Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Earthquake detection and location for Earthquake Early Warning Using Deep Learning

To cite this article: S. Anggraini *et al* 2021 *J. Phys.: Conf. Ser.* **1951** 012056

Earthquake detection and location for Earthquake Early Warning Using Deep Learning

S. Anggraini[1], S. K. Wijaya[1], and Daryono[3]
[1] Physics Department, University of Indonesia, Depok, West Java, Indonesia
[2] Center for Earthquake and Tsunami, Agency for Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia

E-mail : septa.anggraini@ui.ac.id

Abstract. Earthquake Early Warning System (EEWS) is a warning system that provides information about the estimated S wave arrival time, which can cause significant and destructive seismic energy using the information carried by the P wave. Technological advances in analyzing data supported by big data, the interconnection between networks, and high-performance computing systems in the era of the 4.0 industrial revolution have posed challenges to process and analyze earthquake early warning using modern seismological techniques. Early identification of earthquake events is the key to time efficiency to accelerate the dissemination of information. Here, we implement deep learning for early detection and classification of the earthquake P wave and noise signals using raw historical data from 3 component BMKG single station (2014 -2020) in the subduction zone of West Sumatra. The feature selection of the waveform is only selected for earthquakes distance in the cluster close to the station centroid. Statistically, the results of training and testing show good and convergent performance. This result is a preliminary study of deep learning, which is targeted at the classification of earthquakes p wave and noise signals and its association to estimate early earthquake location using 3 component record channels.

1. Introduction

Indonesia is located in an area with a high level of earthquake activity. This is due to the meeting point of three major tectonic plates in the world, the Indian Ocean - Australia in the south, the Pacific Ocean in the east, and Eurasia. The state of tectonic settings and high seismic activity causes Indonesia to become one of the world's most natural disaster-prone countries in the world.

As a result of the plate movement activity, Indonesia has 16 megathrust zone segmentations and 295 active earthquake sources of faults [1]. As a country with high seismic activity, an Earthquake Early Warning System is needed to provide earlier information before a strong earthquake hits an area.

Earthquake Early Warning System (EEWS) is a system for early detection of strong earthquakes based on the predicted arrival time of the S waves, which can cause significant shocks and even destructive seismic energy using the information carried by the P wave. From this EEWS sensor, the record will be sent to the EEWS Center, then the data processed automatically, and the results will be distributed to receivers.

By knowing the potential of destructive earthquakes early, the community as end-users (information recipients) can prepare themselves by making self-rescue efforts, including temporarily stopping vital objects to reduce the impact of a bigger disaster.

The Earthquake Early Warning System uses P wave information to estimate the strength of the destructive S wave. Seismic stations close to the source will record the release of earthquake energy in body waves or P waves and S waves, surface waves, or Love waves. The P wave is the first arrival phase wave that arrives at the seismic sensor for the first time. The arrival of the P wave contains comprehensive information about the fault size and source character. The P wave amplitude is usually smaller than the S wave and is not destructive, while the S wave is a destructive wave whose amplitude is generally more significant than the P wave. The P wave speed is faster than the S wave.

Early warning for earthquake disaster mitigation is currently possible with the development of high-tech and wide-frequency response (broadband) earthquake sensors with the support of dense earthquake observation networks in Indonesia, data acquisition systems, modern data processing computers, and high-capability servers satellite communication systems, and information dissemination system. Digital Video Broadcast (DVB) is also available, which operates 24 hours a day and seven days a week. Indonesian Tsunami Early Warning System (Ina-TEWS) has succeeded in providing information on earthquake parameters and tsunami warnings in less than 5 minutes. However, an earthquake early warning system does not yet exist and has not been developed optimally in Indonesia.

Accuracy is one of the most critical issues in earthquake early warning systems because false alerts can generate unnecessary public concern and cause significant economic losses. Noise signals usually influence earthquake wave readings. So this is a challenge to create an automatic system that is fast and precise in identifying earthquake events in real-time and reducing false alerts.

The solution that can answer this challenge is applying Seismilogy modern technology that allows real-time earthquake monitoring systems and provides a new paradigm in seismic disaster risk mitigation efforts. Currently, technological advances in analyzing data supported by big data, the interconnection between networks, and high-performance computing systems in the era of the 4.0 industrial revolution have begun to lead to a lot of research on earthquake early warning systems using machine learning and deep learning methods.

An extensive collection of methods to derive valuable knowledge from voluminous data is generated by machine learning. Machine learning algorithms, trained with sufficient information, can identify natural objects and make an expert-level decisions in various disciplines [2]. Experts take advantage of the latest developments in artificial intelligence, a highly versatile convolutionary neural network for the identification and location of earthquakes from a single waveform, and use this approach to research the seismicity caused [3].

This study implements a deep learning model for early detection and classification of the earthquake P wave and noise signals using raw historical data from 3 component BMKG single station (2014 -2020) using a deep neural network (DNN) subduction zone of West Sumatra. We used 5 s before and 5 s after P arrival time (containing events and windows free of events) dataset for preprocessing steps. Previous studies cast earthquake detection using a recent deep neural network in a large dataset [2, 3, 4] and applied it to induced earthquake [3]compared with this study with a relatively small dataset.

2. Related Literature and Methods

2.1. Earthquake Early Warning Systems

An earthquake early warning (EEW) system is designed to identify an earthquake, evaluate its parameters (hypocenter, magnitude, and time of origin), issue warnings to locations/areas where the appropriate steps must be taken before the arrival of disruptive seismic energy [5]. Earthquake early-warning systems use earthquake science and monitoring systems technologies to notify devices and people who are supposed to arrive at the affected locations when shaking waves produced by an earthquake. Seconds after the early warning information is obtained, it will allow individuals and structures to take steps to protect life and property from disruptive shaking.

The earthquake early warning system detects non-destructive P waves (Primary waves) rapidly before the destructive S waves (Secondary waves). The delay between the arrival of the P wave and the S wave controls the number of prior warnings that can be given. The farther the earthquake location from the epicenter, the farther the distance between the time intervals of the arrival of waves. This interval ranges between 60 and 90 seconds for earthquakes that are deep, far, and large.

Two Earthquake early-warning parameters are generally used in several countries that have implemented the Earthquake early warning system, the dominant period and the maximum shift amplitude, which is calculated a few seconds after the P wave arrives at the nearest station. Determination of the P wave's arrival time is crucial in determining the Earthquake early-warning parameters that will be used to predict earthquake parameters such as magnitude and hypocenter distance. A small error in determining the P wave's arrival time will significantly affect the accuracy of the magnitude and distance of the hypocenter obtained.

In the measurement of earthquake early warning parameters, the determination of the P wave's arrival time is crucial (the dominant period and the maximum displacement amplitude). The magnitude and the hypocenter are estimated using these parameters [6]. To calculate the P wave's arrival time, which has high precision and can be used to calculate EEW parameters, the Integ High Order Statistical technique is used. To obtain the displacement signal, the integration phase of the velocity broadband signal is applied.

There has lately been an interest in implementing deep learning and data mining algorithms for similarity-based event detection. A deep convolutional neural network (CNN) was trained in the Oklahoma area to detect and locate single-station waveform earthquakes simultaneously [3]. A deep learning Long Short-Term Memory (LSTM) network was equipped by [4] for sequence-to-label classification. To minimize false alarms triggered by impulsive local noise from natural earthquake early warning systems, [2] trained a generative adversarial network (GAN) to learn the characteristics of first-come earthquake P waves.

2.2. Signal-to-noise ratio (SNR)

Intense noise usually disturbs seismic waves that are recorded by near-surface sensors. Therefore, the seismic data collected is often of poor quality; this phenomenon can be explained as a low signal-to-noise ratio (SNR) [7]. The low seismic data SNR can decrease the rate of several subsequent seismological studies, such as inversion and imaging, for example. The reduction of unnecessary seismic noise is also of great significance.

It is possible to use Window D, part of the seismic record, for SNR:

$$D = \left[x_{i,j}\right]_{MxN} \left(0 < \mathrm{M} \le N_x, 0 < \mathrm{N} \le N_t\right) \quad (1)$$

The following assumption: waveform, amplitude, and seismic wavelet phase in window D hold constant in terms of distance “i” noise is randomly distributed "zero mean" along with the position of the survey line being independent (decorrelated) of the signal, such that the signal is independent (decorrelated)

$$x_{i,j} = s_j + n_{i,j} \quad (2)$$

$$\sum_{i=1}^{M} n_{i,j} = 0 \quad (3)$$

In general, these assumptions suggest a restriction to this approach, but they can be fulfilled if the local window is selected in the seismic section's stable signal area. Where $s_j$is amplitude of signal, and$n_{i,j}$ is amplitude of noise. So if the energy of the signal in a window is:

$$E_S = M \sum_{j=1}^{N} s_j^2 = \frac{1}{M} \sum_{j=1}^{N} \left(\sum_{i=1}^{M} x_{i,j} \right)^2 \quad (4)$$

It's possible to quantify the noise energy by

$$E_N = \sum_{j=1}^{N} \sum_{i=1}^{M} x_i^2 - E_S \quad (5)$$

Ultimately, the SNR decibel expression is calculated as:

$$SNR = \frac{E_S}{E_N} = 10 \log_{10} \left(\frac{\sum_{j=1}^{N} \left(\sum_{i=1}^{M} x_{i,j} \right)^2}{M \left(\sum_{j=1}^{N} \sum_{i=1}^{M} x_i^2 - E_S \right)} \right) \quad (6)$$

this SNR definition was introduced by [17].

2.3. Skewness

Skewness describes the normal distribution of the data. Normal distribution for skewness = 0 or data on the same distribution as mean. If the data distribution is larger with the sharpness average of the left-sloping curve, on the contrary, for smaller data, the sharpness of the curve is sloping to the right. Earthquake wave arrival time data are characterized by amplitude values that increase sharply or are more significant than the average. Therefore skewness can be used to identify the arrival of the P wave. The problem is the value.

The optimum curve is not at the wave's time of arrival. The time of arrival corresponds to the maximum value of the curve of skewness that occurs. The technique of differentiation skewness is designed to minimize the delay time.

2.4. Kurtosis

Kurtosis is a measure of the acuity of the A distribution. A Gaussian distribution has a kurtosis distribution of = 0. It is different from the earthquake signal, which has a non-Gaussian distribution that results in high kurtosis values. The normal distribution of earthquakes has a higher curvature than normal. This is used as a tool for identifying non-Gaussian signals. The number of data points in the selected time window (window shifting) depends on the data count (sampling rate).

2.5. Data

We used P waves waveforms and noise waveforms from the subduction zone of West Sumatra. The earthquake data collection consists of PDSI IA seismic network archives of broadband waveforms, operated by BMKG, Indonesia, from January 2014 to September 2020. We used the initial arrival times given by the BMKG Catalog, where appropriate. The recordings were continued three-component waveforms with a standard 20 sample per seconds (east-west, north-south, up-down).

Here, we focus on the megathrust zone in West Sumatra. We did sort seismic events catalogue based on p arrival time in PDSI station. In this region, the Meteorology, Climatology and Geophysics Agency (BMKG) catalogued 1661 seismic events (this catalogue is based on PDSI earthquake records) from January 2014 to September 2020 (Fig. 1). Seismic moment magnitudes range from M 1,6 to M 8,8. We used the continuous broadband records of PDSI stations. PDSI was active from 1 January 2007 to the present. Signals from PDSI stations are recorded at 20 Hz on three channels corresponding to the three spatial dimensions: BHZ (vertical), BHN (north-south), and BHE (west-east).

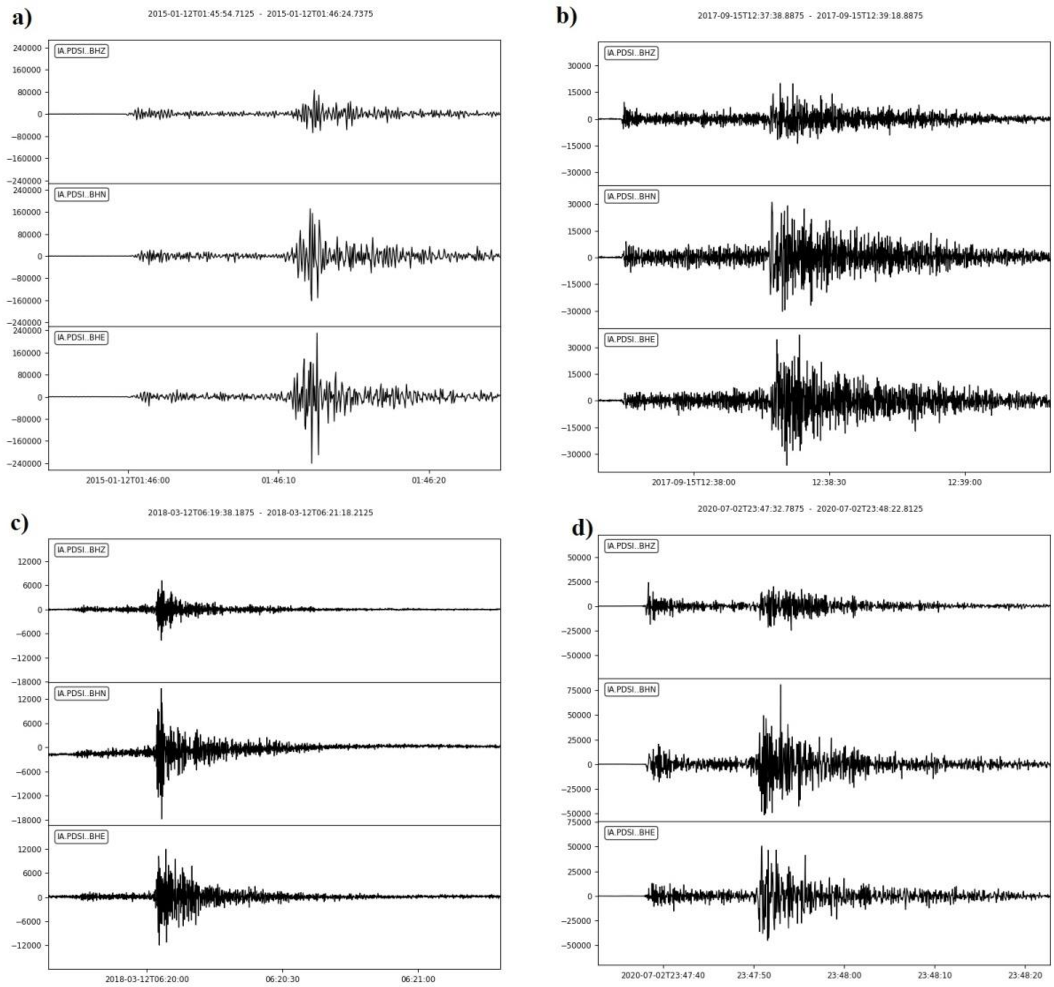

Figure 1. The four seismograms selected are labeled as earthquakes (components Z, N, E from top to bottom of each panel). The station code, waveform start time, origin time and earthquake magnitude (as indicated by the catalog provider) are reported as follows: (a) station code: IA.PDSI, event time: 2015-01-12T01:45:59.70, magnitude: 4; (b) station code: IA.PDSI, event time: 2017-09-15T12:37:43.90, magnitude: 7,7; (c) station code: IA.PDSI, event time: 2018-03-12T06:19:43.20, magnitude: 6,6; (d) station code: IA.PDSI, event time: 2020-07-02T23:47:37.80, magnitude: 7,3.

We perform basic preprocessing steps for both data sets before training our machine learning models, as follows: cut the waveforms 5 s before and 5 s after P arrival time(containing events and windows free of events). All the algorithm training and subsequent analysis are applied to three-component record Chanel, horizontal, and vertical broadband waveforms.

We used high-order statistical algorithms (i.e., kurtosis and skewness and SNR) to identify P-waves with differential skewness and kurtosis. The objective of this method is to obtain accurate calculation results that can be applied automatically. This careful and automatic calculation is indispensable in determining the dominant period and maximum displacement amplitude and the Earthquake Early Warning (EEW) information.

The first three basic features are the kurtosis value of three raw BHZ, BHE, BHN. The next three to nine features are the skewness and SNR values of three raw BHZ, BHE, BHN. The specific features (time series) are derived from ten seconds of P-wave (200 data points) of three raw EW, NS, and UD parts.

2.6. Generating location labels

Cluster analysis of K-means attempts to segment n entities into clusters in a multivariate dataset, where each individual in the dataset is entirely assigned to a particular cluster. K-means cluster analysis as a complex partition algorithm is an iterative operation. Next, the data is partitioned initially. Each group's means are determined, and then the data was partitioned again by assigning each data to the nearest mean location of the cluster[8, 9]. Cluster analysis is a multivariate technique that searches for data set trends by grouping observations data into clusters [10]. This method aims to determine the optimal grouping of related observations or items in each cluster (homogeneous). These clusters, though, are not the same as each other (heterogeneous). The level of data similarity determines the distance between the data. Small data distances suggest a high degree of data similarity, and, on the other hand, large data distances indicate low levels of data similarity.

We split the list of earthquakes in this area into nine regional clusters. We use the K-means algorithm for this, with the Euclidean interval between epicentres as the metric. The cluster centroids we obtain describe eight areas on the map (Fig. 2). The cluster whose centroid is the nearest (that is, each point is assigned to its Voronoi cell) is set to every point on the globe. We find that nine clusters cause the major earthquake sequences to be fairly partitioned. Therefore, our classification includes two codes or classes in the machine learning terminology: Class 0 corresponds to earthquake-free (seismic noise), and Class 1 corresponds to earthquakes occurring in the relevant geographical area.

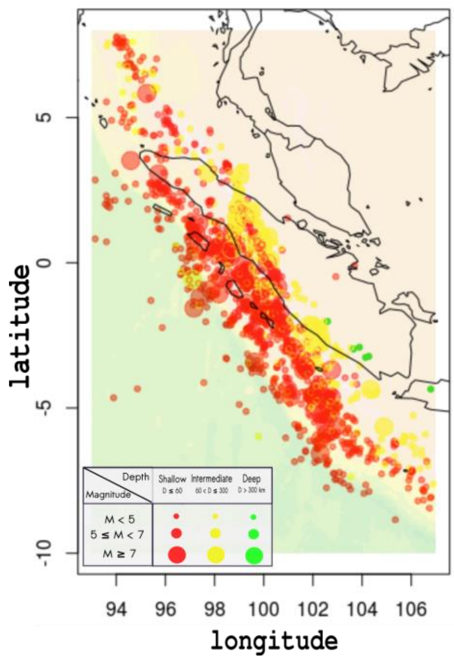

Figure 2. From 1 January 2014 to 30 September 2020, earthquakes and seismic stations in the Area of Interest (West Sumatra).

2.7. Methodology

Deep earthquake learning takes as input a three-channel window of waveform data and estimates a discrete probability in M groups or classes' machine learning terminology. The deep learning method architecture is composed, as seen in Figure 3, of an input, several hidden layers, a completely linked layer, and an output.

The training data set by optimization of the loss function specified as mean square error of output values were trained in this architecture. The self-adaptive architecture of the training and validation data sets was obtained through results.

Deep neural networks composed of dense layers are the most basic neural network architecture of deep learning (fully-connected layers). In this layer, all the inputs and outputs in each layer are connected to all the neurons. It should change the batch size and epochs according to computer memory and computational convergence, respectively. Here each epoch consists of 32 batch and 100 epochs. Commonly used activation functions in the literature are, for instance, sigmoid, hyperbolic tangent, Rectified Linear Units (ReLU) and its variants, see [11, 12, 13, 14]. Here, the training process used the sequential model with two fully-connected layers. There was 18 unit of dense layer, 0,4 probability rate of dropout Layer, Relu activation Function, next 2 unit Dense Layer and softmax for classification.

The benefit of having ReLU over other nonlinear activation functions is that, except for its negative values, it is not bound to output values, whereas other nonlinear activations are either saturated or bound to output values or both. The ReLU gradients intuitively take just two values, i.e., either zeros or ones. These ReLU features make the computations significantly faster than those using sigmoidal nonlinearities, and the mode's efficiency and performance are also improved. This study was programmed with TensorFlow CPU 2.2.0.

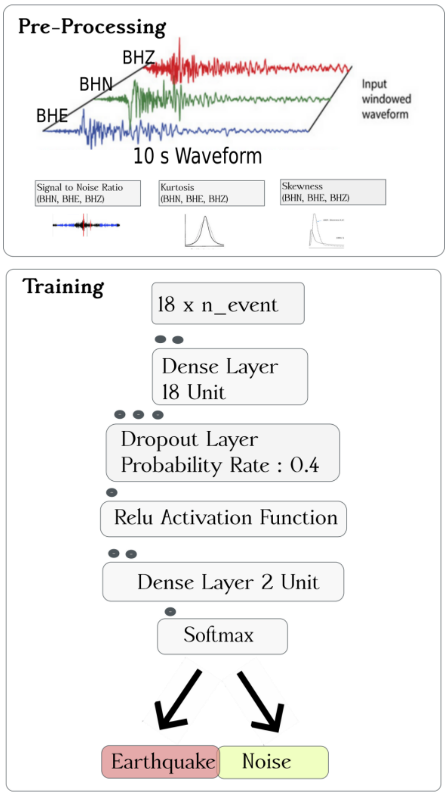

Figure 3. The deep neural network architecture corresponding to the proposed model.

Dropout is a process to prevent overfitting and also speed up the learning process. Dropout refers to removing neurons that are visible hidden or visible layers on the network. By eliminating a neuron, means removing it temporarily from the existing network. The neurons to be removed will be randomly selected. Each neuron will be assigned a probability value between 0 and 1.

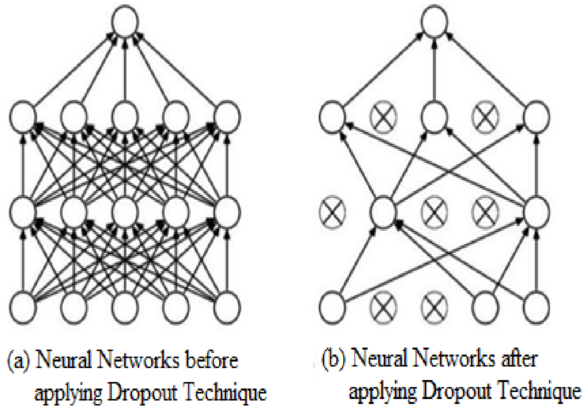

Figure 4. Dropout Technique

The picture above the neural networks (a) is an ordinary neural network with two hidden layers. Whereas in part (b), the neural network has applied dropout regularization technique where several activation neurons are no longer used. This technique is straightforward to implement in a deep learning model and will impact the model's performance in training and reduce overfitting.

For classification strategies with a vast number of classes, such as multinomial logistic regression, multiclass linear discrimination analysis, Naive Bayes Classifier, and Artificial Neural Network, the Softmax function is used. Softmax is a function that converts the K-dimensional vector X, which is an actual value, into a vector with the same shape but with values in the range 0-1, which is 1. The softmax function is used in layers in neural networks and is usually found in the last layer to get the output. Softmax neurons receive input and then do weighting and adding bias. But after that, the neurons in the softmax layer do not apply the activation function but instead use the softmax function. It can be concluded that the softmax layer determines the most significant probability for its class result.

A confusion matrix is used to calculate performance metrics (accuracy) which aims to measure the performance of the model that has been made. The confusion matrix provides information on comparing classification results carried out by the system (model) with the actual classification results. The confusion matrix is in the form of a matrix table that describes the performance of the classification model on a series of test data whose actual value is known. There are four terms as a representation of the result of the classification process on the confusion matrix. The four terms are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive (TP) is positive data that is predicted to be correct. True Negative (TN) Is negative data that is predicted to be correct. False Postive (FP) Is negative data but predicted as positive data. False Negative (FN) Is positive data but predicted as negative data.

3. Result and Discussion

Accuracy appears good result in a more large dataset with a balanced of both noises and events composition in the training dataset is suggested for the model. More investigation is necessary to reduce the false alarm. We were focused on clustering the Kmeans model. The lowest predictability of cluster is identified located beneath clusters of 3 and 5. From that analysis, we can boost the model with additional waveform signal from recognized cluster training or another way augmented dataset to enrich training dataset that indicates the lack that weighted by geographical exposure.

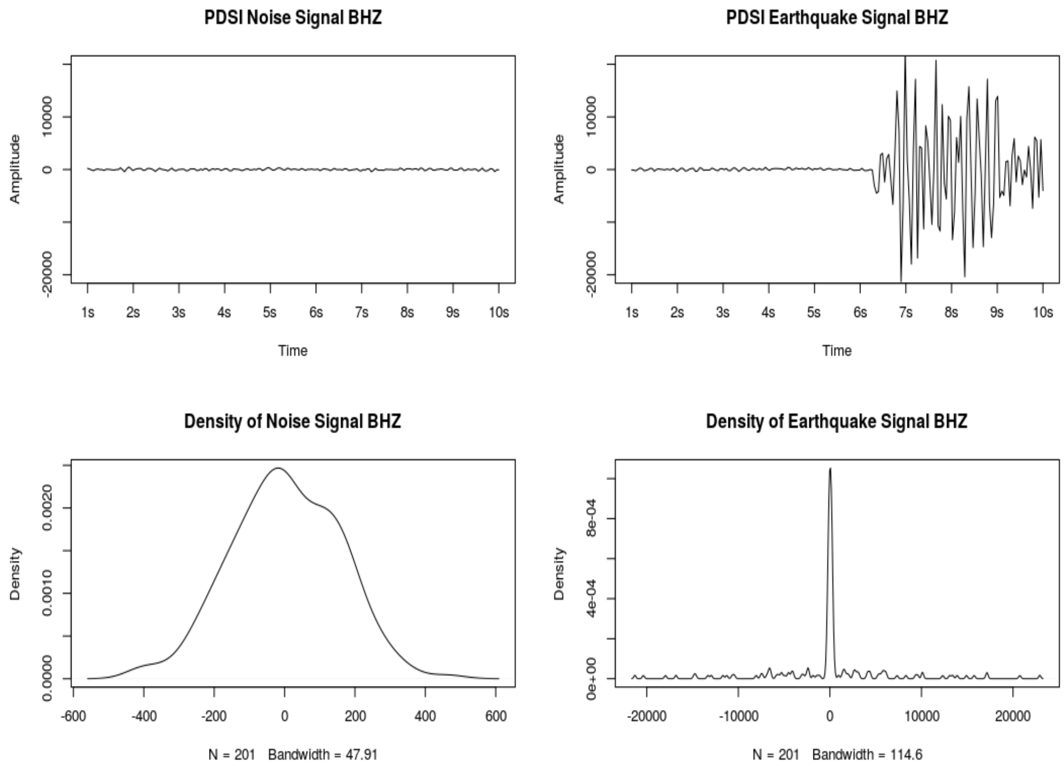

Figure 5. PDSI seismic waveform series

P waves and noises variety can be appropriately classified. As we can see in Figure 5, there is a difference in the density distribution pattern between earthquake and noise so that this data can be modeled for classification. Seismograms include noise and are often entirely covered by the intended signals [15]. Therefore, obtaining a better SNR by choosing a recording location with low ambient noise and data preparation is critical for applied seismology. The effectiveness of the above is primarily based on our comprehension of the disparity between seismic and noise signals.

Fourier analysis, the decomposition of the signal into a sine wave, is the basic mathematical method for this reason. Different mathematical formulas must be used, depending on the type of signal (transient or stationary, continuous or sampled). Generally, misclassified P waves have a relatively irregular waveform or an insignificant wave collection. The tested noise waveform often differs significantly in waveform characteristics, but most of them are appropriately categorized.

The dataset for training and validation was 1455 signals (events and noises) and 624 signals for testing. Our dataset was relatively small compared to the previous research [3]. Our model consisted of 18 units of dense layer, 0,4 probability rate of dropout Layer, Relu activation Function, next 2 unit Dense Layer and softmax for classification. We have trained our network with 100 epochs, with each epoch consisting of 15 iterations. The model was trained using 0,00001 learning rates.

The model shows slightly under-fitted with a loss value of 0.45 and validation loss of 0,38. Accuracy shows greater than 70% for the 30% split training dataset. Model accuracy increases 84% using (70% training : 30% testing). Training data has a different pattern that is identified by eyeball visualization. Accuracy appears good result in a bigger dataset with a balanced of both noises and events composition in the training dataset is suggested for the model. For testing results, the accuracy of the true negative was 93 %.

There is typically a trade-off between false positives (false alerts) and false negatives (missed alerts), as with any classification issue, which the trigger criteria can change. In addition to the proportional number of noise triggers, the discrimination threshold setting may be influenced by two other significant factors: the original step picker's robustness and the tolerance of false alarms by end-users. At the first stage, a robust phase picker will suppress many noise sources, thereby decreasing the discriminator's workload. The tolerance to false alarms varies between users. For instance, industry users with high false warning costs may be less tolerant of false alarms and prefer a high threshold for alerts. Personal users who receive no damage due to false alarms may be more forgiving and may use a comparatively low threshold to reduce the risk of missed alerts.

We then evaluate the accuracy based on the location, see fig. 6. The estimated class (1–9) is contrasted with the geographic label selected from the PDSI p arrival catalogue for each of the detected events. 98% location accuracy was obtained in cluster 9. More investigation is necessary to reduce the false alarm. We were focused on clustering the Kmeans model. The lowest accuracy testing based on clusters are identified located beneath cluster 5, then we find out more about this problem based on SNR value.

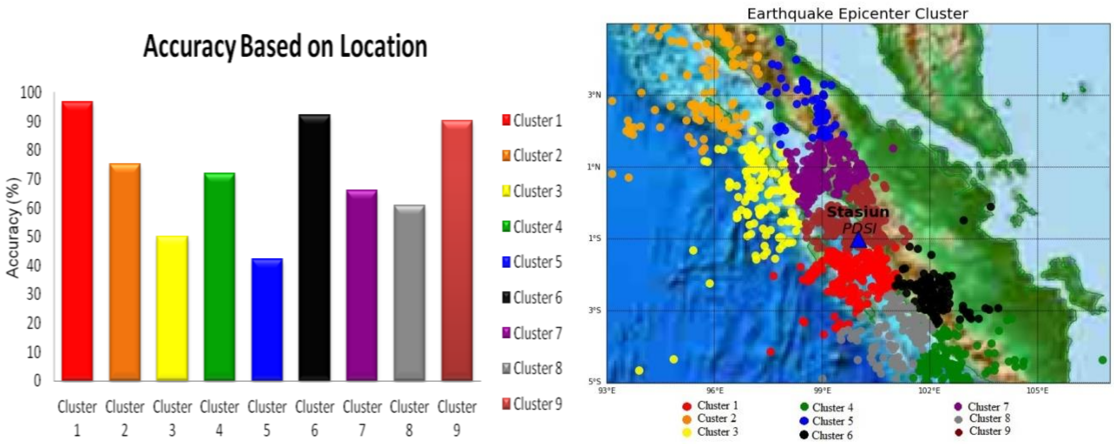

Figure 6. accuracy based on location (cluster)

From 624 tested signals, cluster 9 has the highest accuracy, see fig 6. From that point, and if we connect it to the location of the earthquake source, it turns out that the location of the earthquake source in cluster 9 is in an area close to the PDSI station. The far, the accuracy drops to 42 %. Cluster 3 and 5 have low average SNR value with the specification of SNR waveform see Table 1.

Table 1. SNR Value at low accuracy based on cluster

Culster	Average SNR BHZ	Average SNR BHN	Average SNR BHE
Cluster 3	3,53973221	3,437820694	3,507239733
Cluster 5	3,569937031	3,427483444	3,516632287

The low seismic data SNR can decrease the quality of several subsequent seismological studies, such as inversion and imaging, for example [3]. The reduction of unnecessary seismic noise is also of great significance. So for future work, low accuracy cluster needs specific processing techniques to improve the SNR value. One of the key challenges of applied seismology is maintaining high SNR or enhancing it by appropriate methods of data collection and analysis when conditions are poor. The effectiveness of SNR enhancement primarily depends on our experience of how seismic signals and noise varies. The performance of this method is better for near-source earthquake than the far source. These results are in accordance with previous studies using the different methods [4], but our accuracy was higher for the testing result.

Suppose the suggested technique is applied in EEWS to recognize earthquakes and noise with the near-source location stations detected in one second. In that case, an automated warning may occur well before the epicenter shakes strongly, so people who are close to the earthquake epicenter to be warned in the event of an earthquake [5].

4. Conclusion

This result is a preliminary study of deep learning to classify earthquakes p wave and noise signals and its association to estimate early earthquake location. Accuracy appears good result in a more extensive dataset with a balanced of both noises and events composition in the training dataset is suggested for the model. More investigation is necessary to reduce the false alarm. We are focused on clustering the Kmeans model. The lowest predictability of cluster is identified located beneath the cluster of 3 and 5. From that analysis, we can boost the model with additional waveform signal from identified cluster training or another way augmented dataset to enrich training dataset that indicates the lack that weighted by geographical exposure. For future work, we will add the dataset for training. Concerning increasing the amount of data, deep learning approaches are increasing [16].

References

[1] Pustlitbang PUPR, 2017 *Buku Peta Gempa 2017* .
[2] Li Z Meier M A Hauksson E Zhan Z and Andrews J, 2018 Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning *Geophys. Res. Lett.* **45**, 10 p. 4773–4779.
[3] Perol T Gharbi M and Denolle M A, 2017 Convolutional neural network for earthquake detection and location *arXiv* **2016**, November 2016 p. 2–10.
[4] Kuyuk H S and Susumu O, 2018 Real-time classification of earthquake using deep learning *Procedia Comput. Sci.* **140**, October 2018 p. 298–305.
[5] Tajima F and Hayashida T, 2018 Earthquake early warning: what does “seconds before a strong hit” mean? *Prog. Earth Planet. Sci.* **5**, 1.
[6] Gunawan H Puspito N T Ibrahim G and Harjadi P J P, 2011 Determination of earthquake early warning based on the initial phase of P-wave (case study of West Java) *IUGG Gen. Assem.* July.
[7] Chen Y Zhang M Bai M and Chen W, 2019 Improving the Signal-to-Noise Ratio of Seismological Datasets by Unsupervised Machine Learning *Seismol. Res. Lett.* **90**, 4 p. 1552–1564.
[8] Weatherill G and Burton P W, 2009 Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region *Geophys. J. Int.* **176**, 2 p. 565–588.
[9] Everitt B and Hothorn T, 2011 An Introduction to Applied Multivariate Analysis with R *Springer Sci. Media* p. 163–164.
[10] Novianti P Setyorini D and Rafflesia U, 2017 K-Means cluster analysis in earthquake epicenter clustering **3**, 2 p. 81–89.
[11] Maas A L Hannun A Y and Ng A Y, 2013 Rectifier nonlinearities improve neural network acoustic models *ICML Work. Deep Learn. Audio, Speech Lang. Process.* **28**.
[12] He K Zhang X Ren S and Sun J, 2015 Delving deep into rectifiers: Surpassing human-level performance on imagenet classification *Proc. IEEE Int. Conf. Comput. Vis.* **2015 Inter** p. 1026–1034.
[13] Xu B Wang N Chen T and Li M, 2015 Empirical Evaluation of Rectified Activations in Convolutional Network *arXiv*.
[14] Unterthiner T and Hochreiter S, 2016 Fast And Accurate Deep Network Learning *arXiv* p. 1–14.
[15] Bormann P and Wielandt E, 2013 Seismic Signals and Noise *New Man. Seismol. Obs. Pract. 2* June 2013 p. 1–41.
[16] Alom M Z *et al.*, 2019 A state-of-the-art survey on deep learning theory and architectures *Electron.* **8**, 3 p. 1–67.