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.2024.3469637

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

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

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Characterization of Structural Building Damage in Post-Disaster using GLCM-PCA Analysis Integration

AGUNG TEGUH WIBOWO ALMAIS[1,2], ADI SUSILO[1], AGUS NABA[1], MOECHAMMAD SAROSA[3], ALAMSYAH MUHAMMAD JUWONO[1], CAHYO CRYSDIAN[2], MUHAMMAD AZIZ MUSLIM[4], AND HENDRO WICAKSONO[5]

[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]Department of Electrical Engineering, Universitas Brawijaya, Malang 65145, Indonesia
[5]School of Business, Social & Decision Science, Constructor University, 28759 Bremen, Germany

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 Understanding the characteristics of a building after a natural disaster can be achieved using image analysis techniques. Among these techniques are the Gray-Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA). In the GLCM process, the input image is converted into numerical values using eight different angles and varying pixel distances (1 and 0.5 pixels). The resulting numerical values from GLCM are then fed into the PCA process to reveal information stored within post-disaster building images. Interestingly, the PCA results differ between images processed with GLCM at a 1-pixel distance versus a 0.5-pixel distance. After validation based on surveyor assessments, it was found that the valid and accurate representation of real-world image information corresponds to the GLCM results obtained with a 0.5-pixel distance, indicating severe damage. This conclusion is supported by the fact that PCA results using a GLCM distance of 0.5 produce 2D and 3D visualizations predominantly clustered around severely damaged coordinates, with a range of values (n) where $n \geq 2$. Therefore, integrating image analysis techniques such as GLCM and PCA can be used to determine the level of post-disaster building damage.

INDEX TERMS Characteristics, Building, Post-Natural Disaster, GLCM, PCA.

I. INTRODUCTION

Infrastructure resilience to natural disasters is one of the biggest challenges faced by modern society. Disasters such as earthquakes, storms, and floods can cause significant structural damage to buildings, resulting in huge economic losses and, more importantly, loss of human life [1], [2]. Therefore, post-disaster structural damage assessment is very important to ensure the safety and sustainability of building structures. In this context, an integrated approach using the Gray Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA) offers untapped potential in structural damage assessment. *Wildeman* has shown that non-destructive methods such as PCA and GLCM can provide valuable insight into the structural condition of buildings [3]. In addition, the GLCM-PCA integration can be a decision making to determine the level of building damage after natural disasters with the latest development, in his research, *Mohammad Aljanabi* explained that in civil engineering the latest knowledge to determine building damage is found in the decision-making stage [4]. According to *Aklouche et al.* PCA has been used extensively in multivariate data analysis to reduce data dimensions while retaining most of the information that matters [5]. Meanwhile, GLCM, as a texture analysis tool, has proven effective in identifying patterns of damage to materials [6]. However, integrating these two methods in the context of post-disaster structural damage assessment is rarely explored. Although there have been advances in structural damage assessment techniques, significant research gaps remain. Most previous studies have focused on using individual methods to detect damage, which are often insufficient to capture the complexity of damage incurred in

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

natural disasters [7]. In addition, existing research often does not fully use Product Lifecycle Data (PLD) which can improve the accuracy of damage assessment [8]. This research aims to overcome these shortcomings by integrating GLCM and PCA, thus providing a more robust and accurate method for post-disaster structural damage assessment. Furthermore, this research can also make a significant contribution to the field of post-disaster structural damage assessment with several innovative aspects:

1. Combining GLCM and PCA for the first time as an integrated method in structural damage assessment, provides a new perspective in damage data analysis.
2. The proposed model promises improved accuracy in damage detection, which is critical for emergency intervention and post-disaster reconstruction.
3. Given the large data and computationally intensive requirements of the Convolutional Neural Network (CNN) method, introducing automation in the damage assessment process significantly reduces the time required.
4. Assist stakeholders (the government) in distributing more objective assistance with the condition of victims affected by natural disasters.

The content of this article explains the following: the background to the use of GLCM-PCA to assess the extent of damage to buildings after natural disasters and the use of GLCM-PCA in several studies in "Related Works". Then the "Method" describes the steps of GLCM-PCA. "Results and discussion", describes the experimental process of GLCM-PCA using Python programming language to assess the extent of sector damage after natural disasters and validate results. Finally, the "conclusion" summarizes the results of the experiment and opportunities for future research.

II. RELATED WORKS

In their research, *Almais et al.* explained that the level of damage to buildings after natural disasters was 3 clusters, namely lightly damaged, moderately damaged, and severely damaged by clustering unsupervised data using the PCA technique [9]. TABLE I is the result of a review of previous related studies and the position of this study.

TABLE I
RELATED WORKS TO GLCM AND PCA

Reference	Topic	Method	Subject
[10]	Structural damage assessment	PCA	Cross-correlation analysis in a data-driven approach
[11]	Steel structure damage detection	PCA	Modal frequency variation of the dynamic test of steel structures
[12]	Detection of damage to plate structures	PCA-FRF	Explore PCA-FRF features using Unsupervised Machine Learning
[13]	Focusing on Parkinson's disease	GLCM-PCA	Structural Analysis using GLCM and PCA applications
[14]	Structural Damage Assessment	Convolutional Neural Network (CNN)	CNN's Effectiveness in Assessing Earthquake Damage
[15]	Structural damage detection	GLCM and Machine Learning	Comparing GLCM with machine learning methods in structural damage detection
[16]	Structural health monitoring	PCA and Deep Learning	Integrate PCA with Deep Learning for structural health monitoring.
[17]	Detection of damage to composite structures	GLCM and PCA	Improve damage detection in composite structures by using GLCM and PCA.
[9]	Clustering Data	PCA	Labeling data on the level of damage to buildings after natural disasters
Ours	Assessment of the extent of damage to buildings	GLCM-PCA	Characteristics of the level of damage to buildings after natural disasters

In the existing literature review, several studies have applied GLCM and PCA in different contexts. However, the integrated application of both in post-disaster structural damage assessment is still limited. The study identified that there is still a need to develop a more comprehensive damage assessment model, one that does not rely on just one method, but integrates various analytical techniques to obtain more accurate and reliable results. Therefore, this study seeks to fill this gap by proposing an integrated method that harnesses the strengths of GLCM and PCA.

III. PROPOSED METHOD AND DATA PREPARATION

The methods chapter describes two sub-chapters, namely the proposed method and data preparation. The first sub-chapter, the proposed method, describes the steps of GLCM and PCA. The second sub-chapter, data preparation, describes how to obtain data, data parameters or features, and fill in data on damage to buildings after natural disasters.

A. PROPOSED METHOD

By using input in the form of images of buildings affected by natural disasters, then the images enter the GLCM process to find out the value of each feature in the GLCM. FIGURE 1 is the architecture of the research conducted.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

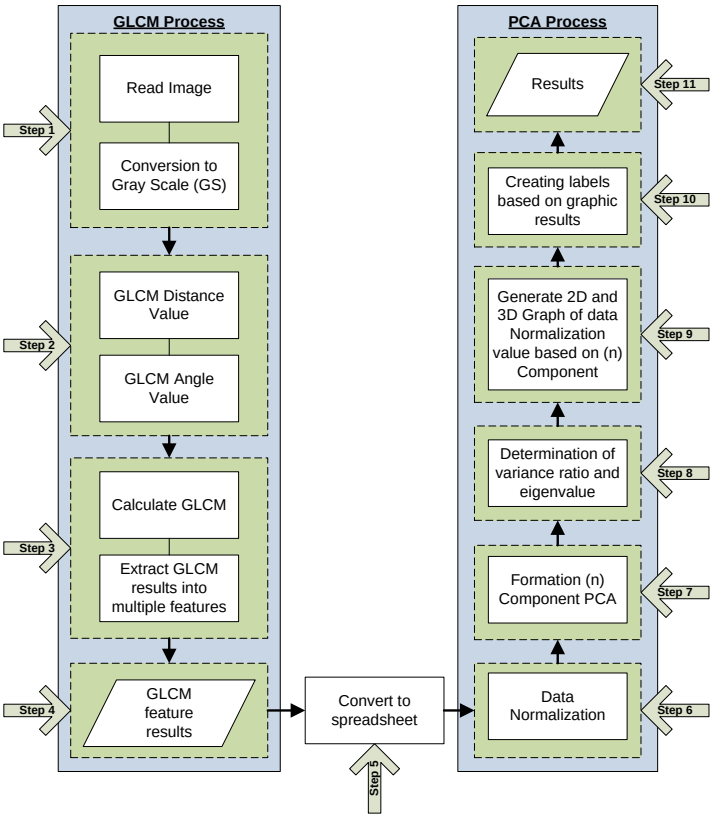


FIGURE 1. **Our main architecture of the whole paper.**

The value of the features in GLCM then enters the conversion process to a spreadsheet to be analyzed using PCA to find out the clustering of GLCM feature values, the results of clustering GLCM features using PCA produce clustering of the level of building damage after natural disasters based on research by *Almais et al.* For GLCM-PCA integration steps to assess the level of damage to buildings after natural disasters in more detail are as follows:

Step 1: Reading the Image and converting to Grey Scale (GS)

Reading images is often related to visual literacy skills, namely the ability to understand, create, and communicate meaning from information conveyed in the form of images [18]. The conversion to GS according to *Chen et al.* that today gray image coloring remains a major task in areas such as animated films, medical image processing, and various computer vision [19]. GS images are one of the data inputs to find out information from an image because the size is not heavy and the information available is enough to help decide a problem.

Step 2: GLCM Distance and Angle Value

GLCM uses a combination of distance and direction to analyze an image, so the approach can produce a high degree of accuracy in recognizing image patterns such as horizontal, vertical, diagonal, checkerboard, and color texture calcification [20], [21]. The study used distance based on the research of *Benco et al.* which used a distance value (d) of 1 [20], but in this study, the comparison experiment used the value of d=0.5. So in this study, the results were compared using different GLCM distances (d), namely 1 and 0.5. As for the angular magnitude, *Srivastava et al.* used the reference for their research which explained that in GLCM 4 angles can be used, namely 0^0, 45^0, 90^0, and 135^0 with the other 4 angles, namely 180^0, 225^0, 270^0, and 315^0 have the same result [21]. However, this study uses 8 angles, namely 0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, and 315^0 for the process of finding feature values in GLCM.

Step 3: Calculate GLCM and extract GLCM result into multiple features

GLCM can extract 8 features significantly distinguishing normal and abnormal brain images [22]. GLCM can also lighten computation and elevate accuracy levels making it more efficient to use for real-time pattern recognition applications [23].

Step 4: GLCM Feature Results

The result of the GLCM feature is the values obtained from image texture analysis using a grayish-level co-occurrence matrix. These features include various aspects of texture such as contrast, dissimilarity, homogeneity, energy, and correlation used for image classification, medical analysis, and other pattern recognition applications [20], [24].

Step 5: Convert to Spreadsheet

The term "Convert to Spreadsheet" is the process of converting data from another format into spreadsheet form, such as Excel. This is often done to facilitate data analysis, especially in research that involves collecting and processing large amounts of data [25].

Step 6: Data Normalization

Data normalization in PCA (Principal Component Analysis) is an interesting topic. In some cases, PCA assumes that data input is normally distributed, which is not always true in the real world. However, applying normalization to input data can change the data structure and affect the results of multivariate analysis and calibration used in data mining [26], [27]. In his research [28] There are six methods to standardize data, namely:

1. Normalization (NR): Changes the data so that it has a mean of 0 and a variance of 1.
2. Standard Scale (SS): Converts data into a z-score, with a mean of 0 and a standard deviation of 1.
3. MinMax Scaling (MM): Converts data into a range of 0 to 1.
4. MaxAbs Scaling (MA): Converts data into a range of -1 to 1.
5. Robust Scaling (RS): Transform data based on median and interquartile range.
6. Quantile Transformer (QT): Converts data into a uniform or normal distribution.

This study uses the Standard Scale (SS) because the type of building damage data after natural disasters

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

uses a standard scale, which is 1/2/3. The equation from StandartScale (SS) uses the following equation (1) [29]:

$$x_{standart} = \frac{x - mean(x)}{standart\ deviation\ (x)} \quad (1)$$

Standard deviation using the equation (2) as follows [29]:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (2)$$

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

Step 7: Formation (n) Component PCA

Principal Component Analysis (PCA) is a commonly used dimension reduction technique in signal processing. PCA looks for projection matrices that minimize Mean Squared Error (MSE) between the reduced dataset and the original dataset [30], [31]. PCA has a wide range of relevant implementations in signal processing and data analysis.

Step 8: Determination of Variance Ratio and Eigenvalue

The variance ratio is a measure of the distribution of data. The equation to determine the variance ratio can use the following equation (3):

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{(n-1)} \quad (3)$$

Equation (3) has the understanding that the distribution of data has a certain size that can determine the amount of data distribution. While eigenvalue is a value that occupies a place in the eigenvector in the form of a matrix [29].

Step 9: Generate Visualization 2D and 3D Graph of Data Normalization Value Based on (n) Component

Visualization is very important in representing a result, especially in PCA results [32]. By using a 3-dimensional (3D) graph to represent it into a range of values whose results illustrate the results in a 2-dimensional (2D) image.

Step 10: Creating Labels Based on Graphic Results

Almais et al. have researched the coordinate value (n) on the PCA graph. Based on these studies, we can interpret the results as follows [9]:

1. If the coordinate value PC1 (n) is in the range ($n < 0$), then the data result label PC1 is lightly damaged.
2. If the coordinate value PC1 (n) is in the range ($0 \leq n < 2$), then the PC2 data result label is moderately corrupted.
3. If the coordinate value PC1 (n) is in the range ($n \geq 2$), then the data result label PC1 is severely damaged.

Step 11: Result and Validation

Results are values that have gone through a certain process. To ensure the correctness of the results, validation is required [33], [34]. Validation plays an important role in determining whether the results obtained conform to existing requirements [35]. According to *Byabazaire et al*., validation of results can be done by testing two different data sets: data trust matrix with model building. The correlation between these two records will give the smallest error value [36]. Here are the steps to validate the results:

1. Visualizing PCA results data: visualize PCA result data in 2 dimensions and 3 dimensions, especially PC1 and PC2 data.
2. Visualizing original target data that has passed surveyor validation: visualize original target data that has passed surveyor validation.
3. Comparing Visualization Results: Compare 2-dimensional visualization results from PC1 based on original target data with 2-dimensional visualization results from PC2 based on original target data.

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

Algorithm: ***GLCM to convert post-disaster building images into feature values***

Input:

- Input data (image) using function **imread** (**'namafile.jpg'**)
- Convert data (image) to Gray Scale using function **rgb2gray**

Process:

- Determining distances with values of 1 and 0.5
- Define angles with values 0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, and 315^0
- Extract images into feature values with the **greycomatrix** function
- Calculate feature values for contrast, dissimilarity, homogeneity, energy, and correlation using the **greycoprops** function

Output:

Displays contrast, dissimilarity, homogeneity, energy, and correlation results using the **print** library

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

Input:

Data input using function **pd.read_csv** (**'namafile.csv'**)

Process:

- Data normalization using the **StandardScaler()** and **fit_transform()** functions
- Generate variance ratio using function **explained_variance_ratio_()**
- Generate eigenvalue using **function explained_variance()**
- Generate PCA components using the **PCA()** function
- Generate dataFrame results of the PCA component using function **pd.DataFrame()**

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

- Visualization of normalization results using **scatterplot()**

Output:
The result of visualization of labeling data using PCA using **scatterplot()**

B. DATA PREPARATION

Using image data of post-natural disaster sector damage on the open data site, namely Kaggle, you can use data directly from camera photos, and or Google search engine search results. In Kaggle there is data of approximately 100 data on building damage after natural disasters that have been classified according to the type of natural disaster. In this study, for all types of natural disasters, and to determine the level of building damage after natural disasters. It does not focus on victims of natural disasters and sectors other than buildings affected by natural disasters.

IV. RESULT

The results and discussion explain the GLCM-PCA process in determining the structure of building damage after natural disasters and test the results with data from experts or surveyors to prove that the combined GLCM-PCA can determine the structure of building damage after natural disasters.

A. Image Reading and Conversion to Grey Scale (GS)

Reading images is the first step to data input using the "**imread**" Python library. The data is in the form of an image that will be converted into a GS image. If the image is already GS then the image will be converted into matrix values based on the GS image results using the Python library "rgb2gray". The image that enters the GLCM process must be in the form of matrix values of the GS form. FIGURE 2 is an illustration of the process of inputting images into matrix values.

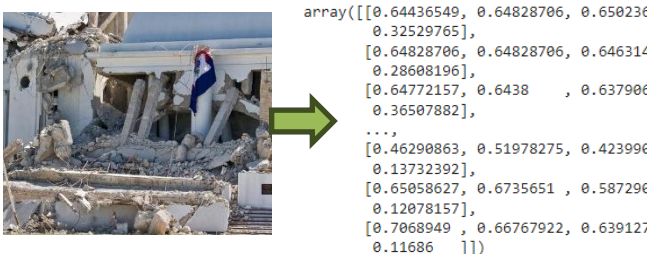


FIGURE 2. The process of converting images to matrix values.

B. GLCM Distance and Angel Value

Use 10 combinations of angle changes and 2 distance values to find out at what angle and distance to get optimal results. Every angle and distance goes through the GLCM and PCA analysis stages to determine damage to building structures after natural disasters. The distance is usually set at 1 pixel but can be smaller or greater than 1 pixel but in this study, it uses 1 and 0.5 for the distance value while the direction uses angles 0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, 315^0. Changes to angle data and distance values are found in TABLE II.

TABLE II
CHANGES IN GLCM ANGLE AND DISTANCE VALUES

Corner (^{0})	Distance	
	1	2
0 45 90 135 180 225 270 315	1	0.5

C. Calculate GLCM and Extract GLCM Result Into Multiple Features

The matrix value derived from GS image conversion is then combined with several angle values and distance values in TABLE II. To generate multiple feature values using the "**greycomatrix**" function. The result will then be the unique value present in each GLCM feature. For the source pieces used in the process of calculating and extracting GLCM into multiple features contained in FIGURE 3 the following:

```
glcm = greycomatrix(gray_image.astype(int),levels=1000,
                    distances=distances, angles=angles,
                    symmetric=True, normed=True)
```

FIGURE 3. Calculate GLCM and extract GLCM result into multiple features

D. GLCM Feature Results and Convert to Spreadsheet

The result of calculating and extracting matrix values using GLCM is a value of each feature in GLCM based on the angle used. This study used 5 types of features, 8 changes in angle values, and 2 distance values as shown in TABLE III. For this type of feature use contrast, dissimilarity, homogeneity, energy, and correlation. The values of each feature based on the resulting angle and distance values are in TABLE III.

TABLE III
GLCM FEATURE VALUE RESULT

Feature Name	Corner (^{0})	Distance	Feature Value
Contrast	0	1	0.000104373239
		0.5	0.00010437
	45	1	0.0000930075743
		0.5	0.
	90	1	0.000110317919
		0.5	0.00011032
	135	1	0.00009882
		0.5	0.
	180	1	0.000104373239
		0.5	0.00010437
	225	1	0.0000930075743

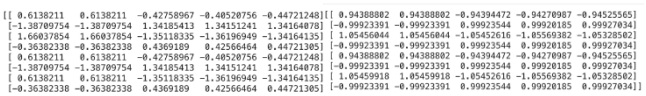
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

Feature Name	Corner (^{0})	Distance	Feature Value
Dissimilarity	0	0.5	0.
		1	0.000104373239
	45	0.5	0.00010437
		1	0.0000930075743
	90	0.5	0.
		1	0.000110317919
	135	0.5	0.00011032
		1	0.00009882
	180	0.5	0.
		1	0.000104373239
	225	0.5	0.00010437
		1	0.0000930075743
	270	0.5	0.
		1	0.000104373239
	315	0.5	0.00011032
		1	0.00009882
		0.5	0.
Homogeneity	0	1	0.99994781
		0.5	0.99994781
	45	1	0.9999535
		0.5	1.
	90	1	0.99994484
		0.5	0.99994484
	135	1	0.99995059
		0.5	1.
	180	1	0.99994781
		0.5	0.99994781
	225	1	0.9999535
		0.5	1.
	270	1	0.99994484
		0.5	0.99994484
	315	1	0.99995059
		0.5	1.
Energy	0	1	0.99989273
		0.5	0.99989273
	45	1	0.99989828
		0.5	0.99994498
	90	1	0.99988969
		0.5	0.99988969
	135	1	0.99989537
		0.5	0.99994498
	180	1	0.99989273
	225	0.5	0.99989273
		1	0.99989828
		0.5	0.99994498
	270	1	0.99988969
		0.5	0.99988969
	315	1	0.99989537
		0.5	0.99994498
Correlation	0	1	0.0525793895
		0.5	0.0525793895
	45	1	0.157848230
		0.5	1.
	90	1	-0.000055162
		0.5	-0.0000551620021
	135	1	0.105213745
		0.5	1.
	180	1	0.0525793895
		0.5	0.0525793895
	225	1	0.157848230
		0.5	1.
	270	1	-0.000055162
		0.5	-0.0000551620021
	315	1	0.105213745
		0.5	1.

E. Data Normalization

Normalizes the value of the features in TABLE III into a value that balances the values in all features. The result of data normalization is data that can be transformed into balanced data by PCA standards. The standard of PCA is that normalized data can produce an optimal number of PCs. FIGURE 4 (a) and (b) are the results of normalization data based on GLCM distances.



(a) (b)

FIGURE 4. Data Normalization Results: (a) Distance 1; (b) Distance 0.5

F. Formation (n) Component PCA and Determination of Variance Ratio

Dividing normalized data into PC is an analysis process that determines the results of PCA analysis. The right number of PCs can determine the optimal analysis results. In their research, Almais et al. determined PC can use a range of eigenvalue values or variance ratios [9]. In this study, 2 PCs were used because 2 PCs had a stable variance ratio value and were optimal enough to analyze the results. The results of trials using different distances, namely 1 and 0.5, produce different variance ratio values. For the variance ratio result value based on the distance difference, the GLCM result can

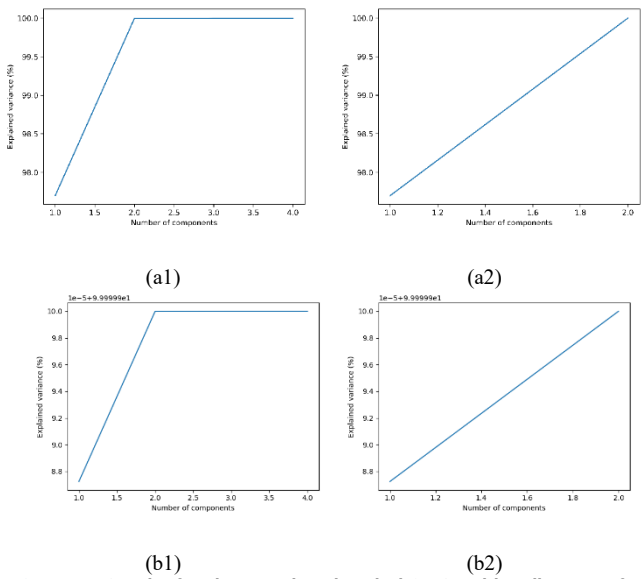
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

be seen in TABLE IV, while the form of the graph of the variance ratio value in each PC based on the GLCM distance value is in FIGURE 5.

TABLE IV
CHANGES IN GLCM ANGLE AND DISTANCE VALUES

Distance	PC	Value Variance Ratio
1	1	97.82%
	2	2.1%
	3	0.0%
0.5	1	100.0%
	2	0.0%
	3	0.0%



(a1) (a2)

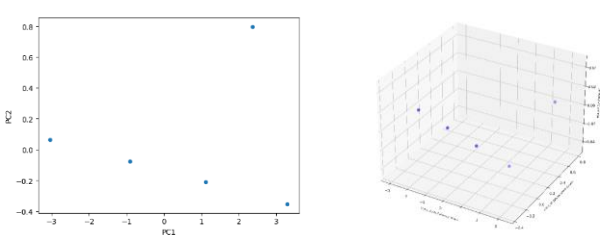
(b1) (b2)

FIGURE 5. Graph of variance ratio value: (a1) 3 PCs with a distance of 1; (a2) 2 PCs with a distance of 1; (b1) 3 PCs with a distance of 0.5; (b2) 2 PCs with a distance of 0.5.

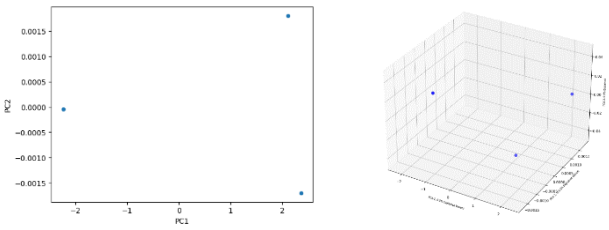
FIGURE 5 is a visual form of the variance ratio value at each GLCM distance value. FIGURE 5(a1) and (b1) is a visualization of the variance ratio values at GLCM distances of 1 and 0.5 with the number of PCs as much as 3. FIGURE 5 (a2) and (b2) is a visualization of variance ratio data at GLCM distances of 1 and 0.5 using 2 PCs.

G. Generate Visualization 2D and 3D Graph of Data Normalization Value Based on (n) Component

After determining the number of PCs based on the variance ratio value, the next step is visualizing 2-dimensional (2D) and 3-dimensional (3D) data on the PC, so that it is clear the data conflict that exists on each PC. FIGURE 6 is a 2D and 3D visualization of 2 PCs at each GLCM distance value.



(a1) (a2)



(b1) (b2)

FIGURE 6. Visualization of data distribution on a PC: (a1) 2D data visualization distance 1; (a2) distance 3D data visualization 1; (b1) 2D data visualization distance 0.5; (b2) 0.5 distance 3D data visualization.

FIGURE 6 (a1) and (a2) are forms of 2D and 3D visualization for PC data distribution using a distance at GLCM of 1. FIGURE 6 (b1) and (b2) are 2D and 3D visualizations for PC data distribution using a GLCM safe distance of 0.

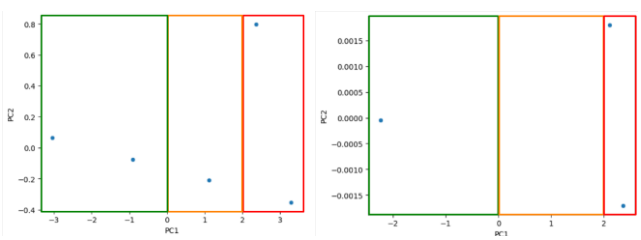
H. Creating labels based on graphic results

The results of the visualization of the distribution of PC data in FIGURE 6 can find out the information if it means using the standard value of the level of damage to buildings after natural disasters that already exist in their research *Almais et al.* explained that there are 3 levels of damage to buildings after natural disasters, namely lightly damaged, moderately damaged, and severely damaged [9]. Each of these levels has a range of values (n) as in TABLE V.

TABLE V
RANGE VALUE (N) LEVEL OF DAMAGE TO POST-DISASTER BUILDING [9]

Range of Values (n)	Damage Rate
$n < 0$	Lightly Damaged
$0 \leq n < 2$	Moderate Damaged
$n \geq 2$	Heavily Damaged

The visualization of the distribution of PC data in FIGURE 6 can be known by clustering the coordinate points using the standard value of the level of damage to buildings after natural disasters in TABLE V. The result is that there are three clusters in the visualization of PC data distribution at every GLCM distance. More details are seen in FIGURE 7.



(a) (b)

FIGURE 7. Clustering results: (a) Distance 1; (b) Distance 0.5.

In FIGURE 7(a) and (b) there are 3 clusters based on color, namely red, orange, and green. Red means severely damaged, orange is moderately damaged, and green is lightly damaged. Each color has a range of values (n) according to TABLE V. The difference between FIGURE 7(a) and (b) is the application of GLCM distance, for FIGURE 7(a) uses a GLCM distance of 1 while FIGURE 7 (b) uses 0.5.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

The clustering results in FIGURE 7 can be seen in more detail by knowing the coordinate point values PC1 and PC2 at each angle and GLCM distance. TABLE VI is the value of each coordinate based on FIGURE 7 with a GLCM distance of 1 and 8 different GLCM angles. While in TABLE VII shows the value of each coordinate in FIGURE 7 using a GLM distance of 0.5 and an angle of 8 different GLCM angles. In addition, TABLES VI and VII also contain colors, range values, and labels that explain the level of damage to buildings after natural disasters.

TABLE VI
COORDINATE POINTS AND DATA LABELS WITH GLCM DISTANCE 1

Color	Corner (^{0})	Distance GLCM (pixels)	Coordinate Point Value		Range of Coordinate Values (n)	Label
			PC1	PC2		
Orange	0	1	1.1205093066092984	-0.21070433409751635	$0 \leq n < 2$	Moderate Damaged
Green	45		-3.0404339062487598	0.06375026570358637	$n < 0$	Lightly Damaged
Red	90		3.2968473779787715	-0.35309091014531 39	$n \geq 2$	Heavily Damaged
Green	135		-0.9115353902403686	-0.07522092293031846	$n < 0$	Lightly Damaged
Orange	180		1.1205093066092982	-0.21070433409751 61	$0 \leq n < 2$	Rusak Sedang
Green	225		-3.0404339062487598	0.06375026570358626	$n < 0$	Lightly Damaged
Red	270		2.3660726017808877	0.7974408927938105	$n \geq 2$	Heavily Damaged
Green	315		-0.9115353902403686	-0.07522092293031846	$n < 0$	Lightly Damaged

TABLE VII
COORDINATE POINTS AND DATA LABELS WITH A GLCM DISTANCE OF 0.5

Color	Corner (^{0})	Distance GLCM (pixels)	Coordinate Point Value		Range of Coordinate Values (n)	Label
			PC1	PC2		
Red	0	0.5	2.1107078754276776	0.0018027059719974018	$n \geq 2$	Heavily Damaged
Green	45		-2.2343575884662035	-0.0000485	$n < 0$	Lightly Damaged
Red	90		2.35798997551556	-0.0017045687207842654	$n \geq 2$	Heavily Damaged
Green	135		-2.2343575884662035	-0.0000485	$n < 0$	Lightly Damaged
Red	180		2.1107078754276785	0.0018027059719971869	$n \geq 2$	Heavily Damaged
Green	225		-2.2343575884662035	-0.0000485	$n < 0$	Lightly Damaged
Red	270		2.3580246274938 97	-0.0017066913292165524	$n \geq 2$	Heavily Damaged
Green	315		-2.2343575884662035	-0.0000485	$n < 0$	Lightly Damaged

V. DISCUSSION

The analysis of PCA results uses the GLCM results in FIGURE 2 with 2 changes in different GLCM distances, namely 1 and 0.5, resulting in a cluster of PC data distribution found in FIGURE 7. In FIGURE 7(a) using a GLCM distance of 1, 5 coordinate points are scattered. The details of the distribution of the 5 coordinate point values are found in TABLE VI, from the number of points that should be 8 coordinate points to 5 coordinates means that there are coordinate points with the same point values. From the 5 coordinate points seen in FIGURE 7, they can be clustered into 3 clusters of post-natural disaster damage levels, namely 2 values of the coordinate point entering the lightly damaged cluster (green), 1 value of the coordinate point entering the moderately damaged cluster (orange), and 2 values of the coordinate point entering the heavily damaged cluster (red). Meanwhile, by using a GLCM distance of 0.5, there are 4 coordinate points. Details for the coordinate point value are found in TABLE VII, from the number of points that should be 8 coordinate points in FIGURE 7(b) it can be seen that there are only 4 coordinate points because the other 4 coordinate points are the same. From the 4 coordinate points in the data cluster using a GLCM distance of 0.5, 3 clusters of building damage levels after natural disasters were produced, namely 1 coordinate point value in the lightly damaged cluster (green), no value at the coordinate point in the moderately damaged cluster (orange), and 3 coordinate values in the heavily damaged cluster (red). TABLE VIII is the total coordinate point data from the cluster analysis using PCA.

TABLE VIII
THE NUMBER OF COORDINATE POINTS OF PCA CLUSTERING RESULTS

GLCM distance (pixels)	Coordinate Point		Number of coordinate points	Color	Value Range (n)	Label
	PC1	PC2				
1	1.12050930660929 84	-0.21070433409751635	2	Orange	$0 \leq n < 2$	Moderate Damaged
	-3.04043390624875 98	0.06375026570358637	2	Green	$n < 0$	Lightly Damaged
	3.29684737797877 15	-0.35309091014531 39	1	Red	$n \geq 2$	Heavily Damaged
	-0.91153539024036 86	-0.07522092293031846	2	Green	$n < 0$	Lightly Damaged
	2.36607260178088 77	0.7974408927938105	1	Red	$n \geq 2$	Heavily Damaged
0.5	2.11070787542767 76	0.0018027059719974018	2	Red	$n < 0$	Heavily Damaged
	-2.23435758846620 35	-0.0000485	4	Green	$n < 0$	Lightly Damaged
	2.35798997551556	-0.0017045687207842654	1	Red	$n \geq 2$	Heavily Damaged
	2.35802462749389 7	-0.0017066913292165524	1	Red	$n \geq 2$	Heavily Damaged

FIGURE 8 is a visualization of the distribution of data in TABLE VIII which is a comparison of the values of PC1, PC2, and GLCM 1 coordinate points.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

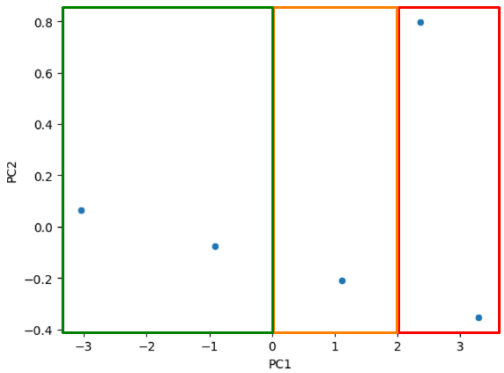


FIGURE 8. The distribution of data at the coordinates of PC1 and PC2 is based on the distance of GLCM 1

In FIGURE 9 is a visualization of the distribution of data in TABLE VIII which is a comparison of the values of PC1, and PC2 coordinate points and a GLCM distance of 0.5.

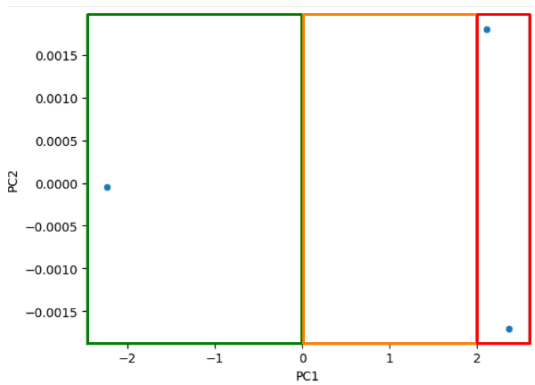


FIGURE 9. The distribution of data at the coordinates of PC1 and PC2 is based on the GLCM distance of 0.5

FIGURES 8 and 9 have 3 different color clusters, namely green, orange, and red. According to *Almais et al.* explained that clustering the level of building damage after a natural disaster has 3 types of colors, namely green (lightly damaged), orange (moderately damaged), and red (severely damaged) [9]. The results of FIGURES 8 and 9 are also the results of visualization of data distribution which has different meanings if in FIGURE 8 there are 2 green dots of coordinates, 1 orange dot of coordinates, and red dots of 2 coordinate points. Meanwhile, in FIGURE 9 there are 1 coordinate point in green, 0 coordinate points in orange, and 2 coordinate points in red. So FIGURE 2 means lightly damaged and severely damaged because the number of coordinate points in the cluster of lightly damaged and severely damaged buildings is the same, namely 2 coordinate points. In FIGURE 9, it is understood that in FIGURE 2 the level of damage to the building is severely damaged because the number of coordinate points in the cluster of the damage to the disaster building is more than in other clusters.

Validation based on existing studies can be used to validate the results of FIGURES 8 and 9 by the actual situation. The following are the validation results of several studies that address the same thing but with different models or completion techniques.

1. According to *Bachriwindi et al.* explained that to determine the level of damage to buildings after natural disasters, 5 criteria can be used, namely 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 [2]. Then the 5 criteria are included in the process of the Multi-Criteria Decision Making (MCDM) method, namely Weighted Product (WP) to find out the alternatives. There are 3 types of alternatives, namely lightly damaged, moderately damaged, and severely damaged. Validating FIGURE 2 based on their research by *Bachriwindi et al.* is to know the alternatives of the image based on 5 existing criteria with a value range of 1 (lightly damaged), 2 (moderately damaged), and 3 (severely damaged) [2]. FIGURE 2 is based on the first criterion, namely the condition of the building has a scale of interest, a heavy scale of interest, a scale rating collapsed, and a damage value of 3. FIGURE 2 is based on the criteria for the condition of the building that the building has collapsed. The second criterion is the condition of the building structure has a scale of interest weight, the scale rating most buildings are damaged, and the damage value is 3. In FIGURE 2, based on the criteria for the state of the building structure, the structure of the building (columns, walls) has been mostly damaged. The third criterion is that the physical condition of the building has a scale of interest weight, a scale rating of >50%, and a damage value of 3. So in FIGURE 2 based on the criteria for the physical condition of the building, the building has collapsed or collapsed by more than 50%. The fourth criterion is that the function of the building has a heavy scale of interest, a dangerous scale rating, and a damage value of 3. In FIGURE 2, based on the criteria for building functions, it has been seen as dangerous for residents affected by natural disasters because many building structures have collapsed and collapsed. The fifth criterion is that other supporting conditions have a heavy scale of interest, a scale rating of total damage, and a damage value of 3. Based on the criteria of other supporting circumstances, FIGURE 2 has been completely damaged because the roof of a large area has collapsed, and the pillars, walls, doors, and roof have also collapsed or collapsed. In his research *Bachriwindi et al.* after all the values of building damage in each criterion are known, the next step is to determine alternative levels of building damage using the WP method. After the WP method process, the value of the 3 alternatives is known, then the value of the 3 alternatives is ranked to find out the highest score of each alternative to get the optimal alternative [2]. In this case, the highest alternative value is the heavily damaged alternative.
2. According to *Almais et al.*, using the same method as his research, *Bachriwindi et al.* but different from another type of MCDM method, namely TOPSIS [37]. Based on their research, *Almais et al.* the results of FIGURE 2 are included in the heavily damaged alternatives, because the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

TOPSIS method is one of the methods of MCDM, so if in their research *Bachriwindi et al.* using the WP method produces heavy damage, then the TOPSIS method also produces heavily damaged alternatives. Only the final result value of each alternative will be different on the WP and TOPSIS methods but the ranking of the alternatives is the same.

3. Based on their research, *Almais et al.* used a prediction model by applying the Neural Network (NN) method to predict the level of building damage after natural disasters [38], in FIGURE 2 including the types of clusters of building damage levels after natural disasters that are severely damaged. According to *Almais et al*., if the 5 criteria for determining the level of damage to buildings after a natural disaster have a value of 3, then the meaning is heavy damage [38]. The process in *Almais et al.'s* research to intelligently assess the level of building damage after natural disasters is to apply the Neural Network Forwardpropagation (NNFP) method to predict the level of building damage after natural disasters using training data derived from the TOPSIS method process. Then the training data was tested using testing data based on new data from surveyors.

From the validation results of 3 existing studies, valid validation results have been obtained that FIGURE 2 is included in the cluster of severely damaged buildings. So the results of the surveyor's validation are by the results of the PCA analysis using the GLCM distance of 0.5 because, in 2D and 3D visualization, the results of the PCA analysis at the GLCM distance of 0.5 produce more or dominant coordinate point values in the cluster of the level of damage to severely damaged buildings (in red). From the results of the study, it has been proven to find information on the level of building damage after a natural disaster on a building with images that can be lighter and faster using the integration of the GLCM-PCA method compared to the CNN method. Because by using the GLCM-PCA method integration, it does not require a lot of data and high computing. As in his research, *Kumar et al.* needed a lot of data and high computing because they applied CNN for brain segmentation classification with image data of 514 images for 233 patients and 3064 images for 73 patients using a 25-layer CNN model [39].

Meanwhile, in the integration process of the GLCM-PCA method, it is possible to find out the information of an image by looking for the feature values in each image using GLCM, then the feature values of each post-natural disaster building image enter the PCA process to find out the distribution of feature values on the x and y axis graphs. To analyze the results of PCA, we can use the reference standard for the coordinate value of the post-disaster building damage level in the research of *Almais et al.* which explained that there are 3 clusters of post-natural disaster building damage levels, each of which has its coordinate value [9]. The results of the PCA analysis in the form of labels and clustering coordinate values for the level of building damage after natural disasters are found in TABLE VIII Based on the results of the study, it is known that to find out the information in an image in addition to using the CNN method which requires a lot of data and high computing, it is possible to use the integration of the GLCM-PCA method.

However, to implement the integration of the GLCM-PCA method, it is necessary to first have a clear standard in the previous study that explains the standard coordinate values in each cluster using PCA in the same case, to get maximum results and get the actual image information.

VI. CONCLUSION

Determining the level of damage to buildings after natural disasters using image input of buildings affected by disasters can use the integration of digital image analysis techniques, namely GLCM and PCA. GLCM images in the form of numerical data can be known information using image analysis techniques, namely PCA. PCA can cluster a data set based on its parameters or attributes to know the information. As a result, images that enter the GLCM process are distinguished into 2, namely images using distances of 1 and 0.5. In the PCA process, GLCM results using different distances also have different results. From the PCA analysis using 2D and 3D visualization the GLCM results using distance 1 produce different results from the results of the surveyors, but by using the GLCM distance 0.5 the results are the same as the surveyors, which are severely damaged. If you look at the results of the variance ratio value in the PCA process using GLCM distances of 1 and 0.5, it produces a stable variance ratio value in the PCA process using GLCM 1 distance. However the PCA clustering results do not match the surveyor, that is, the distribution of data is identical to all clusters, so they cannot know the information in the picture. Based on this, the results of PCA analysis techniques are not only seen from the stability of the variance ratio value but the results of 2D and 3D visualization of the distribution of data from PCA analysis can find out the information stored in a digital image input. So in the future to find out the information on an image can use the integration of image analysis techniques, namely GLCM and PCA. In the future, it is possible to develop the integration of GLCM and PCA not only to determine the level of damage to buildings after natural disasters but also to find information from a picture in all areas of life. In addition, it can develop in terms of integration with the CNN method but must be prepared with adequate high computing needs. For further research development by adding image augmentation techniques to improve image quality and avoid potential biases such as varying image quality, types of natural disasters, and different building materials.

ACKNOWLEDGMENT

The author would like to thank the research partners for their contributions that indirectly contributed to the implementation of this research, especially Universitas Brawijaya Malang which has provided facilities to explore the ideas and process

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

of forming this publication. So that we can find out the integration of image analysis techniques, namely GLCM and PCA to determine the level of post-disaster building damage. In addition, for the Database Laboratory of the Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang which has helped provide facilities for research to produce a publication.

REFERENCES

[1] A. H. Safitri, A. T. W. 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, Jun. 2022, doi: 10.31961/eltikom.v6i1.532.

[2] 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. 12019, Jun. 2019, doi: 10.1088/1742-6596/1413/1/012019.

[3] V. Wildemann, "Experimental Research Methods of Mechanical Behavior of Structural Materials Under Complex Thermomechanical Influences," 2021, pp. 469–480. doi: 10.1007/978-981-15-9121-1_32.

[4] Mohammad Aljanabi, "Navigating the Landscape: A Comprehensive Bibliometric Analysis of Decision-Making Research in Civil Engineering," *Mesopotamian Journal of Civil Engineering*, vol. 2023, May 2023, doi: 10.58496/MJCE/2023/005.

[5] B. Aklouche, T. Benkedjouh, H. Habbouche, and S. Rechak, "Damage assessment of composite material based on variational mode decomposition and BiLSTM," *The International Journal of Advanced Manufacturing Technology*, vol. 129, no. 3–4, pp. 1801–1815, Nov. 2023, doi: 10.1007/s00170-023-12371-4.

[6] A. Pierorazio, N. E. Cherolis, M. Lowak, D. J. Benac, and M. T. Edel, "Assessment of Damage to Structures and Equipment Resulting from Explosion, Fire, and Heat Events," in *Analysis and Prevention of Component and Equipment Failures*, ASM International, 2021, pp. 134–145. doi: 10.31399/asm.hb.v11A.a0006804.

[7] U. Ajay Shah, "A review on structural integrity assessment procedures," *International Journal of Structural Integrity*, vol. 5, no. 4, pp. 328–338, Nov. 2014, doi: 10.1108/IJSI-01-2014-0004.

[8] Y.-L. Zhou, C. Hongyou, N. Zhen, and M. Abdel Wahab, "Review on structural damage assessment via transmissibility with vibration-based measurements," *J Phys Conf Ser*, vol. 842, p. 012016, May 2017, doi: 10.1088/1742-6596/842/1/012016.

[9] A. T. W. Almais *et al.*, "Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters," *IEEE Access*, p. 1, 2023, doi: 10.1109/ACCESS.2023.3275852.

[10] J. Camacho Navarro, M. Ruiz, R. Villamizar, L. Mujica, and J. Quiroga, "Features of Cross-Correlation Analysis in a Data-Driven Approach for Structural Damage Assessment," *Sensors*, vol. 18, no. 5, p. 1571, May 2018, doi: 10.3390/s18051571.

[11] O. Caselles, A. Martín, Y. F. Vargas-Alzate, R. Gonzalez-Drigo, and J. Clapés, "Estimators for Structural Damage Detection Using Principal Component Analysis," *Heritage*, vol. 5, no. 3, pp. 1805–1819, Jul. 2022, doi: 10.3390/heritage5030093.

[12] P. Y. Siow, Z. C. Ong, S. Y. Khoo, and K.-S. Lim, "Damage Sensitive PCA-FRF Feature in Unsupervised Machine Learning for Damage Detection of Plate-Like Structures," *International Journal of Structural Stability and Dynamics*, vol. 21, no. 02, p. 2150028, Feb. 2021, doi: 10.1142/S0219455421500280.

[13] S. Tomer, K. Khanna, S. Gambhir, and M. Gambhir, "Comparison Analysis of GLCM and PCA on Parkinson's Disease Using Structural MRI," *International Journal of Information Retrieval Research*, vol. 12, no. 1, pp. 1–15, Oct. 2021, doi: 10.4018/IJIRR.289577.

[14] P. D. Ogunjinmi, S.-S. Park, B. Kim, and D.-E. Lee, "Rapid Post-Earthquake Structural Damage Assessment Using Convolutional Neural Networks and Transfer Learning," *Sensors*, vol. 22, no. 9, p. 3471, May 2022, doi: 10.3390/s22093471.

[15] V. Ahmadian, S. B. Beheshti Aval, M. Noori, T. Wang, and W. A. Altabey, "Comparative study of a newly proposed machine learning classification to detect damage occurrence in structures," *Eng Appl Artif Intell*, vol. 127, p. 107226, Jan. 2024, doi: 10.1016/j.engappai.2023.107226.

[16] A. Fernandez-Navamuel, F. Magalhães, D. Zamora-Sánchez, Á. J. Omella, D. Garcia-Sanchez, and D. Pardo, "Deep learning enhanced principal component analysis for structural health monitoring," *Struct Health Monit*, vol. 21, no. 4, pp. 1710–1722, Jul. 2022, doi: 10.1177/14759217211041684.

[17] B. Huang, B.-H. Koh, and H. S. Kim, "PCA-based damage classification of delaminated smart composite structures using improved layerwise theory," *Comput Struct*, vol. 141, pp. 26–35, Aug. 2014, doi: 10.1016/j.compstruc.2014.05.011.

[18] OECD, *PISA 2018 Results (Volume I): What Students Know and Can Do*. OECD, 2019. doi: 10.1787/5f07c754-en.

[19] C. Chen, J. Wei, C. Peng, and H. Qin, "Depth-Quality-Aware Salient Object Detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 2350–2363, 2021, doi: 10.1109/TIP.2021.3052069.

[20] M. Benco, R. Hudec, P. Kamencay, M. Zachariasova, and S. Matuska, "An Advanced Approach to Extraction of Colour Texture Features Based on GLCM," *Int J Adv Robot Syst*, vol. 11, no. 7, p. 104, Jul. 2014, doi: 10.5772/58692.

[21] D. Srivastava, B. Rajitha, S. Agarwal, and S. Singh, "Pattern-based image retrieval using GLCM," *Neural Comput Appl*, vol. 32, no. 15, pp. 10819–10832, Aug. 2020, doi: 10.1007/s00521-018-3611-1.

[22] Z. Indra and Y. Jusman, "Performance of GLCM Algorithm for Extracting Features to Differentiate Normal and Abnormal Brain Images," *IOP Conf Ser Mater Sci Eng*, vol. 1082, no. 1, p. 012011, Feb. 2021, doi: 10.1088/1757-899X/1082/1/012011.

[23] P. Mohanaiah, P. Sathyanarayana, L. Gurukumar, and A. Professor, "Image Texture Feature Extraction Using GLCM Approach," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, 2013, [Online]. Available: www.ijsrp.org

[24] S. K. P.S and D. V.S, "Extraction of Texture Features using GLCM and Shape Features using Connected Regions," *International Journal of Engineering and Technology*, vol. 8, no. 6, pp. 2926–2930, Dec. 2016, doi: 10.21817/ijet/2016/v8i6/160806254.

[25] D. Brennan, "Simple export of journal citation data to Excel using any reference manager," *J Med Libr Assoc*, vol. 104, no. 1, pp. 72–75, Jan. 2016, doi: 10.3163/1536-5050.104.1.012.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

[26] I. Dinc, M. Sigdel, S. Dinc, M. S. Sigdel, M. L. Pusey, and R. S. Aygun, "Evaluation of normalization and PCA on the performance of classifiers for protein crystallization images," in *IEEE SOUTHEASTCON 2014*, IEEE, Mar. 2014, pp. 1–6. doi: 10.1109/SECON.2014.6950744.

[27] M. A. Salama, A. E. Hassanien, and A. A. Fahmy, "Reducing the influence of normalization on data classification," in *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, IEEE, Oct. 2010, pp. 609–613. doi: 10.1109/CISIM.2010.5643523.

[28] 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.

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

[30] S. Wang, Y. Wang, Y. Chen, P. Pan, Z. Sun, and G. He, "Robust PCA Using Matrix Factorization for Background/Foreground Separation," *IEEE Access*, vol. 6, pp. 18945–18953, 2018, doi: 10.1109/ACCESS.2018.2818322.

[31] G. D. Pelegrina and L. T. Duarte, "A Novel Approach for Fair Principal Component Analysis Based on Eigendecomposition," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 3, pp. 1195–1206, Mar. 2024, doi: 10.1109/TAI.2023.3298291.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] J. Byabazaire, G. M. P. O'Hare, and D. T. Delaney, "End-to-End Data Quality Assessment Using Trust for Data Shared IoT Deployments," *IEEE Sens J*, vol. 22, no. 20, pp. 19995–20009, Oct. 2022, doi: 10.1109/JSEN.2022.3203853.

[37] A. T. W. Almais *et al.*, "SDDS: Damage Level Determination System for Post-Natural Disaster Sector Based on Building Characteristics," in *The 18th IMT-GT International Conference on Mathematics, Statistics and their Applications*, Sciendo, 2024, pp. 23–28. doi: 10.2478/9788367405713-005.

[38] A. T. W. Almais *et al.*, "SASSD: A Smart Assessment System For Sector Damage Post-Natural Disaster Using Artificial Neural Networks," in *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, Aug. 2023, pp. 96–101. doi: 10.1109/COSITE60233.2023.10249540.

[39] S. Kumar and D. Kumar, "Human brain tumor classification and segmentation using CNN," *Multimed Tools Appl*, vol. 82, no. 5, pp. 7599–7620, Feb. 2023, doi: 10.1007/s11042-022-13713-2.

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

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, Universitas 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.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4

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.2024.3469637

ALAMSYAH MUHAMMAD JUWONO obtained his bachelor's degree from Gajah Mada University, Indonesia, in 1987. He obtained a master's degree from the University of Tasmania, in 1997, and a doctorate (Ph.D) from Queensland University of Technology, in 2015. He is also a lecturer at the Department of Physics, Universitas Brawijaya Malang. His research interests include Theoretical Physics and Geophysics.

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.

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 Universitas Brawijaya, Malang. His research interests include Computational Intelligence, Control Systems, and Robotics.

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

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4