Journal of Information Technology and Cyber Security 3(1) January2025: 33-44 EISSN 2987-386X | PISSN 2987-3878
DOI: 10.30996/jitcs.12270

Clustering of Post-Disaster Building Damage Levels Using Discrete Wavelet Transform and Principal Component Analysis

Putri Purnamasari [1], Mochamad Imamudin [2], Syahiduz Zaman [3], A'la Syauqi [4], and Agung Teguh Wibowo Almais [5,*]

[1,2,3,4] Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia
[5] Department of Electrical Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia
* Corresponding author: agung.twa@ti.uin-malang.ac.id

Received: 02 December 2024
Accepted: 18 January 2025
Revised: 18 January 2025
Available online: 28 January 2025

To cite this article: Purnamasari, P., Almais, A. T. W., Imamudin, M., Zaman, S., & Syauqi, A. (2025). Clustering of Post-Disaster Building Damage Levels Using Discrete Wavelet Transform and Principal Component Analysis. *Journal of Information Technology and Cyber Security, 3*(1), 33-44. https://doi.org/10.30996/jitcs.12270

Abstract

Damage assessment of buildings after natural disasters is generally performed manually by a team of experts at the disaster site, making it prone to human error and resulting in low accuracy in classifying the level of damage. This research aims to develop a more efficient and accurate method in post-disaster building damage assessment by integrating Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) techniques. The main contribution of this research is the use of DWT as well as the application of this method on more than one image to improve the accuracy of damage level classification. A total of nine unlabelled images of post-disaster buildings were used in this study, which were obtained from the Regional Disaster Management Agency or Badan Penanggulangan Bencana Daerah (BPBD) of Malang City, Indonesia. The methods applied include data pre-processing, DWT decomposition for image analysis to identify features, and clustering using PCA to cluster the level of building damage into light, medium, and heavy categories, which are then evaluated based on accuracy. The results showed that the method yielded 100% accuracy with validation results from surveyors, as evidenced through 2D and 3D visualisations based on principal components (PC1-PC3). These findings confirm that the integration of DWT and PCA can be an effective alternative in improving the accuracy of post-disaster building damage assessment, as well as supporting decision-making in rehabilitation and reconstruction after natural disasters.

Keywords: clustering of building damage, Discrete Wavelet Transform, post-disaster building damage assessment, Principal Component Analysis.

1. Introduction

Natural disasters often cause significant damage to building infrastructure, resulting in large material losses and heavy casualties (Almais, et al., 2024; Hadiguna, Kamil, Delati, & Reed, 2014). In 2023, the National Disaster Management Agency or Badan Nasional Penanggulangan Bencana (BNPB) recorded more than 5,400 natural disaster events in Indonesia, further emphasizing the urgency of developing faster and more accurate building damage assessment methods (Rosyida, et al., 2024). Currently, post-disaster building damage assessment is generally carried out manually by a team of experts at the disaster site. The manual method used by BNPB to determine the level of damage to buildings after natural disasters is still subjective and prone to human error (Almais, et al., 2023; Amri, et al., 2016), resulting in a low level of accuracy in classifying the level of damage to buildings after natural disasters (Fu'adah, Almais, & Syauqi, 2023; Krichen, Abdalzaher, Elwekeil, & Fouda, 2024). Therefore, a novel, more objective and reliable approach is needed to address this issue.

This research aims to overcome this problem by utilizing digital image processing technology, which can offer a more efficient and accurate solution in building damage assessment. The object of this research is images of buildings damaged by natural disasters in Malang City, Indonesia, taken by the Malang City Regional Disaster Management Agency or Badan Penanggulangan Bencana Daerah (BPBD). The selection of this location is based on the complexity and diversity of the types of damage that occur, which presents its own challenges in the process of identifying and classifying the level of damage.

Several previous studies have developed methods for post-natural disaster building damage assessment using image processing techniques, such as Gray-Level Co-occurrence Matrix (GLCM) combined with Principal Component Analysis (PCA), which showed good ability in clustering damage levels with high variance in the resulting data (Almais, et al., 2024). In the study by Biswal, Sanyal, & Mohapatra (2022), Discrete Wavelet Transform (DWT) is combined with deep learning methods for damage detection, and achieved an accuracy of 94.8%. However, these methods tend to be limited to certain dataset and have not been able to provide optimal accuracy on high contrast images.

While some of these studies have made significant contributions, there is a major gap that needs to be addressed, namely the lack of methods that can handle images with high noise and sharp contrast variations. In addition, there is still a lack of providing an objective and automatic classification of damage levels. This research aims to fill the gap by combining DWT and PCA, which is expected to overcome the noise problem and improve accuracy and objectivity in post-disaster building damage grouping.

The main contribution of this research is the development of a more efficient building damage assessment system, utilizing a combination of DWT and PCA. This system is not only expected to reduce reliance on manual assessment, but also offer a more objective approach in damage classification based on digital image analysis. This research also has the potential to provide new insights into the application of image processing technology in the context of natural disaster mitigation and management.

2. Literature Review

2.1. Discrete Wavelet Transform (DWT)

The Wavelet Transform method is a signal processing technique that allows analysis in two domains, time and frequency. This technique has been widely used in image processing because of its ability to isolate specific features at various frequency levels, which is very useful for the extraction of important information from digital images (Huda, Karyatanti, & Dewantara, 2019). In this study, Discrete Wavelet Transform (DWT) is chosen to reduce noise in images of buildings damaged by natural disasters and to extract important features that indicate the level of damage. Previous research by Biswal, Sanyal, & Mohapatra (2022) showed that the use of DWT in image processing can improve the accuracy of damage detection in post-disaster building structures, especially in images that have high noise levels.

DWT works by breaking the image into several frequency subbands using low-pass and high-pass filtering processes. This process produces several image components, such as approximation components and detail components, which can be analyzed in more depth (Khairuna, Sriani, & Triase, 2017). This research uses the Daubechies wavelet type because of its ability to handle unsymmetrical signals, which are often found in images of buildings with heavy damage due to natural disasters.

2.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of data while retaining the greatest variance of the original data. PCA works by transforming multivariate data into a number of uncorrelated principal components, known as principal components (Christaki, et al., 2022). In this study, PCA is applied to classify the level of building damage based on features that have been extracted using DWT.

Study conducted by Almais, et al. (2023) shows that PCA can be effective in classifying building damage data into several categories, such as light, medium, and heavy damage, with a high level of accuracy. The PCA process begins with data normalization, then proceeds with the calculation of the covariance matrix to determine the variance of each component. The main component selection is done based on the obtained Eigenvalue, which is then used to classify the level of building damage.

2.3. Wavelet and PCA Transformation Integration

The combination of Wavelet Transform and PCA has been proven effective in improving the accuracy of digital image processing, especially in the classification of building damage after natural disasters. Wavelet Transform is used to extract important features from images by isolating certain frequency patterns, while PCA is applied to reduce the data dimension and group the processed data into certain categories (Almais, et al., 2024). In addition of being able to reduce large-dimensional data, PCA can also classify data (Tazi, Isnaini, Mutmainnah, & Ainur, 2019; Troccoli, Cerqueira, Lemos, & Holz, 2022). In this research, a combination of these two methods is used to build a system that is able to automatically categorize the level of building damage based on available images. The cluster results are then validated by a team of experts to ensure optimal accuracy.

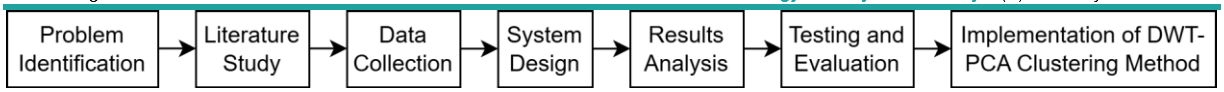


Fig. 1. Research stages.

The use of the Wavelet Transform helps to separate important details from the image such as the edges of damaged buildings, while PCA is used to perform clustering based on the principal components that represent the damage. By combining these two methods, this research is expected to provide more accurate results compared to manual methods that are often prone to subjectivity and error.

2.4. Decision Support System (DSS)

This research also focuses on developing a Decision Support System (DSS) based on image processing technology. The developed DSS utilizes the results of image processing from the DWT and PCA methods to provide recommendations regarding repair priorities and resource allocation in dealing with post-disaster building damage (Putra, Hermawan, & Hatmoko, 2020). Research conducted by Junaidi (2019) shows that the application of DSS based on image processing technology can increase the effectiveness and efficiency in decision making, especially in emergency situations such as handling natural disasters. The DSS proposed in this study is designed by considering the flexibility and accuracy in classifying the level of damage, so that it can be adapted to various types of natural disasters. The system is expected to reduce reliance on manual assessment methods, while improving objectivity and consistency of assessment results across different disaster locations and conditions.

3. Methods

3.1. Problem Identification

Post-disaster rehabilitation and reconstruction in Indonesia is managed by BNPB and BPBD, which play a role at the national and regional levels respectively. BPBD Malang City, as one of the institutions handling post-disaster building damage assessment, faces challenges in producing accurate assessments. The assessment process carried out by the assessment team often relies solely on estimates, without clear reference criteria. It makes the assessment results vulnerable to being influenced by differences in viewpoints between assessors.

Fig. 1 shows the research stages starting from problem identification to result analysis. After the problem identification stage, the research continues with a literature study to obtain a theoretical and methodological basis that will support the design and implementation of the system in the next stage.

3.2. Literature Studies

At the literature study stage, the researcher refers to various studies that are relevant to this topic. Some of the reviewed studies applied similar methods, but in different cases, while other studies discussed the same case or object with different method approaches. The selected literature remains focused on methods and objects related to this research, in order to support the development of a comprehensive system.

3.3. Data Collection

This research uses secondary data in the form of 9 (nine) images of post-disaster building conditions obtained from BPBD Malang City. These images document various types of building damage due to natural disasters. Each image is accompanied by information such as the location of the incident, the type of building, and the level of damage observed, which facilitates further analysis. There are 3 levels of damage to buildings after natural disasters, namely light, medium and heavy damage, based on the observations of the surveyor team.

This secondary data is an important fundamental of this research as it provides in-depth insight into post-disaster building damage patterns. With structured visual documentation and accurate supplementary information, this research was able to utilize the data to analyze the causes and severity of damage more comprehensively. This analysis not only helps identify damage patterns, but also contributes to the development of more objective methods in the assessment of building damage due to natural disasters.

3.4. System Design

This subsection describes the workflow of DWT and PCA in digital image processing. This process includes the initial stages, such as image conversion to grayscale and filter application, to the final stage, which is principal component analysis to produce output in the form of graphs and labels. Fig. 2 presents an illustration that maps the relationship between processes in the system, providing an overall depiction of the integration of DWT and PCA in image data processing.

3.4.1. Image preprocessing

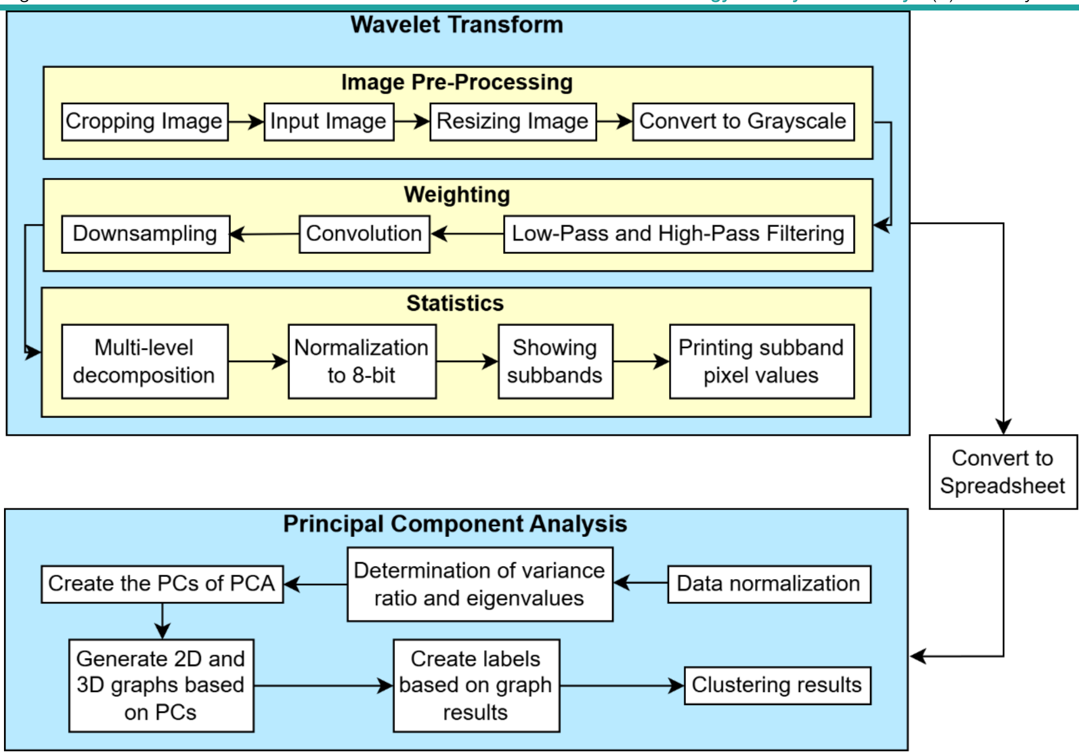


Fig. 2. System design.

(a) (b)

Fig. 3. (a) Original image before cropping and (b) Cropped image.

Once the cropping process is complete, the image is fed into the system for further processing. The image is then resized to 128×128 pixels for standardization, as the smaller size makes the decomposition process more efficient in terms of time and memory usage. Next, the image was converted to grayscale to simplify the data without losing important structural information. Fig. 3 shows the images before and after cropping, while Fig. 4 shows the image that has been converted to grayscale.

3.4.2. Weighting

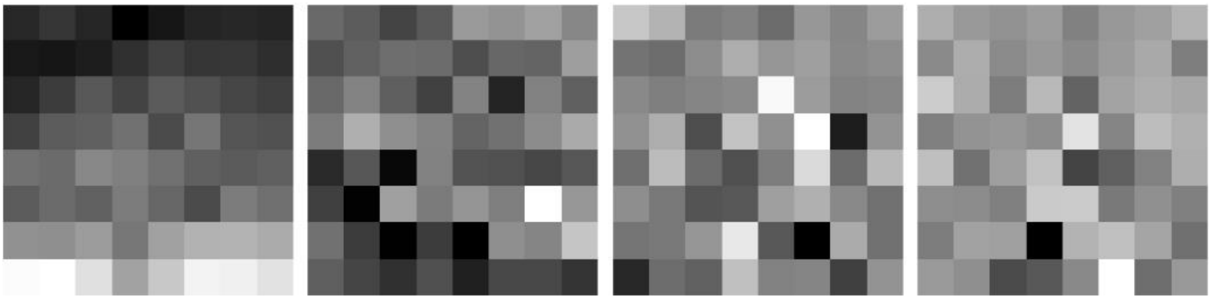
The grayscale image is processed with low-pass and high-pass filters that are weighted. For the low-pass filter, a weight of [0.5, 0.5] is used, while for the high-pass filter a weight of [-0.5, 0.5] is used. The low-pass filter serves to capture low-frequency information or coarse details of the image, while the high-pass filter is used to extract high-frequency details, such as image edges. Next, a convolution process is performed by applying these filters to the rows and columns of the image to extract important features. After that, downsampling of the rows and columns is performed to reduce the size of the data that needs to be processed in the next stage, resulting in a smaller image that retains the relevant information.

3.4.3. Statistics

The image is decomposed using wavelet transform into several subbands (LL, LH, HL, & HH). The LL subband, which represents the basic structure of the image, is used for PCA analysis because it contains

(a) (b)

Fig. 4. (a) Resized image (128×128 pixels) and (b) Grayscale image.



(a) (b) (c) (d)

Fig. 5. Level 4 Wavelet decomposition results: (a) Approximation (LL), (b) Horizontal details (LH), (c) Vertical details (HL), and (d) Diagonal details (HH).

Table 1
Grouping of the level of building damage after natural disasters.

Damage Rate	Level	Subband
Heavy	2	4
Moderately	3	3
Minor	4	2

low-frequency information. This decomposition allows the system to separate horizontal, vertical and diagonal detail elements, as shown in Fig. 5, which shows the results of the visual representation of level 4 wavelet decomposition. In this study, different decomposition levels are used to categorize the level of building damage, as presented in Table 1.

For the category of heavy damage, decomposition levels 2 and 4 subbands were used; for medium damage, decomposition levels 3 and 3 subbands; and for light damage, decomposition levels 4 and 2 subbands. With this approach, each damage category can be analyzed more precisely based on the details generated from each decomposition level.

After decomposition of the processed grayscale image, the matrix values of each subband (LL, LH, HL, HH) are saved in an Excel file to facilitate further analysis. Each subband is saved in a separate sheet. Tables 2, 3, 4 and 5 show the results of extracting the matrix values for each subband. These four subbands provide different and complementary information about the analyzed image. Each subband has 8 parameters, namely 0, 1, 2, 3, 4, 5, 6, and 7, resulting from feature extraction using the wavelet transform. The value of each subband is different, as shown in Table 2 to Table 5. By utilizing the values of all these subbands, assessment of building damage levels can be done more efficiently and accurately, as each subband provides a different perspective on the condition of the building. The use of Wavelet Transform allows for a more in-depth and detailed analysis, thus increasing the effectiveness in post-disaster damage assessment.

Table 2
Matrix approximation (LL).

0	1	2	3	4	5	6	7
41	54	38	0	22	37	39	37
25	22	29	48	66	54	55	46
38	61	88	67	91	82	70	63
66	92	96	113	75	118	85	82
113	107	136	128	113	96	91	96
92	107	97	124	103	75	124	112
145	143	156	119	161	178	179	171
252	255	224	162	200	243	240	226

Table 3
Matrix horizontal details (LH).

0	1	2	3	4	5	6	7
105	92	68	89	154	147	161	137
80	98	110	108	78	103	100	158
105	132	95	65	131	37	130	96
124	175	143	131	100	113	140	170
42	87	8	130	83	81	71	87
64	2	160	123	149	127	255	151
113	60	1	60	0	143	133	197
95	69	49	80	33	72	72	52

Table 4
Matrix vertical details (HL).

0	1	2	3	4	5	6	7
199	180	123	130	108	151	134	156
115	109	141	174	150	159	136	142
137	127	133	138	249	156	132	134
145	174	78	196	144	255	29	148
11	188	108	80	125	218	108	185
142	122	84	88	159	176	137	114
117	122	150	232	88	0	171	114
41	109	97	192	130	133	64	148

Table 5
Matrix diagonal details (HH).

0	1	2	3	4	5	6	7
174	153	146	156	129	152	160	180
138	173	139	152	138	155	169	126
204	172	125	185	99	163	174	167
128	148	151	141	227	133	189	176
191	113	159	194	68	96	130	174
142	140	128	202	203	121	144	128
125	162	160	0	189	192	165	112
154	143	75	88	137	255	110	153

3.4.4. Data normalization

Data normalization is an important step to ensure that each feature has the same scale, so that no feature dominates the PCA. In this study, data from the LL and LH subbands were selected because they contain important information about the image. These two subbands were then combined into a 1-dimensional (1D) representation to simplify the analysis process, as shown in Fig. 6 which illustrates the 1D normalization results. The normalization process is performed by subtracting the mean of each feature and dividing it by the standard deviation, so that the resulting data has a mean of zero and a variance of one. This step aims to improve the accuracy of PCA, which is sensitive to the scale of the data. Data normalization (x_{norm}) based Standard Deviation can be calculated based on Eq. (1), where x is the data value of the feature to be normalized, μ is the mean of the dataset (the average value of all data in one feature), and σ is the Standard Deviation of the dataset (a measure of the spread or dispersion of data values) (Ahn, et al., 2020).

$$x_{norm} = \frac{x-\mu}{\sigma} \tag{1}$$

```
Normalized Data:
[-1.127014 -0.890883 -1.181506 -1.871735 -1.472128 -1.19967  -1.163342
 -1.19967  -1.417637 -1.472128 -1.344981 -0.999867 -0.672916 -0.890883
 -0.872719 -1.036195 -1.181506 -0.763736 -0.27331  -0.654752 -0.218818
 -0.382294 -0.600261 -0.727408 -0.672916 -0.200654 -0.127999  0.180788
 -0.509441  0.271607 -0.327802 -0.382294  0.180788  0.071804  0.598558
  0.453246  0.180788 -0.127999 -0.218818 -0.127999 -0.200654  0.071804
 -0.109835  0.380591 -0.000851 -0.509441  0.380591  0.162624  0.762033
  0.725705  0.961836  0.289771  1.052656  1.361442  1.379606  1.234295
  2.705572  2.760063  2.196982  1.070819  1.761048  2.542096  2.487605
  2.23331   0.035476 -0.200654 -0.636588 -0.255146  0.925508  0.798361
  1.052656  0.580394 -0.418621 -0.091671  0.126296  0.089968 -0.454949
 -0.000851 -0.055343  0.998164  0.035476  0.525902 -0.146163 -0.69108
  0.507738 -1.19967   0.489574 -0.127999  0.380591  1.30695   0.725705
  0.507738 -0.055343  0.180788  0.671213  1.216131 -1.10885  -0.291474
 -1.726423  0.489574 -0.36413  -0.400458 -0.582097 -0.291474 -0.709244
 -1.835407  1.034492  0.362427  0.834689  0.435082  2.760063  0.871016
  0.180788 -0.7819   -1.853571 -0.7819   -1.871735  0.725705  0.544066
  1.706556 -0.146163 -0.618424 -0.981703 -0.418621 -1.272325 -0.563933
 -0.563933 -0.927211]
```

Fig. 6. 1D normalization result.

Table 6
Results of variance ratio and eigenvalue of each major component.

PC	Variance Ratio (%)	Eigenvalue (%)
PC1	65.27	11.07
PC2	15.22	2.58
PC3	9.91	1.68
PC4	4.15	0.7
PC5	2.60	0.44
PC6	2.09	0.35
PC7	0.77	0.13
PC8	0.00	0.00

3.4.5. Determination of variance ratio and eigenvalue

In the stage of determining the variance ratio and Eigenvalue, the initial step is to calculate the covariance matrix of the normalized data to understand the correlation between features. This covariance matrix is then used to determine the direction of the largest variance in the data. Eigenvalues are calculated to measure the amount of variance in each principal component, as illustrated in Eq. (2) (Jolliffe, 2002; Molin, 2021; Susilo, Isnanto, & Riyadi, 2020), where s^2 is the variance that measures how dispersed the data is compared to the average. The larger the variance value, the more dispersed the data is. $\sum_{i=1}^{n}(x_i - \mu)^2$ is the sum of squares of the difference between each data value (x_i) and the average (μ), which is used to calculate the variance. n is the amount of data, and $n-1$ is used to correct for bias in calculating the sample variance to make it more accurate. If the data being calculated is the entire population, the denominator is n. Meanwhile, the Eigenvectors give direction to the principal components.

Table 6 shows the variance ratio and Eigenvalue for each principal component. In Table 6 there are variance ratio values and Eigenvalues from PC1 to PC8. According to (Almais, et al., 2023), if the Eigenvalue of the main component (PC) is greater than 0, then that component is the most efficient and optimal for use in data clustering. In Table 6, PC1 to PC3 have Eigenvalues greater than 0, which indicates that the three PCs are the most efficient and optimal for data analysis.

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1} \quad (2)$$

3.4.6. Creating the Principal Components (PC)

The next step is the creation of Principal Components (PCs) in PCA. Principal components are calculated based on the Eigenvalues that represent the largest variance in the data, as well as the Eigenvectors that have been obtained. These components are the most important and significant features, used to summarize the main information of the data. PCs are linear combinations of the original features in the data, with weights determined by the Eigenvalues. This process ensures that the first principal component (PC1) has the greatest contribution in explaining the variance, followed by PC2, PC3, and so on. PCA retains only a few PCs that explain most of the variance, so the dimensionality of the data can be

```
PCA Components (2D):
[[-4.642417  0.293644  0.882196 -0.899682 -0.09826   0.184437 -0.559476
  -0.       ]
 [-1.533552  0.217857  1.02948  -0.238611  0.55866   0.743108  0.615132
  -0.       ]
 [ 0.071951 -0.095431  0.84439   0.852138  1.026016 -0.876605 -0.115774
  -0.       ]
 [ 6.766506 -0.465249 -0.205337 -0.893591  0.301976  0.259563 -0.211725
  -0.       ]
 [ 0.381532 -0.919384 -0.263672  1.53004  -0.422424  0.750592 -0.232461
  -0.       ]
 [ 1.402184 -0.079889  1.4359   -0.07177  -1.190908 -0.597141  0.250224
  -0.       ]
 [-0.067595  3.379009 -1.629291  0.128485 -0.130648 -0.108211  0.055896
  -0.       ]
 [-2.37861  -2.330556 -2.093665 -0.407009 -0.044412 -0.355743  0.198185
  -0.       ]]
```

Fig. 7. Result of Principal Components.

Table 7
Coordinate point (n) level of damage to buildings after natural disasters (Almais, et al., 2023).

Coordinate Point (n)	Damage Level
$n < 0$	Minor damage
$0 \leq n < 2$	Moderately damaged
$n \geq 2$	Heavy damage

reduced without losing important information. The results of these principal components can be seen in Fig. 7. The equation for obtaining PCA data ($data_{PCA}$) is presented in Eq. (3), where x_{norm} is the normalized data and $eigenvector\_subset$ is a subset of the relevant Eigenvectors (Jolliffe, 2002).

$$data_{PCA} = x_{norm} \times eigenvector\_subset \quad (3)$$

3.4.7. 2D and 3D graphics visualization from PC

The next stage is to visualize the results of the PCA analysis. After the data is normalized and the principal component (PC) values are calculated, the data is projected into the principal component space (PC1, PC2, and so on). At this stage, the data that has been reduced in dimension is visualized in the form of 2D and 3D graphs to facilitate the identification of patterns and clusters in the data. Fig. 8 shows the visualization results in the form of 2D and 3D graphs.

The 2D graph uses two coordinate points, PC1 and PC2, while the 3D graph uses three coordinate points: PC1, PC2, and PC3. This visualization maps data points based on the value of each PC, which illustrates the distribution of data related to the level of building damage after a natural disaster.

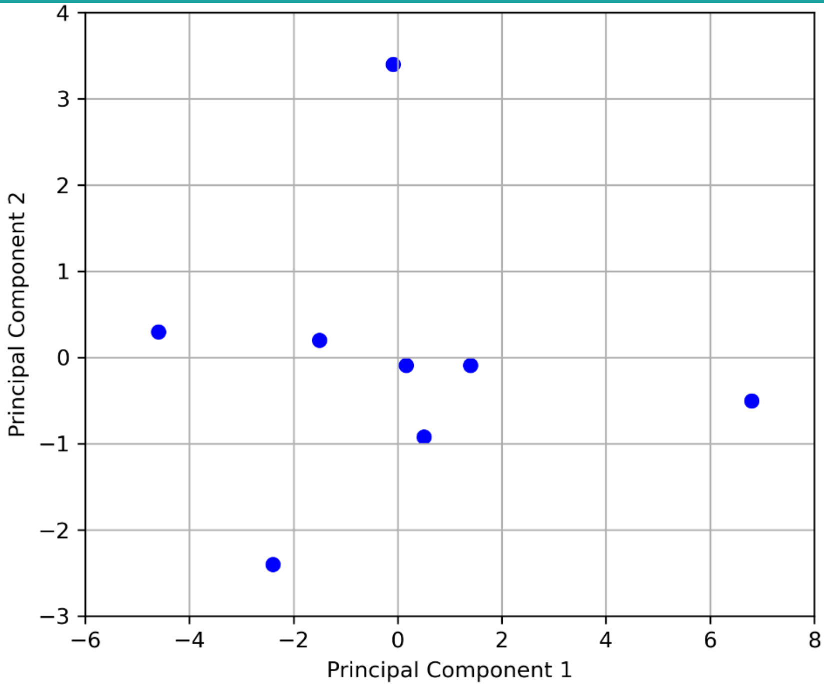
3.4.8. Create labels based on graph results

Based on 2D or 3D graph visualization, the data is grouped and labeled according to the level of building damage (light, medium, heavy) based on the position of the coordinate points on the graph in Fig. 8. The determination of the standard post-disaster building damage level is adapted from the research of Almais, et al. (2024), which uses data analysis of coordinate points on the main component (PC) in PCA to classify the level of building damage. The damage category is determined as follows: If the coordinate value (n) is below 0 ($n < 0$), then the building is categorized as lightly damaged; if n is between 0 and 2 ($0 \leq n < 2$), then the damage category is moderate; and if n is greater than or equal to 2 ($n \geq 2$), then the building is categorized as severely damaged (Almais, et al., 2023). Table 7 shows the standard values for the level of damage to buildings after natural disasters.

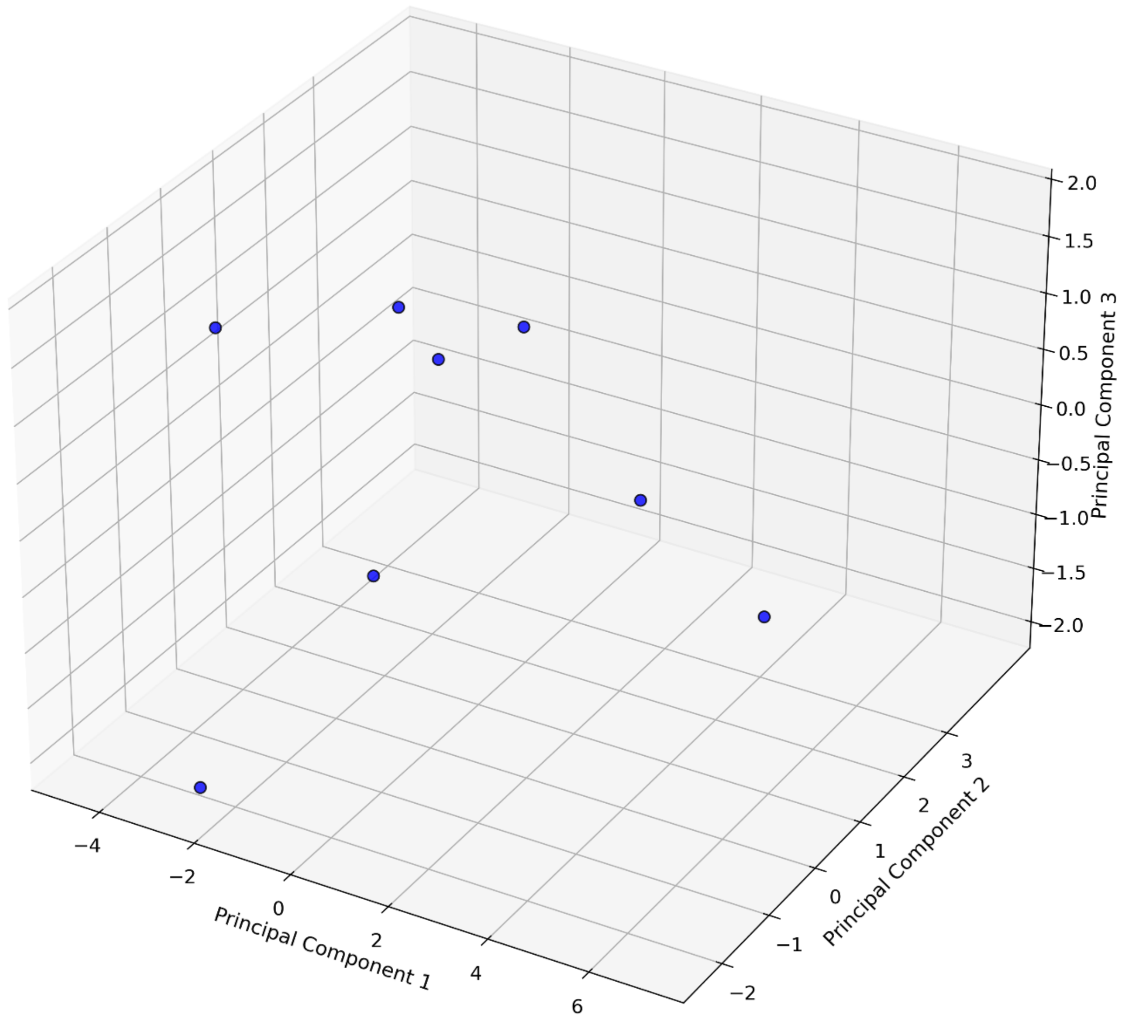
Fig. 8, which is translated based on the standards in Table 7, shows the distribution of coordinate points on PC1 and PC2 for the 2D graph. In Fig. 9, four coordinate points (n) in green is below 0 ($n < 0$), three coordinate points (n) in yellow have values between 0 and 2 ($0 \leq n < 2$), and one coordinate point (n) in red are above 2 ($n > 2$). Therefore, Fig. 9 shows that image 3 depicts a building with a light degree of damage (lightly damaged), as most of the coordinate points (n) is at a value $n < 0$, which corresponds to the category of minor damage according to the standards shown in Table 7.

3.4.9. Results

The data that has been analyzed using PCA produces groupings that can be used to assess the level of damage to buildings after a disaster, while retaining important information from the original data.

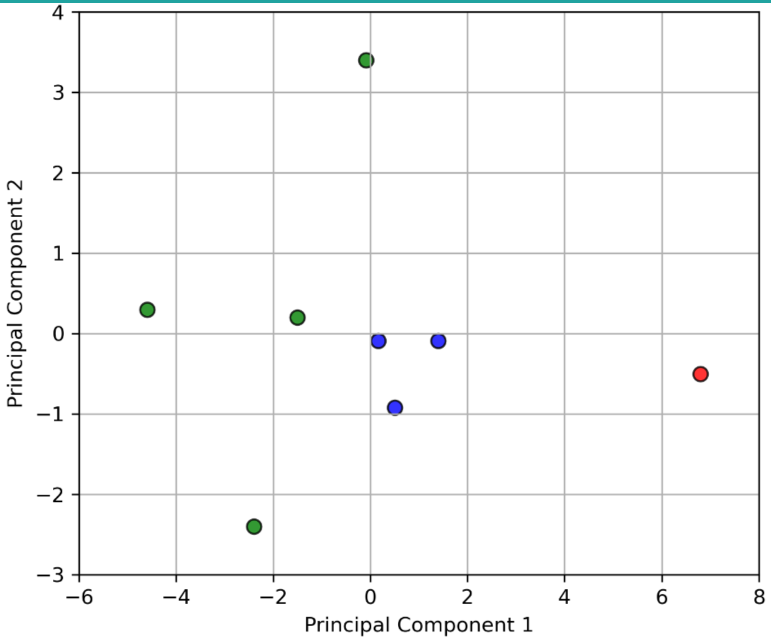


(a)

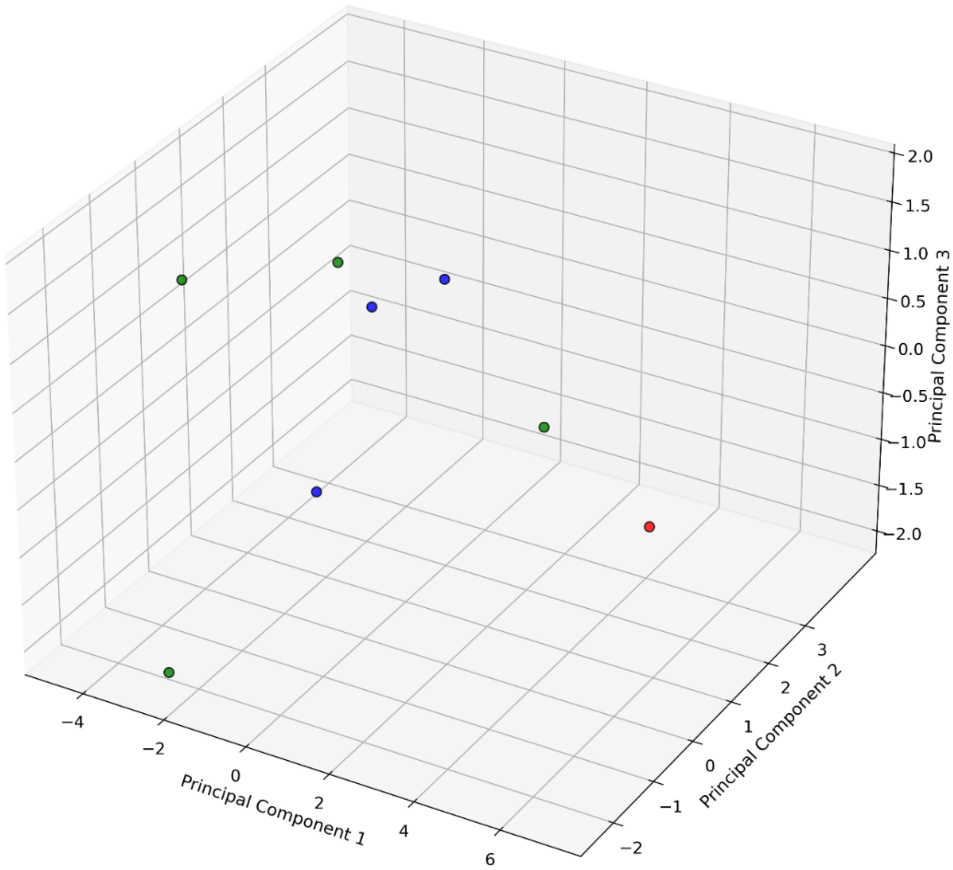


(b)

Fig. 8. PCA visualization of subband features: (a) 2D PCA and (b) 3D PCA.



(a)



(b)

Fig. 9. Clustering results: (a) 2D PCA visualization, (b) 3D PCA visualization.

4. Results and Discussion

This research successfully categorizes the damage level of buildings after natural disasters with high accuracy using DWT and PCA methods. With this approach, the developed system showed consistent and effective results in separating damage levels into three clear categories: lightly damaged, moderately damaged, and heavily damaged.

One important aspect of this research is external validation, which is conducted by comparing the system results with the manual assessment conducted by a team of surveyors from BPBD Malang City. The validation results show that the developed system produces a grouping (Fig. 9) that matches human judgement with a very high accuracy rate of 100%. This proves that the applied method is reliable in assessing damage to buildings after natural disasters with very high accuracy.

This research not only succeeded in producing an accurate classification of damage levels, but also showed the great potential of the system in supporting decision-making for post-disaster building rehabilitation and reconstruction. The system can be an effective tool in accelerating the evaluation and recovery process, providing faster and more accurate information compared to manual methods that require a lot of time and money.

However, the system requires further development, including further research using a larger and more diverse dataset covering different types of disasters and building conditions. This will ensure that the system can be adapted for more complex situations and damage types. In addition, system development could focus on integration with other technologies, such as real-time monitoring systems, which could support more efficient damage assessment in the field. The use of more advanced methods for feature extraction and data clustering can also improve the accuracy and ability of the system to handle more complex data, thus expanding its application in the disaster and infrastructure fields.

5. Conclusions

This research successfully developed a system to classify the level of post-disaster building damage using the DWT and PCA methods. This system is able to classify the level of damage to buildings after natural disasters with high accuracy, which is in accordance with the manual assessment of the BPBD Malang City surveyor team. The results are consistent, reliable, and provide accurate information, critical in the post-disaster building rehabilitation and reconstruction process. The applied method not only improves the speed and efficiency of assessment, but also supports more informed and data-driven decision-making. The external validation results show a very high concordance with the manual assessment, making it an effective alternative to replace the time-consuming and costly manual method.

6. CRediT Authorship Contribution Statement

Putri Purnamasari: Data curation, Formal analysis, Investigation, Visualization, Project administration, and Methodology. **M. Imamudin:** Supervision, and Writing - review & editing. **Syahiduz Zaman:** Review & editing. **A'la Syauqi:** Software, and Writing - review & editing. **Agung Teguh Wibowo Almais:** Conceptualization, Resources, Supervision, Validation, Methodology, Funding, Initial draft writing, and Writing - review & editing.

7. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Data Availability

Data will be made available on request.

9. References

Ahn, J., Ji, S.-H., Ahn, S. J., Park, M., Lee, H.-S., Kwon, N., . . . Kim, Y. (2020). Performance evaluation of normalization-based CBR models for improving construction cost estimation. *Automation in Construction, 119*. doi:https://doi.org/10.1016/j.autcon.2020.103329

Almais, A. T., Susilo, A., Naba, A., Sarosa, M., Crysdian, C., Tazi, I., . . . Hendro, W. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters. *IEEE Access, 11*, 74590 - 74601. doi:https://doi.org/10.1109/ACCESS.2023.3275852

Almais, A. T., Susilo, A., Naba, A., Sarosa, M., Juwono, A. M., Crysdian, C., . . . Wicaksono, H. (2024). Characterization of Structural Building Damage in Post-Disaster Using GLCM-PCA Analysis Integration. *IEEE Access, 12*, 146190-146201. doi:https://doi.org/10.1109/ACCESS.2024.3469637

Amri, M. R., Yulianti, G., Yunus, R., Wiguna, S., Adi, A. W., Ichwana, A. N., . . . Septian, R. T. (2016). *Risiko Bencana Indonesia (RBI).* Jakarta, Indonesia: Badan Nasional Penanggulangan Bencana. Retrieved from https://inarisk.bnpb.go.id/pdf/Buku%20RBI_Final_low.pdf

Biswal, P. R., Sanyal, B., & Mohapatra, R. K. (2022). Building Damage Detection Using Discrete Wavelet Transforms and Convolutional Neural Networks. *Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications. 2nd*, pp. 743–753. Jaipur, India: Springer, Singapore. doi:https://doi.org/10.1007/978-981-16-6332-1_60

Christaki, M., Vasilakos, C., Papadopoulou, E.-E., Tataris, G., Siarkos, I., & Soulakellis, N. (2022). Building Change Detection Based on a Gray-Level Co-Occurrence Matrix and Artificial Neural Networks. *Drones, 6*(12), 414. doi:https://doi.org/10.3390/drones6120414

Coburn, A. W., Spence, R. J., & Pomonis, A. (1994). *Vulnerability and Risk Assessment: Disaster Management Training Programme* (2nd ed.). Cambridge, United Kingdom: UNDP Disaster Management Training Programme. Retrieved from https://books.google.co.id/books/about/Vulnerability_and_Risk_Assessment.html?id=zS24mQEACAAJ

Fu'adah, A. F., Almais, A. T., & Syauqi, A. (2023). Klasifikasi Tingkat Kerusakan Sektor Pasca Bencana Alam Menggunakan Metode MULTIMOORA Berbasis Web. *JISKA (Jurnal Informatika Sunan Kalijaga), 8*(1), 222-230. doi:https://doi.org/10.14421/jiska.2023.8.3.222-230

Hadiguna, R. A., Kamil, I., Delati, A., & Reed, R. (2014). Implementing a web-based decision support system for disaster logistics: A case study of an evacuation location assessment for Indonesia. *International Journal of Disaster Risk Reduction, 9*, 38-47. doi:https://doi.org/10.1016/j.ijdrr.2014.02.004

Huda, F. I., Karyatanti, I. D., & Dewantara, B. Y. (2019). Identifikasi Gangguan Belitan Stator Motor Induski Metode Wavelet. *Prosiding Seminar Nasional Sains dan Teknologi. 10.* Semarang, Indonesia: Fakultas Teknik Universitas Wahid Hasyim. doi:https://doi.org/10.36499/psnst.v1i1.2901

Jolliffe, I. T. (2002). *Principal Component Analysis.* New York, United States: Springer. doi:https://doi.org/10.1007/b98835

Junaidi, M. (2019). *Sistem pendukung keputusan untuk rehabilitasi dan rekonstruksi fisik aksi setelah bencana alam menggunakan metode wp-topsis.* Malang, Indonesia: Universitas Islam Negeri Maulana Malik Ibrahim. Retrieved from http://etheses.uin-malang.ac.id/id/eprint/16610

Khairuna, K., Sriani, S., & Triase, T. (2017). Pendekomposisian Citra Digital dengan Algoritma DWT. *Algoritma: Jurnal Ilmu Komputer dan Informatika, 1*(1). Retrieved from https://moraref.kemenag.go.id/documents/article/98021043410574412

Krichen, M., Abdalzaher, M. S., Elwekeil, M., & Fouda, M. M. (2024). Managing natural disasters: An analysis of technological advancements, opportunities, and challenges. *Internet of Things and Cyber-Physical Systems, 4*, 99-109. doi:https://doi.org/10.1016/j.iotcps.2023.09.002

Molin, S. (2021). *Hands-On Data Analysis with Pandas: A Python data science handbook for data collection, wrangling, analysis, and visualization.* Birmingham, UK: Packt Publishing.

Putra, I. S., Hermawan, F., & Hatmoko, J. U. (2020). Penilaian Kerusakan dan Kerugian Infrastruktur Publik Akibat Dampak Bencana Banjir di Kota Semarang. *Wahana Teknik Sipil: Jurnal Pengembangan Teknik Sipil, 25*(2), 86-97. doi:http://dx.doi.org/10.32497/wahanats.v25i2.2154

Rosyida, A., Maulani, M. A., Firmansyah, Y., Setiawan, T., Pangesti, K. P., & Ichsan, F. K. (2024). *Data Bencana Indonesia Tahun 2023* (Vol. 3). Jakarta, Indonesia: Pusdatinkom BNPB. Retrieved from https://bnpb.go.id/buku/buku-data-bencana-indonesia-tahun-2023

Susilo, V., Isnanto, R. R., & Riyadi, M. A. (2020). Herbal Leaf Pattern Analisis Using Principal Component Analisis (PCA) and Canberra Distance. *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). 7*, pp. 100-104. Semarang, Indonesia: IEEE. doi:https://doi.org/10.1109/ICITACEE50144.2020.9239235

Tazi, I., Isnaini, N. L., Mutmainnah, M., & Ainur, A. (2019). Principal Component Analysis (PCA) Method for Classification of Beef and Pork Aroma Based on Electronic Nose. *Indonesian Journal of Halal Research, 1*(1), 5-8. doi:https://doi.org/10.15575/ijhar.v1i1.4155

Troccoli, E. B., Cerqueira, A. G., Lemos, J. B., & Holz, M. (2022). K-means clustering using principal component analysis to automate label organization in multi-attribute seismic facies analysis. *Journal of Applied Geophysics, 198*. doi:https://doi.org/10.1016/j.jappgeo.2022.104555