Identification of earthquake types based on seismogram data

D N K Hardani*, D A Megapiengga, and M T Tamam

Department of Electrical Engineering, Faculty of Engineering and Sciences, Universitas Muhammadiyah Purwokerto. Indonesia

*Email: diannova.kh@ump.ac.id

Abstract. Indonesia is one country that has many active volcanoes, and it can have a high potential for earthquakes due to volcanic activity. Central Java is an area that has an active volcano, the Mount Slamet. Various kinds of earthquake signals, earthquake strength, and frequency are recorded using a seismogram. The process of recognizing seismic signal patterns using short-wave transformations has a higher chance of success. Several types of earthquakes that have occurred on the Mount Slamet. There are three types of earthquake data tested, Shallow Volcanic Earthquakes (VB), Gust, and Tremors. This study aims to identify the type of earthquake vibration signals recorded on the seismograph. Earthquake image processing system consists of several parts. The image is normalized to get the image in the time domain. Then the image is processed with two processes to determine the characteristics of the earthquake. The Fast Fourier Transform process is used to determine the strength of earthquake signals based on the frequency. The quantization process is used to take samples of each data in the time domain. In this study, the method used for identification is pattern recognition and decision trees. The identification system can recognize signals that are approached using the Root Mean Square, Average Power Spectrum, and statistical features. The results of tests carried out obtain 100% accuracy of each method.

1. Introduction

Indonesia is one country that has many active volcanoes, and it can have a high potential for earthquakes due to volcanic activity. Central Java is an area that has an active volcano, Mount Slamet. The area of Mount Slamet starting from the peak to its feet divided into five districts. The west-northwest sector includes the Brebes Regency area, the north sector including the Tegal Regency region, the northeast-southeast sector including the Purbalingga Regency region, and the south-southwest sector including the Banyumas Regency region [1]. The earthquake frequency of an area refers to the type and size of earthquakes experienced over a while. Various kinds of earthquake signals, earthquake strength, and frequency are recorded using a seismogram. Earthquake vibrations recorded on the seismograph can be identified to determine the category of the earthquake that occurred.

Seismogram data in monitoring is raw data, which contains various information about the observed condition. Short-wave switching is a method that can be used to present data and operator functions into different frequency components. The process of recognizing seismic signal patterns using short-wave transformations has a higher chance of success. With the transfer of this short-wave range, it is possible to locate time-frequency. Thresholding method aims to limit and eliminate parts of the signal.
that are considered not to contain much valuable information. By determining the value of the parameter data, then the restricted parts can be regarded as mixed noise [2].

The automatic classification process of seismic events is fundamental because of the large amount of data that is received continuously. Seismic signal analysis that classifies activities can do by visual inspection and calculation of signal characteristics. This process is subjective and requires significant hard work and time and experience. Reliable automated classification systems can reduce the effort needed and make classification faster and more objective. The approach taken for the classification process by developing an expert classification system based on fuzzy rules can imitate human reasoning and combine analyst knowledge of the classification of seismic events. The classification results on real seismic data show the robustness of the classifier and its ability to operate in on-line classification [3].

Detection and classification of seismic signals can also use an Artificial Neural Network (ANN) methods. A neural network that is trained to recognize actual seismic events makes it possible to save about 70% of computing time in an automatic detector. Automatic associative networks trained in real-time create an internal model of the input signal and errors made in reproducing the signal increase when some modification of the input signal prevents it from matching in the internal model [4]. The unsupervised (neural network) training method consists of a Self-Organizing Map (SMO). The SOM is constructed as a neural classifier and complementary reliability estimator to distinguish seismic events, and was employed for varying map sizes. The SOM achieved a discrimination reliability that could be employed routinely in observatory practice; however, about 6% of all events were classified as ambiguous cases [5]. But in other research prove that the resulting SOM map separated into different areas, each one containing the events of a defined type. It means that the SOM discriminates well the four classes of seismic signals. Moreover, the system will classify a new input pattern depending on its position on the SOM map. This approach can be an efficient instrument for the real-time automatic analysis of seismic data, especially in the case of possible volcanic unrest [6]. A computer-based classifier to automatically identify four seismic event classes of the Llaima volcano are implemented using support vector machines. Results indicate that the features used for recognition of the events of Villarica volcano also provide good recognition results for the Llaima volcano, yielding classification exactitude of over 80% [7].

The new approach combines nonparametric smoothing and classification techniques, which are applied directly to the seismic data, with different graphical representations of the intermediate steps introduced. For each sensor position, potential horizon locations identified along the corresponding time-series traces. These candidate locations then examined across all traces and when consistent patterns occur the points are linked together to form coherent horizons [8].

Classifying seismic signals into the appropriate type of volcanic earthquake is one of the most important tasks for monitoring volcanic activity. Such tasks must be carried out routinely. This activity causes a significant workload for personnel. Pattern recognition gives practitioners volcanic seismology theories and methods for designing classification systems with digital signal processing techniques. It will present promising and challenging opportunities for automatic identification of volcanic earthquakes. This study aims to identify the type of earthquake vibration signals recorded on the seismograph by comparing pattern recognition and decision trees analysis model.

2. Method
The morphology of Mount Slamet divided into old volcano morphology and young volcano morphology. Geological structures that develop in the Slamet Mountain and surrounding areas, generally in the form of normal faults that often founded in the Old Slamet group. Traces of these faults in the field based in the form of breccias, scratch lines, fault zones, straightness of hills and valleys, straight and steep escapes and sharp contact between rock units [1]. It is composed of bedrock, old volcano rock, and young volcano rock. It is also cut off by a particular fault line. The geological identification of Mount Slamet is carried out through field observations of the landscape, rocks, and
geological alignment [9]. The geological structure of an area can affect the occurrence of active volcanoes and are prone to earthquake disasters.

2.1. Signals database
The data used in this work to evaluate the classification system is data collected at Mount Slamet Monitoring Post, Center for Volcanology and Geological Disaster Mitigation Gambuhan, Pemalang. Signals recorded at three observation stations, namely Jurangmangu station, Blambangan station, and Gunung Cilik station. The type of earthquake recorded on Mount Slamet consists of volcanic A, volcanic B, tectonic, tremor, and gust earthquake. The gust earthquake and the tremor earthquake were earthquakes that dominated the seismograph recordings on Mount Slamet. In this study, based on recorded data, there are three types of earthquake signal data tested, shallow volcanic earthquakes (VB), the gust earthquake, and the tremor earthquake. Figure 1 shows one of the signals recorded by a seismogram from several Mount Slamet monitoring centers during an earthquake. In the spectrogram image, the X-axis shows the time of the earthquake, and the Y-axis represents the frequency.

![Figure 1](image1.png)

**Figure 1.** Example of seismic data (a) One of the signal data when the gust earthquake from several observation centers (b) A spectrogram of the gust earthquake.

2.2. Signal processing
The method adopted in this study can be divided into three main stages, pre-processing, feature extraction, and classification, as shown in Figure 2. This part explains each block in the identification system. Algorithms developed using Matlab, including signals handling, filtering, segmentation, resampling, normalization, and quantization.

![Figure 2](image2.png)

**Figure 2.** Block diagram of earthquake types identification system

2.2.1. Pre-processing.
Earthquake image processing system consists of several parts. The pre-processing step aims to prepare data that will be used for further processing. The process at this step consists of resampling and normalization. Resampling step is done by resampling existing data by cutting the data based on the
time to be processed and eliminating noise contained in the sample data. Then the data is converted into grayscale so that the color data that owned only has a value of 0 and 1. Resampling has the aim of the image cut following the period that can be processed to the next stage.

Normalization is the process of converting two-dimensional image data into one-dimensional signals. Normalization is done by calculating the height and width of the image and the DPI (Dot per Inch) of the image data. The goal of normalization is that the amplitude of the earthquake signal is in the region of 0 to 1, and the earthquake signal will be calculated based on the period.

Then the image is processed with two processes to determine the characteristics of the earthquake. The Fast Fourier Transform process is used to determine the strength of earthquake signals based on the frequency. The quantization process is used to take samples of each data in the time domain. Next, to find out the characteristics of an earthquake, use the Root Mean Square process which is used to calculate the average root FFT and the Average Power Spectrum used to calculate the average.

2.2.2. Feature extraction

Feature extraction aims to get specific characteristics from a signal. Optimal feature extraction will produce different characteristics for each class of data so that it will facilitate the identification process. Feature extraction has methods commonly used for recognition. The extraction method used for earthquake signal processing is Root Mean Square (RMS) and Average Power Spectrum (AVG). Root Mean Square (RMS) is a statistical measure of the size of a variable that varies. RMS is useful when there are positive and negative variations, for example sinusoidal signals. RMS is used in various fields and is most often used in the signal field. RMS is the averaged signal magnitude of each window and its equation is shown in equation (1). In the equation, \( x_i \) is the value of the \( i^{th} \) sample in the processing window and \( N \) is the window size [10]. RMS is related to the constant force and non-fatiguing contraction.

\[
RMS = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}
\]  

In the equation (1), \( x_i \) is the value of the \( i^{th} \) sample in the processing window, and \( N \) is the sample size, that is, the number of observations in the sample and FFT coefficients.

Average Power Spectrum is a process to measure the average power of a deterministic periodic signal. It is a type of continuous signal in the time domain that produces a discrete power spectrum. First of all, the signal filter to be extracted is processed by the Hamming Window method to avoid aliasing. Aliasing is undesirable in signal processing. Aliasing is a new signal where it has a different frequency from the original signal. This effect can occur because of the low number of sampling rates. To reduce the possibility of aliasing, it must go through a windowing process. The window function used in this system is the Hamming window. This process also serves to calculate the average power spectrum, other than that the noise produced is not too large. Equation (2) is the formula for the Hamming window method in sample M (\( \alpha=0.54 \)) [11, 12],

\[
w(n) = \{\alpha - (1 - \alpha)\cos\left(\frac{2\pi n}{M - 1}\right)\}, \quad 0 \leq n \leq M - 1
\]

where \( M \) is number of points in the output window. If zero or less, an empty array is returned. And \( w(n) \) is the window, with the maximum value normalized to 1. After the windowing process uses the Hamming window, converting these values into a logarithmic value use equation (3).

\[
AVG\ power = 10 \times \log_{10}\left(\frac{w(n)}{2}\right)
\]

Another method often used for feature extraction is statistical. This characteristic was chosen based on prior knowledge about the data. The series of features selected contain components: mean, standard deviation, correlation, variance, skewness, and kurtosis [13, 14].
2.3. Classification
The identification process used classification method when the whole process has completed. Each earthquake gets the characteristics and produces a value from each earthquake signal image, which arranged into a dataset. There are two approaches used to identify the type of seismogram signal, including the pattern recognition approach and decision tree analysis. The aim is to compare the two classifier's performance.

2.3.1. Pattern recognition system
Pattern recognition in the field of engineering study theories and methods for designing machines that can recognize patterns. Many techniques and methods in the field of pattern recognition are borrowed from other basic and applied disciplines such as Digital Signal Processing, statistics, and machine learning. The principle of pattern recognition in this study is a matching process that will state the process ends when \( dt \) is equal to the number sought, as equation (4). But, when the value found is not contained in the dataset that created, then \( dt \) cannot be the same as the amount requested. The process to be carried out is the approach with the smallest error. It will continue to work until \( p, dt \) and \( q \) are in one place.

\[
dt = \frac{(p + q)}{2}
\]  

where \( dt \) is the number sought, \( p \) is the initial value and \( q \) is the final value.

2.3.2. Decision tree models
In decision analysis, decision trees are used to represent decisions in decision making. The decision tree is one of the most important classification methods. It constructed by data provided, data values, and characters. The results of tree development affected by the number and type of attribute value significantly. The decision tree requires two types of data: training and testing data. Training data is the most significant part of the tree building data and procedures based on it. The more training data, the higher the decision accuracy. Testing data provides the level of accuracy and misclassification of decision trees. There are many decision tree algorithms. The algorithm used in this study is C4.5 and Random Forest.

C4.5 algorithm uses a learning mechanism. The choice of algorithm attributes based on assumptions that depend on the complexity of the decision tree and the amount of information. C4.5 extends the range of classifications to many attributes. This algorithm based on the entropy of data contained by nodal points generated from the decision tree. Entropy is representative of the degree of object disturbance in systematology. Small entropy induces small disorder [15].

Random forest is one of the popular tree-based algorithms that provide the best results for all types of problems. It is part of the ensemble learning method in which two or more weak algorithms are combined to form a robust modeling algorithm. The idea is to make too many trees in the forest. Based on the voting tree output will be done. Some functions generated by Random Forest are used by the "bagging" ensemble to overcome the problem of over-fitting when confronted with small data sets [14].

3. Result
One component that can take from the signal is its frequency spectrum by transforming a signal from the time zone to the frequency region so that a collection of information about the frequency contained in the signal will obtain, the transformation used is the Fourier transform. Getting the frequency spectrum from a seismogram signal can be done using Fourier transform, but directly calculating using FFT is inefficient because it is caused by computational time which will be very large. Therefore a method is used to obtain the signal frequency spectrum, which is the calculation of the estimated power spectrum, as shown in Figure 3. Spectral density has random properties which means it is not periodic, so the signal spectrum is based on Fourier analysis.
Data collection aims to make earthquake character references. It is done by taking the value of the feature extraction from each earthquake. Data sampling was carried out at three stations, namely Jurangmangu station, Bambangan station, and Gunung Cilik station. Sample data are grouped into one test data. There are three types of earthquake data tested, shallow volcanic earthquakes, gusts, and tremors. Samples of feature data obtained from the feature extraction are shown in Table 1.

| Data                  | Features  | RMS     | AVG     |
|-----------------------|-----------|---------|---------|
| Shallow Volcanic Earth| vb1jrmN.mat | 10.4253 | -9.162  |
|                       | vb1bbgN.mat | 7.6644  | -9.162  |
|                       | vb1clkN.mat | 6.5649  | -9.162  |
| Gust Earthquake       | hb1jrmN.mat | 13.9635 | -9.1626 |
|                       | hb1bbgN.mat | 14.9888 | -9.1626 |
|                       | hb1clkN.mat | 13.283  | -9.1626 |
| Tremor Earthquake     | lt1jrmN.mat | 6.5675  | -9.1629 |
|                       | lt1bbgN.mat | 7.0331  | -9.1629 |
|                       | lt1clkN.mat | 8.5671  | -9.1625 |

In the pattern recognition process, the extraction results of the \( \text{rms} \) and \( \text{avg} \) feature values used as material for the pattern approach according to the type of earthquake. As an illustration, shown in Figure 4, which shows the value approach process. The approach process is used to determine the outcome of signal processing. For example, for example some data obtained from the \( \text{rms} \) feature extraction are \( A = (10.4253, 7.5329, 3.658, 13.7372, 17.2639, 14.9888, 15.1844, 8.9468, 13.283) \). Then the data is sorted first so that searches are faster. Sort \( \text{rms} \) data on \( A \) starting from the smallest to
the largest value, then obtained \( B = (3.658, 7.5329, 8.9468, 10.4253, 13.283, 13.7372, 14.9888, 15.1844, 17.2639) \).

\[
\begin{array}{cccccc}
10.4253 & 3.658 & \rightarrow p & 3.658 & \rightarrow p & 3.658 \\
7.5329 & 7.5329 & \rightarrow p & 7.5329 & \rightarrow p & 7.5329 \\
3.658 & 8.9468 & \rightarrow d & 8.9468 & \rightarrow d & 8.9468 \\
13.7372 & 10.4253 & \rightarrow d & 10.4253 & \rightarrow d & 10.4253 \\
17.2639 & 13.283 & \rightarrow d & 13.283 & \rightarrow d & 13.283 \\
14.9888 & 13.7372 & \rightarrow d & 13.7372 & \rightarrow d & 13.7372 \\
15.1844 & 14.9888 & \rightarrow d & 14.9888 & \rightarrow d & 14.9888 \\
8.9468 & 15.1844 & \rightarrow q & 15.1844 & \rightarrow q & 15.1844 \\
13.283 & 17.2639 & \rightarrow q & 17.2639 & \rightarrow q & 17.2639 \\
\end{array}
\]

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

**Figure 4.** An illustration of the pattern approach according to the type of earthquake. (a) RMS values are sorted from smallest to largest. (b) Determine the initial position, final position, and middle value. (c) Find the middle value based on the new initial position. (d) Find the middle value based on the new end position.

First, take the initial position \( p = 1 \) and the final position \( q = N \), then find the middle data position with equation (4). If you want the value to match is \( 13.889 \), then \( dt = \frac{(1 + 9)}{2} = 5 \). The middle value obtained is the 5th data that is 13.283. Because the value is \( 13.889 > 13.283 \), then \( p = dt + 1 = 5 + 1 = 6 \).

And then, look for the middle value, so: \( dt = \frac{(6 + 9)}{2} = 7 \). The middle value obtained is the 7th data that is 14.9888. Because the value of 13.889 < 14.9888, then \( q = dt - 1 = 7 - 1 = 6 \). So the value of \( dt \) must be found again \( dt = \frac{(6 + 6)}{2} = 6 \). The final result obtained is 13,889 approaching the 6th data that is 13,737. Likewise, to find the avg value approach. Each earthquake has different rms and avg values. That is what will be used as a reference in the classification process. For example, the results of earthquake signal image processing get \( rms \) value = 10.4253 and \( avg \) = 9.1612, the two values obtained can be matched in the dataset.

The next classification method used to identify the type of earthquake is the C4.5 algorithm. The data used as input to this classification process are statistical features consisting of components: mean, standard deviation, correlation, variance, skewness, and kurtosis. The classification process uses the 10-fold cross-validation test mode. Five leaves build the classification system, and the tree size is nine, as shown in Figure 5. The classification results show an accuracy of 100% with the time taken to build model 0.04 seconds.
The following algorithm used for the classification process is Random Forest. Bagging is done with 100 iterations and a basic learning process. Trial mode is carried out with 10-fold cross-validation. The time needed to build the model is only 0.17 seconds. The accuracy of this algorithm is 100%. It shows that the algorithm used is excellent. Based on system testing, the results of identification of the three methods are shown in Table 2. It shows that the method used well with the accuracy of each method by 100%.

### Table 2. Identification Results

| Classifier          | Accuracy (%) |
|---------------------|--------------|
| Pattern Recognition System | 100          |
| Decision Tree Model | 100          |
| Random Forest       | 100          |

**4. Conclusion**

Based on the discussion, it can be concluded that signal processing using FFT can be applied to applications and helps in earthquake signal matching. The identification system can recognize signals that are approached using the Root Mean Square and Average Power Spectrum. Also, statistical features such as mean, standard deviation, correlation, variance, skewness, and kurtosis can be used for input data on the classification process. The classification process using the pattern recognition method, C4.5, and Random Forest shows the results of tests conducted with 100% accuracy of each method. It shows that the identification process works excellent.

**References**

[1] P V. dan M. B. G 2014 Badan Geologi [Online]. Available: http://www.vsi.esdm.go.id/index.php/gunungapi/data-dasar-gunungapi/529-gergaji-durian?start=2.

[2] Utari E L 2013 Pengenalan pola gelombang seismik dengan menggunakan wavelet pada aktivitas gunung merapi J Teknol Inf vol VIII no 23 pp 51–6.

[3] Ait Laasri E H, Akhouayri E S, Agliz D, Zonta D and Atmani A 2015 A fuzzy expert system for automatic seismic signal classification Expert Syst Appl vol 42 no 3 pp 1013–27.

[4] Romeo G 1994 Seismic signals detection and classification using artificial neural networks Annals of Geophysics vol 37 no 3.

[5] Kuyuk H S, Yildirim E, Dogan E and Horasan G 2011 An unsupervised learning algorithm:
Application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul Nat Hazards Earth Syst Sci vol 11 no 1 pp. 93–100.

[6] Esposito A M, D’Auria L, Giudicepietro F, Caputo T and Martini M 2013 Neural analysis of seismic data: Applications to the monitoring of Mt. Vesuvius Ann Geophys vol 56 no 4.

[7] Curilem M, Fernando H, Daniel B, Caesar S M, Gistavo F, Luis F, Carlos C, Gonzalo A, Max C , Salman K, Nestor B Y 2014 Pattern recognition applied to seismic signals of the Llaima volcano (Chile): An analysis of the events’ features J Volcanol Geotherm Res vol 282 pp. 134–47.

[8] Aykroyd R G and Hamed F M O 2014 Horizon detection in seismic data: An application of linked feature detection from multiple time series Adv Stat.

[9] Widagdo A, Candra A, Iswahyudi S and Abdullah C I 2013 Pengaruh struktur geologi gunung slamet muda dan tua terhadap pola sebaran panas bumi IRWNS 2013 vol 4 no 2, pp 204–7.

[10] Xiao Z and Menon C 2017 Counting grasping action using force myography: An exploratory study with healthy individuals JMIR Rehabil Assist Technol vol 4 p e5.

[11] Islam S, Islam S, Rahman S K and Ferdous N S 2014 Design of FIR filter using hamming window Int. J Emerg Res Manag Technol vol 3 no 2 pp 13–6.

[12] Yadav P, Sahu L D, Maravi and Rahi P K 2017 Design of low pass filter using Hanning and Hamming Window techniques Int J Res Appl Sci Eng Technol vol 5 no 6 pp 1983–87.

[13] Esmael B, Arnaout A, Fruhwirth R K and Thonhauser G 2013 A statistical feature-based approach for operations recognition in drilling time series Int J Comput Inf Syst Ind Manag Appl vol 5 pp 454–61.

[14] Hardani D N K, Wahyunggoro O, Nugroho H A and Faisal N 2014 Analysis of emotional condition based on electrocardiogram signals in 2014 International Conference on Electrical Engineering and Computer Science (ICEECS) pp 152–57.

[15] Smith M R, Wang X and Rangayyan R M 2009 Evaluation of the sensitivity of a medical data-mining application to the number of elements in small databases Biomed Signal Process. Control vol 4 no 3 pp 262–68.