachelor's degree from the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, in 2008, and the master's degree in communication and informatics systems from the Department of Electrical Engineering, Universitas Brawijaya, Malang, in 2016, where he is currently pursuing the Ph.D. degree in geophysics with the Department of Physics. He is also a Lecturer with the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include databases, decision support systems, machine learning, and disaster management.

**ADI SUSILO** received the bachelor's and master's degrees from the Department of Geophysics, Gajah Mada University, Yogyakarta, Indonesia, in 1989 and 1997, respectively, and the Ph.D. degree from James Cook University, Australia, in 2004. He is currently a Professor in disaster and exploration of natural resources with the Department of Physics, University of Brawijaya, Malang. His research interests include the disaster and exploration of natural resources and geosciences, groundwater mangrove, disaster management, geophysics, and geology.

**AGUS NABA** received the bachelor's degree from the Department of Physics, University of Brawijaya, Malang, Indonesia, in 1995, the master's degree from the Bandung Institute of Technology, in 2000, and the Doctorate (Dr.Eng.) degree from the University of Tsukuba, in 2007. He is currently a Lecturer and a Professor in intelligent science with the Department of Physics, University of Brawijaya. His research interests include smart systems and signal processing.

**MOECHAMMAD SAROSA** received the bachelor's degree from Universite De Nancy I, France, in 1990, and the master's and Ph.D. degrees from the Department of Electrical Engineering, Bandung Institute of Technology, in 2002 and 2007, respectively. He is currently a Lecturer and a Professor in computer science and in electrical engineering with the State Polytechnic of Malang. His research interests include information communication technology, artificial intelligence, mobile computing, embedded systems, and the Internet of Things.

**CAHYO CRYSDIAN** received the bachelor's degree from the Department of Electrical Engineering, University of Brawijaya, Malang, Indonesia, in 1997, the master's degree from the Department of Computer Science, University of Technology Malaysia, in 2003, and the Ph.D. degree from the University of Technology Malaysia, in 2006. He is currently a Lecturer with the Informatics Engineering Department, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include intelligence systems, computer vision, and image processing.

**IMAM TAZI** received the bachelor's degree from the Department of Physics, University of Brawijaya, Malang, Indonesia, in 1998, the master's degree from the Department of Physics, Surabaya Institute of Technology, in 2005, and the Ph.D. degree from the Department of Physics, Gajah Mada University, in 2014. He is currently a Lecturer with the Department of Physics, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include intelligent sensors and physics (sensors).

**MOKHAMAD AMIN HARIYADI** received the bachelor's degree from the Department of Informatics Engineering, Gajayana University, Malang, Indonesia, in 1987, the master's degree from the Department of Informatics Engineering, Sepuluh Nopember Institute of Technology, in 2000, and the Ph.D. degree from the Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, in 2014. He is currently a Lecturer with the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include intelligence systems and smart systems.

**MUHAMMAD AZIZ MUSLIM** received the bachelor's and master's degrees from the Department of Electrical Engineering, Surabaya Institute of Technology, Indonesia, in 1998 and 2001, respectively, and the Ph.D. degree from the Kyushu Institute of Technology, in 2008. He is currently a Lecturer with the Department of Electrical Engineering, University of Brawijaya, Malang. His research interests include computational intelligence, control systems, and robotics.

**PUSPA MILADIN NURAIDA SAFITRI ABDUL BASID** received the bachelor's degree from the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, in 2014, and the master's degree from the Department of Computer Science, University of Brawijaya, Malang, in 2017. She is currently a Lecturer with the Department of Informatics, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Her research interest includes software engineering.

**YUNIFA MIFTACHUL ARIF** received the Diploma 4 (D4), master's, and Ph.D. degrees from the Department of Electrical Engineering, Surabaya Institute of Technology, Indonesia, in 2007, 2010, and 2022, respectively. He is currently a Lecturer with the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interest includes game technology.

**MOHAMMAD SINGGIH PURWANTO** received the bachelor's and master's degrees from the Department of Geophysics Engineering, Surabaya Institute of Technology, Indonesia, in 2005 and 2009, respectively. He is currently pursuing the Ph.D. degree with the Department of Physics, Field of Geophysics, Brawijaya University Malang. He is also a Lecturer with the Department of Geophysical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests include aero geophysics, geographic information systems, artificial intelligence, and recharge areas.

**DIYAN PARWATININGTYAS** received the bachelor's degree from the Department of Physics, Padjadjaran University, Bandung, Indonesia, in 2001, and the master's degree from the Department of Engineering Physics, Bandung Institute of Technology, in 2007. He is currently pursuing the Ph.D. degree with the Department of Physics, Field of Geophysics, Universitas Brawijaya Poor. He is also a Lecturer with the Department of Physics, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia. His research interest includes applied geophysics.

**SUPRIYONO** received the bachelor's degree from the Department of Computer Science, University of Brawijaya, Malang, Indonesia, in 2005, and the master's degree from the Department of Informatics Engineering, Surabaya Institute of Technology, in 2009. He is currently a Lecturer with the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interest includes software engineering.

**HENDRO WICAKSONO** received the bachelor's degree from the Department of Computer Science, Institute of Technology Bandung, Indonesia, in 2002, the master's degree from the Department of Information and Communication Engineering, University of Karlsruhe, Germany, in 2006, and the Doctorate (Dr.Ing.) degree from the Department of IT for Engineering, Karlsruhe Institute of Technology, Germany, in 2016. He is currently a Lecturer and a Professor in industrial engineering with Constructor University, Bremen. His research interests include knowledge management, sustainability management, applied machine learning, data management, and industry 4.0.