This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

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], HENDRO WICAKSONO[4], IMAM TAZI[2], MOKHAMAD AMIN HARIYADI[2], MUHAMMAD AZIZ MUSLIM[5], PUSPA MILADIN NURAIDA SAFITRI ABDUL BASID[2], YUNIFA MIFTACHUL ARIF[2], MOHAMMAD SINGGIH PURWANTO[6], DIYAN PARWATININGTYAS[7], AND SUPRIYONO[2]

[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]Electrical Engineering, State Polytechnic of Malang Malang, Malang 65141, Indonesia
[4]Intelligent Systems and Production Engineering, Karlsruhe Institute of Technology, Karlsruhe, German
[5]Department of Electrical Engineering, Universitas Brawijaya, Malang 65145, Indonesia
[6]Department of Geophysical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[7]Department of Physics, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia, 10430 Indonesia

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

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, “This work was supported in part by the U.S. Department of Commerce under Grant BS123456.”

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 the 216 post-disaster sector damage data were processed using Principal Component Analysis (PCA) 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 the data normalization are then made into a 2-dimensional graph to see the visualization of the PCA results data. 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 with the original target data, and the PC2 results with the original target data. The result is that there are 2 updates, the first is that the distribution of PC1 data is very good in grouping the data when comparing the distribution of data with PC2, because the variance ratio and eigenvalue values of PC1 are greater than 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 cluster 1 ($n<0$), while the original target data which has a value of 2 (damaged moderately) occupies cluster 2 coordinates ($0 \leq n < 2$), and for the original target data the value 3 (heavily damaged) occupies cluster 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 an appropriate data label for further processing.

INDEX TERMS Clustering, Label, Post Disaster, Principal Component Analysis, Sector.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

I. INTRODUCTION

Natural disasters are a big problem for countries with tropical climates, such as Indonesia [1]–[3]. This country 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 [9], [10], tsunamis [11], and volcanic eruptions [12]. The natural disaster had a significant impact on damage, besides that the disaster also caused damage to all sectors including infrastructure [13], settlement or housing [14], economy or property [15], and local sectors such as religion, education, and social [16].

Determining the level of damage to the sector after natural disasters involves surveyors who have been assigned by the local government [17], on his research [18] 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. In his research [19] by applying existing criteria in research [18] then combined with the Decision Support System (DSS) method can be used to determine the level of damage to the sector after a natural disaster. Results of post-disaster sector damage levels in research [17]–[20] 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 level of damage to the post-natural disaster sector, but according to government regulations it already exists, but not in numerical form. 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 Principal Component Analysis (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 Principal Component Analysis (PCA) processes the 216 data to produce numerical values based on the results of the Principal Component Analysis (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 [21] using Principal Component Analysis (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*. Principal Component Analysis (PCA) is a technique for reducing the dimensions of a data set, increasing interpretability, and minimizing the loss of information in data [22]. In addition, using Principal Component Analysis (PCA) can reduce image complexity and execution time [23]. Principal Component Analysis (PCA) in addition to grouping a set of data into a normal data group, in research [24]–[26] uses Principal Component Analysis (PCA) to find the error 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 Principal Component Analysis (PCA).
2. Using Principal Component Analysis (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 Principal Component Analysis (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 Principal Component Analysis (PCA) to create labels and the use of Principal Component Analysis (PCA) in several studies in "Related Work". Then "Method and Data Preparation" explains the steps of Principal Component Analysis (PCA) and data acquisition. The "results and discussion", explains the Principal Component Analysis (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 I. 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. [27]. 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 [28], [29]. 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

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

model [30]. 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. [31]. In their research, *Santos Ferreira et al.* use two methods to label data on grass types in agriculture, these methods Deep Representations and Image Clusters (JULE) and Deep Clustering for Unsupervised Learning of Visual Features (DeepCluster) [32].

TABLE I
RELATED WORK FOR PCA AND CLUSTERING

References	Topic	Method	Subject
[21]	Clustering	K-Means Clustering and PCA	Grouping high-dimensional data into a low-dimensional space
[27]	Mapping data	PCA and SVM	Mapping high-dimensional data into a low-dimensional space.
[28], [29]	Separate data	KPCA (Kernel Principal Component Analysis)	Separating the differences in the value of attributes and variables on the product.
[30]	Reduce information	PCA and SVM	Eliminate overlapping and redundant information in the data.
[31]	Estimating data	PCA	Estimating the size and shape characteristics of the particles.
[32]	Label data	JULE and DeepCluster	Labeling of data types of grass in agriculture.
[33]	Sharing data	Clustering	Distribute the questions to respondents.
[34]	Extract data	Probabilistic PCA	Extract information on abundant, non-linear, and dynamic data.
[35]	Characterizing information	GWPCA (Geographically Weighted Principal Component Analysis)	Characterizing heavy metals to determine the potential of soil.
[36]	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.

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 [33]. 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 [34]. *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) [35]. *Argota-Perez et al.* also explained that PCA can model changes in data and evaluate data quantitatively [36].

Previous studies have implemented PCA using a framework for handling, reducing data, extracting information or data, dividing data, separating data, grouping data, extracting data, and mapping data. However, to label, a complex 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. Therefore, this study 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. METHOD AND DATA PREPARATION

Our method, 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 about data preparation which explains how to obtain and process sector damage data after natural disasters.

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 [21], [37]. According to *Deisenroth et al*. that PCA can project the original dataset to a new database with lower dimensions [38]. Projecting data into lower dimensions is called PC or principal components, to present as much information as possible beforehand [39].

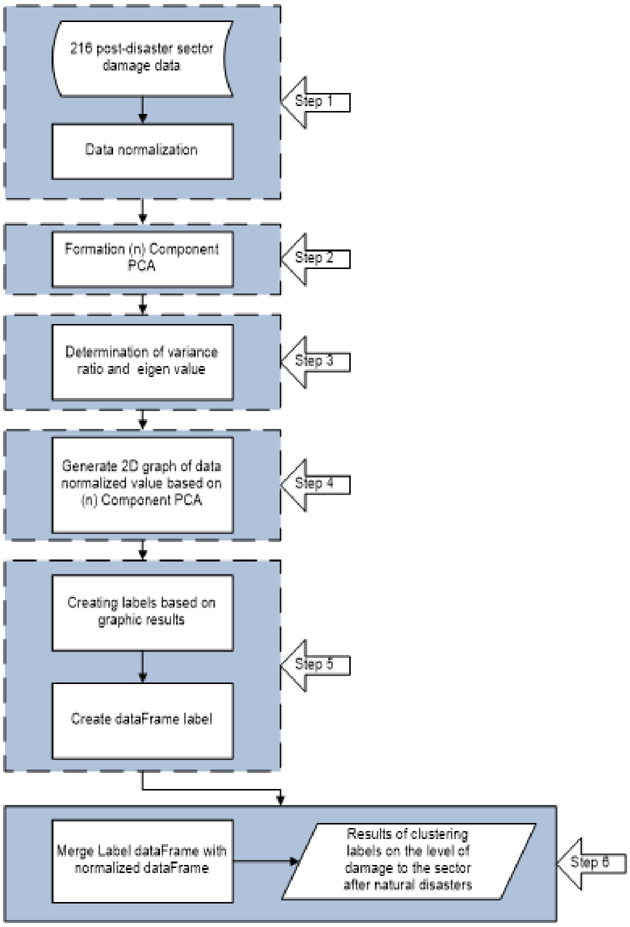


FIGURE 1. Our main architecture of the whole paper.

PCA is used for clustering data on sectoral damage after natural disasters, then the clustering results will be labeled according to the range of values from the PCA results. 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.
Data on damage to buildings or the post-natural disaster sector used is the result of analysis from surveyor data to determine the level of damage to buildings in the field, totaling 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 [40]. In his research [41] 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) [42]:

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

The standard deviation uses the equation (2) as follows [42]:

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

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):

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

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 [42].

Step 4: Determine the number of principal components (PC).
In his research [43] 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 [44].

Step 5: Creating the visualization.
Visualization is a very important thing in representing a result [45], 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

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

and minimizing the loss of information caused by data reduction [46]. In our research, to create a range of values using 2 coordinate points, namely the coordinate points from PC1 and the coordinate points from PC2.

Step 7: Clustering data.
Clustering is a learning technique that divides data into several parts of unsupervised data into several homogeneous data groups [47], [48]. In our research for clustering data using PC1 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 [49], [50]. Validation is a key role in determining whether the results obtained are by existing requirements [51]. In this study, to validate the results, we used a comparison of the results of PC1 data clustering with the original target data.

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

Algorithm: Clustering Using PCA for Labeling Data

```
Input    :  Using libraries
            pd.read_csv('filename.CSV)
Process  :  Data normalization using the
            library StandardScaler() and
            fit_transform()
            Generate variance ratio using a
            library explained_variance_ratio_()
            Generate eigenvalues using a
            library explained_variance()
            Generate PCA components using
            libraries PCA()
            Generate data frame resulting from
            PCA components using a library
            pd.DataFrame()
            Visualization of normalization
            results using scatterplot()
            Generate labels using branching (if
            else) and branching (for)
Output   :  Merge the data frame results of PCA
            components with the results of
            using labels concat()
            Visualization of data labeling
            results using PCA with validation
            training data from surveyors using
            scatterplot()
```

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 II.

TABLE II
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

The data in Table II 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 here is the target or result of the type of damage to the post-natural disaster sector. Table II 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 III has a value of 1/2/3, this value has an understanding that refers to research [18] which is as shown in Table III.

TABLE III
LEVEL OF DAMAGE TO BUILDINGS AFTER NATURAL DISASTERS

Value	Description
1	Light Damage
2	Moderate Damage
3	Heavy Damage

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 get the level of accuracy of the results of labeling using PCA.

A. DATA STANDARDIZATION

Standardize the data contained in Table II so that the data has the same weight when forming a main component (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 library which produces values like figure 2.

```
[[-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**

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

In the data standardization process, it produces a value that is different from the original value, in Table II 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.

B. VALUE OF VARIANCE RATIO AND EIGENVALUE

Determining the value of the variance ratio can use equation (3) by applying the library from the Python programming language "PCA.explained_variance_ratio_.cumsum()", then determine the eigenvalue using the library "PCA.explained_variance_ratio_". This study has eigenvalues and variance ratios for each feature as shown in Table IV which explains the 5 features after being reduced using PCA and dividing them into 5 main components which produce different values of explained variance ratio and cumulative explained variance ratio for each of its features.

TABLE IV
EIGEN VALUE AND VARIANCE RATIO

Number of Components	Eigenvalue	Variance ratio (%)
1.0	85.16	85.16
2.0	9.35	94.52
3.0	4.56	99.09
4.0	0.73	99.82
5.0	0.17	100.0

To see the percentage of the variance ratio for each main component is shown in Figure 3, which illustrates the increase in the value of the variance ratio for each main component.

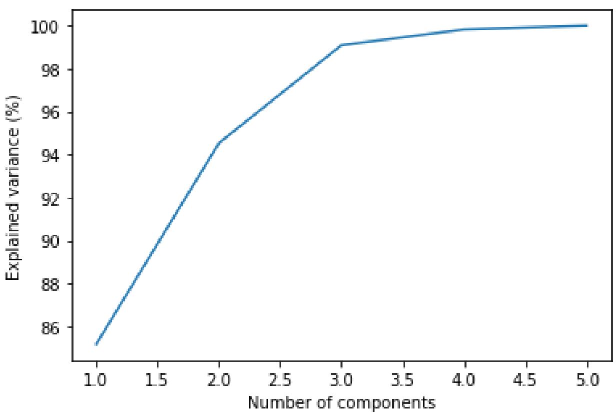


FIGURE 3. Variance Ration Chart

C. NUMBER OF MAIN COMPONENTS (PC)

To determine the number of principal components (PC), look at the eigenvalue and variance ratio. If you look at Table IV, the highest eigenvalue is found in main component 1 (PC1) with an eigenvalue of 85.16 and a variance ratio of 85.16%. But the highest eigenvalue cannot be used as a reference in this study because the goal 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 smaller eigenvalue than PC1 which is 9.35, and a variance ratio value of 94.52%. Using PC2 for comparison when measuring the level of accuracy for the testing process in labeling data. To see the variance ratio values for PC1 and PC2 are shown in Figure 4, explaining that the variance ratio values for PC1 and PC2 form a straight line which means that PC2 data has a wider distribution of data when compared to the distribution of data on PC1.

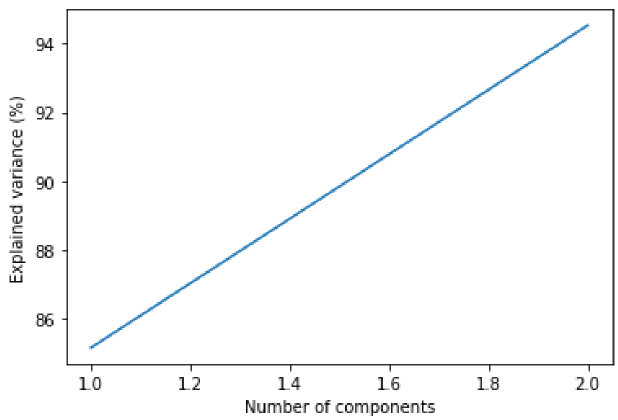


FIGURE 4. Graph of PC1 and PC2 Variance Ration Value

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.

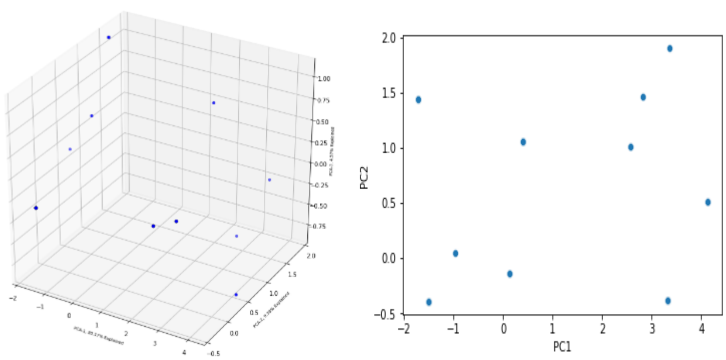


FIGURE 5. 3D and 2D Visualization of Data Distribution on Main Components

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.

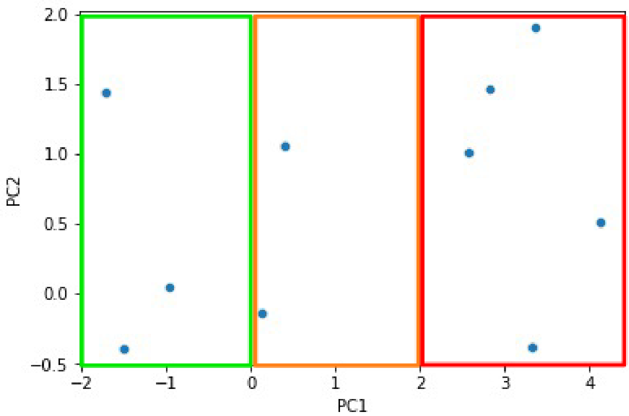


FIGURE 6. Coordinate Points of PC1 and PC2 Data Distribution

The coordinates of each data distribution, it is divided into 3 groups as shown in Figure 7.

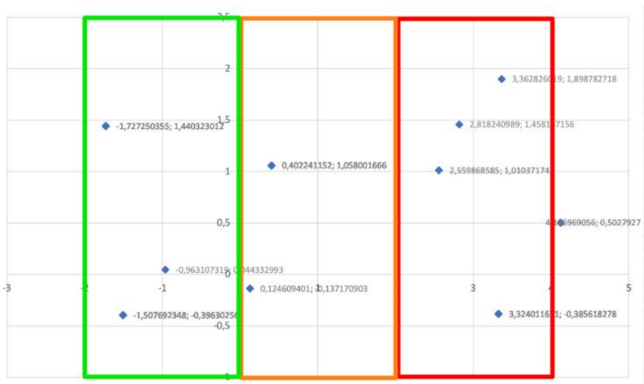


FIGURE 8. PC1 Data Clustering Results Based on Original Target Data

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

TABLE V
POINT AND COORDINATE VALUE PC1 PC2

Color Clustering	Coordinate Point PC1	Coordinate Point PC2	Coordinate Value Range (n)
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	
	2,818240989	1,458147156	
	3,362826019	1,898782718	

Based on the results of the visualization in Figure 7 and the elaboration of the coordinate points in Table V, 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.

F. VALIDATION HASIL CLUSTERING DATA

Testing the validation of data clustering results by using a comparison of 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.

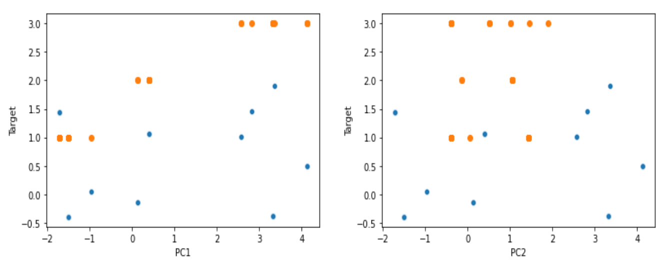


FIGURE 7. Coordinate Points of PC1 and PC2 Data Distribution

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. For PC2 the data distribution when compared to the original data target cannot be as good as using PC1 because

PC1 and PC2 have eigenvalues and variance ratios which cause the distribution of the data to also be different. Figure

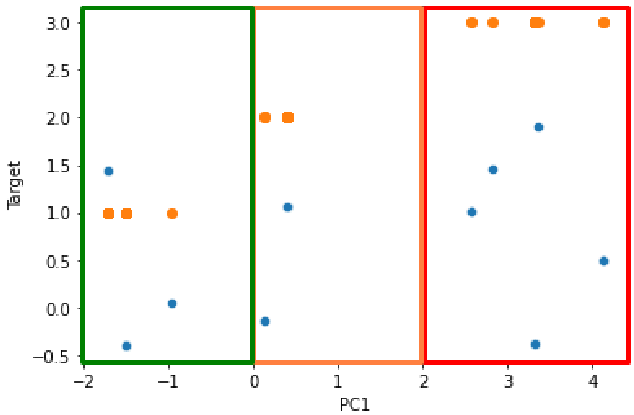


FIGURE 9. Results of Distribution of PC1 and PC2 Data Based on Original Target Data

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 VI.

TABLE VI
PC1 COMPARISON RESULT DATA LABEL BASED ON ORIGINAL TARGET DATA

Comparison Result		Coordinate Value Range (n)	Label
PC1	Target Original Data		
-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		

Table VI 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 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 [21]. To measure the results of labeling data with PCA, you can look at Table VI 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 for PC1 and 9.35 for PC2. So label the data on the level of damage to the post-disaster sector using the distribution of PC1 data, which produces 3 clustering data based on the distribution of the original target data with PC1 data. As a result, the original target data which has a value of 1 (lightly damaged) is grouped at the coordinate point of PC1 (n) with a range of values $n < 0$, for 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 author 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 at the 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 *2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, Jan. 2021, pp. 201–204. doi: 10.1109/IoTaIS50849.2021.9359695.

[2] T. Yudistira *et al.*, "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, May 2021, doi: 10.1093/gji/ggab104.

[3] P. Supendi *et al.*, "Relocated aftershocks and background seismicity in eastern Indonesia shed light on the 2018 Lombok and Palu earthquake sequences," *Geophys J Int*, vol. 221, no. 3, pp. 1845–1855, Jun. 2020, doi: 10.1093/gji/ggaa118.

[4] F. A. Maulana, "Coordinating Functions of the Sidoarjo Regional Disaster Management Agency (BPBD) in Disaster Management in Sidoarjo Regency," Surabaya, 2020.

[5] Y. Widyani and H. Laksmiwati, "Preliminary design of spatio-temporal disaster database in Indonesia to support emergency response," in *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, Aug. 2015, pp. 517–522. doi: 10.1109/ICEEI.2015.7352555.

[6] Badan Nasional Penanggulangan Bencana, "2021 Disaster Management Year End Reflection Note," 2021. https://www.bnpb.go.id/berita/catatan-refleksi-akhir-tahun-penanggulangan-bencana-2021#:~:text=Sepanjang 2021 BNPB mencatat 3.092,dan erupsi gunung api 1. diakses pada 1 Juni 2022 Pukul 21:40 WIB

[7] A. H. Tantri and N. A. Rakhmawati, "Designing A Natural Disaster Ontology for Indonesia," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Jul. 2019, pp. 130–134. doi: 10.1109/ICTS.2019.8850979.

[8] H. Riza, E. W. Santoso, I. Gunawan Tejakusuma, and F. Prawiradisastra, "Advancing Flood Disaster Mitigation in Indonesia Using Machine Learning Methods," in *2020 International Conference on ICT for Smart Society (ICISS)*, Nov. 2020, pp. 1–4. doi: 10.1109/ICISS50791.2020.9307561.

[9] P. Supendi *et al.*, "Relocated aftershocks and background seismicity in eastern Indonesia shed light on the 2018 Lombok and Palu earthquake sequences," *Geophys J Int*, vol. 221, no. 3, pp. 1845–1855, Jun. 2020, doi: 10.1093/gji/ggaa118.

[10] A. Fariza, N. P. Abhimata, and J. A. Nur Hasim, "Earthquake disaster risk map in east Java, Indonesia, using analytical hierarchy process — Natural break classification," in *2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC)*, Nov. 2016, pp. 141–147. doi: 10.1109/KCIC.2016.7883638.

[11] R. Y. Setiawan *et al.*, "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.

[12] P. Lundgren, T. Girona, and M. G. Bato, "Volcano Monitoring with Geodetic and Thermal Remote Sensing Time Series," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul. 2021, pp. 977–979. doi: 10.1109/IGARSS47720.2021.9554167.

[13] S. Uhlemann *et al.*, "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.

[14] K. D. Priyono, "Risk Analysis Of Landslide Impacts On Settlements In Karanganyar, Central Java, Indonesia," *International Journal of GEOMATE*, vol. 19, no. 73, Sep. 2020, doi: 10.21660/2020.73.34128.

[15] 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 Geophy*, vol. 184, p. 104236, Jan. 2021, doi: 10.1016/j.jappgeo.2020.104236.

[16] 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, p. 106004, May 2021, doi: 10.1016/j.enggeo.2021.106004.

[17] A. T. Wibowo 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.

[18] 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, p. 012019, Nov. 2019, doi: 10.1088/1742-6596/1413/1/012019.

[19] A. T. W. Almais, . Fatchurrohman, 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 *Proceedings of the International Conferences on Information System and Technology*, 2019, pp. 272–277. doi: 10.5220/0009909002720277.

[20] A. H. Safitri, A. T. Wibowo Almais, A. Syauqi, and R. I. Melani, "Testing Optimization and Non-Optimization Query Topsis Methods to Determine Damage Level in Natural Disaster Sector," *Jurnal ELTIKOM*, vol. 6, no. 1, pp. 89–99, Jan. 2022, doi: 10.31961/eltikom.v6i1.532.

[21] 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 Geophy*, vol. 198, p. 104555, Mar. 2022, doi: 10.1016/j.jappgeo.2022.104555.

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

[23] 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," *Machine Learning with Applications*, vol. 5, p. 100044, 2021.

[24] 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*, Nov. 2021, doi: 10.1016/j.isatra.2021.11.002.

[25] N. Taki, C. Delpha, D. Diallo, W. ben Hassen, and N. Ravot, "Soft fault diagnosis in wiring networks using reflectometry and Principal Component Analysis," *Measurement*, vol. 198, p. 111378, Jul. 2022, doi: 10.1016/j.measurement.2022.111378.

[26] 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 Safety and Environmental Protection*, vol. 157, pp. 397–410, Jan. 2022, doi: 10.1016/j.psep.2021.11.029.

[27] 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, p. 109036, Jun. 2022, doi: 10.1016/j.anucene.2022.109036.

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

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

[30] N. Liang, Y. Tuo, Y. Deng, and T. He, "PCA-based SVM classification for simulated ice floes in front of sluice gates," *Polar Sci*, p. 100839, Mar. 2022, doi: 10.1016/j.polar.2022.100839.

[31] 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, p. 117461, May 2022, doi: 10.1016/j.powtec.2022.117461.

[32] 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 Agric*, vol. 165, p. 104963, Oct. 2019, doi: 10.1016/j.compag.2019.104963.

[33] 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, p. 108229, Jun. 2022, doi: 10.1016/j.dib.2022.108229.

[34] 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.

[35] 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, p. e08039, Sep. 2021, doi: 10.1016/j.heliyon.2021.e08039.

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

[37] S. Gupta, "A Review of Face Detection and Recognition using Hybrid Approaches," in *International Journal of Computer & Mathematical Sciences* , May 2017, pp. 78–85.

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

[39] 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 (Amst)*, vol. 612, p. 128086, Sep. 2022, doi: 10.1016/j.jhydrol.2022.128086.

[40] V. Susilo, R. R. Isnanto, and M. A. Riyadi, "Herbal Leaf Pattern Analisis Using Principal Component Analisis (PCA) and Canberra Distance," in *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Sep. 2020, pp. 100–104. doi: 10.1109/ICITACEE50144.2020.9239235.

[41] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *Technologies (Basel)*, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.

[42] L. I. Smith, *A tutorial on Principal Components Analysis*. 2002.

[43] T. Gu, "Detection of Small Floating Targets on the Sea Surface Based on Multi-Features and Principal Component Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 809–813, May 2020, doi: 10.1109/LGRS.2019.2935262.

[44] Iswan and I. Garniwa, "Principal component analysis and cluster analysis for development of electrical system," in *2017 15th International Conference on Quality in Research (QiR) : International Symposium on Electrical and Computer Engineering*, Jul. 2017, pp. 439–443. doi: 10.1109/QIR.2017.8168526.

[45] Z. Kai *et al.*, "Sparse flight spotlight mode 3-D imaging of spaceborne SAR based on sparse spectrum and principal component analysis," *Journal of Systems Engineering and Electronics*, vol. 32, no. 5, pp. 1143–1151, Oct. 2021, doi: 10.23919/JSEE.2021.000098.

[46] M. Lambers, H. Nies, and A. Kolb, "Interactive Dynamic Range Reduction for SAR Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 507–511, Jul. 2008, doi: 10.1109/LGRS.2008.922732.

[47] J.-S. Wu, W.-S. Zheng, J.-H. 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.

[48] D. Huang, C.-D. Wang, H. Peng, J. Lai, and C.-K. 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.

[49] 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 Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 3, pp. 948–952, Mar. 2021, doi: 10.1109/TCSII.2020.3018016.

[50] F. Huang *et al.*, "Assessment of FY-3E GNOS-II GNSS-R Global Wind Product," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 7899–7912, 2022, doi: 10.1109/JSTARS.2022.3205331.

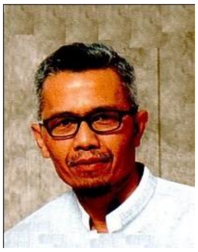
[51] F. Basile and L. Ferrara, "Validation of Industrial Automation Systems Using a Timed Model of System Requirements," *IEEE Transactions on Control Systems Technology*, vol. 31, no. 1, pp. 130–143, Jan. 2023, doi: 10.1109/TCST.2022.3173890.

AGUNG TEGUH WIBOWO ALMAIS earned a bachelor's degree from the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, in 2008, and a master's degree from the Department of Electrical Engineering, Field of Interest in Communication and Informatics Systems, Universitas Brawijaya Malang, in 2016. Currently pursuing a Doctoral degree in the Department of Physics, Field of Geophysics, Universitas Brawijaya Malang. He is also a Lecturer in 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 earned a bachelor's degree from the Department of Geophysics, Gajah Mada University, Yogyakarta, Indonesia, in 1989, a master's degree from the Department of Geophysics, Gajah Mada University, in 1997, a Ph.D. from James Cook University Australia, in 2004. He is also a professor in the Field of Disaster and Exploration of Natural Resources in the Department of Physics, University of Brawijaya Malang. His research interests are Disaster and Exploration of Natural Resources and Geosciences, Groundwater mangrove, Disaster management, Geophysics, Geology

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3275852

AGUS NABA earned a bachelor's degree from the Department of Physics, Brawijaya University, Malang, Indonesia, in 1995, a master's degree from the Bandung Institute of Technology, in 2000, and a doctorate (Dr.Eng) from the University Of Tsukuba, in 2007. He is also a Lecturer and Professor in the Field of Intelligent Science in the Department of Physics, University of Brawijaya Malang. His research interests include Smart Systems and Signal Processing.



MOECHAMMAD SAROSA received a bachelor's degree from Universite De Nancy I, France, in 1990, a master's degree from the Department of Electrical Engineering, Bandung Institute of Technology, in 2002, and a Doctorate from the Department of Electrical Engineering, Bandung Institute of Technology, in 2007. He is also a Lecturer and Professor of Computer Science in Electrical Engineering, at 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 earned a bachelor's degree from the Department of Electrical Engineering, University of Brawijaya, Malang, Indonesia, in 1997, a master's degree from the Department of Computer Science, the University of Technology Malaysia, in 2003, and a Doctoral degree from the University of Technology Malaysia, in 2006. He is also a Lecturer at the Informatics Engineering Department, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include Intelligence Systems, Computer Vision, and Image Processing.

HENDRO WICAKSONO earned a bachelor's degree from the Department of Computer Science, Institute of Technology Bandung, Indonesia, in 2002, a master's degree from the Department of Information and Communication Engineering, University of Karlsruhe, Germany in 2006, and a Doctorate (Dr.Ing) from the Department of IT for engineering, Karlsruhe Institute of Technology, Germany, in 2016. He is also a Lecturer and Professor of Industrial Engineering at Jacobs University Bremen. His research interests include Knowledge Management, Sustainability Management, Applied Machine Learning, Data Management, and Industry 4.0.

IMAM TAZI earned a bachelor's degree from the Department of Physics, Brawijaya University, Malang, Indonesia, in 1998, a master's degree from the Department of Physics, Surabaya Institute of Technology, in 2005, and a Doctorate from the Department of Physics, Gajah Mada University, in 2014. He is also a Lecturer at the Department of Physics, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include Intelligent sensors, Physics (sensors).

MOKHAMMAD AMIN HARIYADI earned a bachelor's degree from the Department of Informatics Engineering, Gajayana University, Malang, Indonesia, in 1987, a master's degree from the Department of Informatics Engineering, Sepuluh Nopember Institute of Technology, in 2000, and a Doctorate from the Department of Electrical Engineering, Sepuluh Nopember Institute of Technology, in 2014. He is also a Lecturer in the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interests include Intelligence Systems and Smart Systems.

MUHAMMAD AZIZ MUSLIM earned a bachelor's degree from the Department of Electrical Engineering, Surabaya Institute of Technology, Indonesia, in 1998, a master's degree from the Department of Electrical Engineering, Surabaya Institute of Technology, in 2001, and a Ph.D. from Kyushu Institute Of Technology, in 2008. He is also a Lecturer in the Department of Electrical Engineering, at the University of Brawijaya, Malang. His research interests include Computational Intelligence, Control Systems, and Robotics.

PUSPA MILADIN NURAIDA SAFITRI ABDUL BASID earned a bachelor's degree from the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, in 2014, and a master's degree from the Department of Computer Science, Brawijaya University Malang, in 2017. He is also a Lecturer at the Department of Informatics, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interest is in the field of Software Engineering.

YUNIFA MIFTACHUL ARIF earned a Diploma 4 (D4) degree from the Department of Electrical Engineering, Surabaya Institute of Technology, Indonesia, in 2007, a master's degree from the Department of Electrical Engineering, Surabaya Institute of Technology, in 2010, and a Doctoral degree from the Department of Electrical Engineering, Surabaya Institute of Technology, in 2022. He is also a Lecturer in the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interest is Game Technology.

MOHAMMAD SINGGIH PURWANTO earned a bachelor's degree from the Department of Geophysics Engineering, Surabaya Institute of Technology, Indonesia, in 2005, and a master's degree from the Department of Geophysics Engineering, Surabaya Institute of Technology, Indonesia, in 2009. Currently pursuing a Doctoral degree in the Department of Physics, Field of Geophysics, Brawijaya University Malang. He is also a Lecturer in the Department of Geophysics Engineering at the Surabaya Institute of

Technology, Indonesia. His research interests include Aero Geophysics, Geographic Information Systems, Artificial Intelligence, and Recharge Areas.

DIYAN PARWATININGTYAS earned a bachelor's degree from the Department of Physics, Padjadjaran University, Bandung, Indonesia, in 2001, and a master's degree from the Department of Engineering Physics, Bandung Institute of Technology, in 2007. Currently pursuing a Doctoral degree in the Department of Physics, Field of Geophysics, Universitas Brawijaya Poor. He is also a Lecturer in the Department of Informatics Engineering, at Indraprastha University PGRI Jakarta. His research interest is Applied Geophysics.

SUPRIYONO earned a bachelor's degree from the Department of Computer Science, Brawijaya University, Malang, Indonesia, in 2005, and a master's degree from the Department of Informatics Engineering, Surabaya Institute of Technology, in 2009. He is also a Lecturer in the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang. His research interest is in the field of Software Engineering.