International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 1
ISSN: 2088-5334

Tsunami Potential Prediction using Seismic Features and Artificial Neural Network for Tsunami Early Warning System

Astri Novianty [a,d,*], Carmadi Machbub [a,d], Sri Widiyantoro [b,d], Irwan Meilano [c,d], Daryono [e]

[a] *School of Electrical Engineering and Informatics, Bandung Institute of Technology, Jl Ganesha 10 Bandung, 40116, Indonesia*
[b] *Faculty of Mining and Petroleum Engineering, Bandung Institute of Technology, Jl Ganesha 10 Bandung, 40116, Indonesia*
[c] *Faculty of Earth Sciences and Technology, Bandung Institute of Technology, Jl Ganesha 10 Bandung, 40116, Indonesia*
[d] *Center for Earthquake Science and Technology, Research Center for Disaster Mitigation, Bandung Institute of Technology, Indonesia*
[e] *Indonesian Agency for Meteorology Climatology and Geophysics, Jl. Angkasa 1 No. 2 Jakarta, 10610, Indonesia*
Corresponding author: [*]astri.akademiks3@gmail.com

***Abstract*— Tsunamis are categorized as geophysical disasters because tectonic earthquakes triggered most of their occurrences. The high number of deaths due to tsunami catastrophe has made many countries develop a tsunami early warning system (TEWS), especially countries prone to tectonic earthquakes. One of the crucial subsystems in a TEWS is the tsunami potential prediction subsystem. To provide an early warning of tsunami, the prediction must be carried out based on early observation of the seismic event before the tsunami. In this short time of computation, the calculation of seismic parameters can only produce some roughly estimated features. Hence, a proper inference method that can decide accurate predictions upon the features is urgently needed for the TEWS. Some existing TEWSs are using rule-based inference to decide the prediction and often overestimate the prediction of tsunami potential. This study tries to develop a tsunami-potential prediction system using the machine learning approach as its inference method. Seismic features extracted from P-wave seismic signals are used as input data for learning and classification using a backpropagation artificial neural network (ANN). The accuracy result is then validated by K-fold cross-validation. Our simulation results show that the utilization of backpropagation ANN has given better accuracy in tsunami prediction compared to one of the existing TEWS that does not use machine learning for its prediction. At least for some seismic events that occurred during 2010-2017, the proposed system results in fewer overestimated predictions than the existing TEWS referred.**

***Keywords*— Tsunami; prediction; backpropagation ANN; seismic; early warning system.**

Manuscript received 14 Jan. 2021; revised 28 Apr. 2021; accepted 21 Jun. 2021. Date of publication 28 Feb. 2022.

I. INTRODUCTION

A tsunami is a huge sea wave that is stimulated by an impulsive disturbance on the seabed in a short time. There are mainly three geological phenomena that can trigger the disturbance and generate tsunami events, i.e., tectonic earthquakes, volcanic eruptions, and landslides on the seabed [1], [2]. Among those phenomena, earthquakes are the most common cause of tsunami. Not every tectonic earthquake generates a tsunami, but tectonic earthquakes cause most tsunamis around the world.

Sea-level rise observation is an often-used method to detect tsunami events. However, the observation's conclusion may be too late for a tsunami early warning. Alternatively, an earlier prediction for tsunami early warning can be provided by observing each earthquake event that occurs. When an earthquake occurs, its tsunami potential can be predicted based on its seismic characteristics. Some studies on the characteristics of earthquake events that generate tsunamis are available for tsunami prediction research [3].

Usually, a tsunami early warning system utilizes three basic seismic parameters to predict tsunami potential, whether a recently occurred earthquake will trigger a tsunami event or not [2], [4], [5]. They are hypocenter location, hypocenter depth, and magnitude. Some simple rules on those parameter values generate the prediction. The rule states that a tsunami potentially occurs if the following conditions are satisfied:

- The earthquake magnitude is 7.0 or higher.
- The hypocenter depth is less than 100 km.
- The hypocenter location is under the sea.

Tsunami-potential predictions issued by a tsunami early warning system often overestimate the real fact of tsunamis events. From disaster management's point of view, the overestimated predictions are better to ensure the safety and

alertness of all related communities, such as government agencies, in responding to the earthquake. However, in scientific and engineering views, overestimate predictions indicate that the system fails to detect the tsunami potential accurately and give an accurate warning for the public community. If this situation continues, it will have an impact on public trust in the early warning issued by the government at a certain point.

Two essential things influencing the accuracy of a prediction are relevant features and the inference method's ability to generalize the pattern based on the determined features. Therefore, one or both of them can cause frequent overestimated predictions.

Previous studies in Lomax and Michelini [3] proposed other features besides the three mentioned earlier to better measure tsunami potential prediction. It recommended two new parameters, i.e., rupture duration and P-wave dominant period. It is claimed that the parameters could represent the tsunami potential of an earthquake better than the previous parameters. However, the study still used a rule-based inference method to decide the prediction. It determined a specific threshold value for each parameter to decide the prediction.

The present study utilizes a machine learning approach to predict the tsunami potential of an earthquake event based on its seismic features. Machine learning has resolved various problems of pattern recognition with good performance [6]–[11]. Hence, it is expected to have more ability to generalize complex seismic patterns to result in more accurate predictions. This approach has been widely used in previous studies related to event seismic classification [12]–[18]. All of them use seismic features for the classifications.

Some previous studies also discuss the utilization of machine learning for tsunami identification purposes [19], [20]. They used basic seismic parameters for the classification. In recent years, machine learning methods are often used to solve many problems related to natural disaster management [21], [22].

The machine learning method used for the prediction in this study is artificial neural network (ANN). ANN has been widely used for many pattern recognition problems [23], including in the field of tsunami and earthquakes. In 2017, a simulation utilized ANN and SVM to identify tsunamigenic earthquakes based on some seismic parameters provided in the IRIS database [24]. Other studies focused on predicting the tsunami damage mapping and recognition using ANN [25], [26].

In this study, the data used for learning and classification consists of some seismic features extracted from seismic signals of tectonic earthquake events. The proposed system will classify the data into two classes, i.e., tsunami potential and no-tsunami potential. The performance of ANN will be validated using K-fold cross-validation.

II. Materials and Methods

The proposed system uses backpropagation ANN for predicting the tsunami potential. The system design can be seen in Fig. 1, which shows the block diagram of the system. The input data of the proposed system is seismic signals obtained from multi stations of seismometer networks. This study involved data from 61 earthquake events that had a magnitude size 7.0 or more and occurred from 1999 to 2017. The data consist of 20 seismic events that triggered tsunami events and 41 seismic events with no tsunami events triggered. All data are obtained from GEOFON [27] and IRIS web service [28].

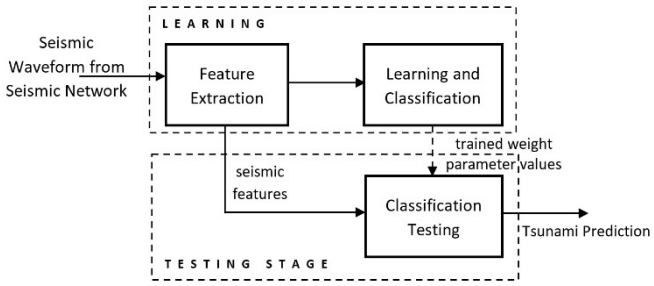

Fig. 1 The proposed system block diagram

The first process in the system is feature extraction to obtain some seismic features, as explained earlier in this study. The results of feature extraction are used as input for the learning process using backpropagation ANN. The learning process results in weight parameter values that can discriminant the seismic features data into two classes related to tsunami potential. These weight values then are utilized in the testing process using testing data.

A. Feature Extraction to Obtain Seismic Features

Seismic features used in this study are extracted from seismic signals obtained by seismometers. The seismic signal consists of three kinds of waves. The first is P-wave, which arrives first at the seismometer receiver when an earthquake occurs. The second part is the S-wave that arrives sometime after the P-wave, and the last is the surface-wave that lastly arrives and is the most destructive part [29].

Most of the seismic parameters reach their stable and final numerical value after the long-period wave of seismic signals come to the sensors [3]. For instance, most of the methods used to determine the earthquake magnitude are using the S-wave to get more accurate value [30]. Hence, the determination of some seismic parameters commonly must wait until the S-wave arrives [31], which is late for the early warning.

However, the determination of seismic parameters should be done earlier in order to provide an early warning of tsunami. Some previous studies have proposed alternative methods to calculate some specific parameters based on the P-wave as in [3], [30], so the seismic parameters can be obtained earlier compared to the calculation based on the S-wave.

There are some seismic features related to tsunami-potential prediction that can be calculated from the P-wave, i.e., rupture duration, P-wave dominant-period, and earthquake moment magnitude. The proposed system used these features as input data for learning and classification using the backpropagation ANN.

1) Rupture duration: Rupture duration is a significant parameter that can be used to predict the tsunami-potential of an earthquake event [3]. Rupture duration shows how long the area of fault needs time to rupture when an earthquake occurs. The procedure for calculating the rupture duration used in this study is the one developed by Lomax in [3]. To get the P-wave

part of the seismic signals, seismograms obtained from the seismometer need to be filtered by high-pass filter (HPF) because P-wave has higher frequency than other parts of the seismic signal. The result of the filter then is squared to get velocity-squared time series. To obtain the envelope function of the signal, the smoothing method should be done. In the procedure [3], triangle function is used to smooth the signal. After the envelope function is obtained, the delay between the first arrival of P-wave ($P0$) and the last point which the envelope function decreased below 90 (T_{90}), 80 (T_{80}), 50 (T_{50}), and 20 (T_{20}) percent of its peak value of amplitude, is calculated. Hence, the apparent rupture duration (T_0) is calculated using formula [3]:

$$T_0 = (1 - W) \cdot T_{90} + W \cdot T_{20} \quad (1)$$

with formula of the weight (W) is as follows:

$$W = \frac{\left[\frac{T_{80}+T_{50}}{2} - 20\right]}{40} \quad (2)$$

where T_n is delay duration between the first arrival of P-wave ($P0$) and the last point at which the envelope function decreased below $n\%$ of its amplitude peak value (in seconds).

2) P-wave dominant period: The P-wave dominant period reflects when the seismic event's slip is still growing or has not stopped yet [3]. This parameter is essential to determine the size of an earthquake. The method proposed by Wu and Kanamori in [30] for calculating the parameter is described as follows. A seismogram, which is the velocity signal, should be integrated to obtain the ground-displacement signal. P-wave dominant period (τ_c) is calculated by:

$$\tau_c = \frac{2\pi}{\sqrt{r}} \quad (3)$$

where r is the ratio between integrated velocity signal and integrated ground-displacement signal over some time period.

3) Period-Duration Discriminant: In addition to two seismic features described above for predicting tsunami potential, rupture-duration and P-wave dominant period, Lomax and Michellini [3] recommend a discriminant feature, i.e., duration-period discriminant ($T_d T_0$) which is product result of rupture duration and P-wave dominant period. Numerous expressions of combination between rupture duration (T_0) and P-wave dominant period (T_d) have been examined in [3] and it was found that $T_d T_0$ gives best agreement with rupture area parameter and tsunami importance of earthquake event that generates tsunami. This discriminant also identifies tsunami potential better than other expressions examined in [3].

4) Period-Duration Discriminant: The commonly used earthquake magnitude scale now is the seismic moment magnitude (M_W) which can represent the size of earthquake better than the previous scale, namely local magnitude (M_L) and surface-wave magnitude (M_S). Here, the moment magnitude scale is selected among others magnitude scale to be used as one of the seismic features, which can be calculated by [30]:

$$M_w = 4.525 \cdot log\, \tau_c + 5.036 \quad (4)$$

where is the average of the P-wave dominant period as described earlier.

B. ANN and its Performance Validation

ANN is known as a robust method for classifying or predicting some phenomena, which involves target function with real, discrete, or even vector values. It is also known as the most effective method for learning complex data from real-world sensor data [32], [33].

Backpropagation ANN is a neural network with multilayer architecture that involves multi perceptron in its learning process. The learning process conducted by the backpropagation ANN has supervised learning and consists of forward-pass and backward-pass.

The ANN architecture developed in this study only used one hidden layer. Hence, the synapsis weight between input-layer and hidden-layer, which connects input-neuron unit i and hidden-neuron unit j, is denoted as u_{ij}. While the synapsis weight between the hidden-layer and output-layer, which connects hidden-neuron unit j and output-neuron unit k, is denoted as w_{jk}. The number of input-neuron units is n_{in}, number of hidden-neuron units is n_{hidd}, and number of output-neuron units is n_{out}. The input value in each input neuron is denoted as x_i with $i = 0, .., n_{in} - 1$. The actual output of the network is y_k with $k = 0, .., n_{out} - 1$.

The complete function represented by a single hidden layer network for the forward pass in backpropagation ANN can be seen in the following formula.

$$y_k = \sigma\left(\sum_{j=0}^{n_{hidd}-1} w_{jk} \cdot \sigma\left(\sum_{i=0}^{n_{in}-1} u_{ij} \cdot x_i\right)\right) \quad (5)$$

The sigmoid function $\sigma(x)$ is the commonly used function for threshold function, which is stated as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (6)$$

Accuracy is the most widely used parameter to represent the performance of a machine learning method, including the ANN method. This study used the K-fold cross-validation to validate the accuracy of the proposed backpropagation ANN architecture.

In Wong and Yang [34] method, all data used in the research is partitioned into disjoint subsets, and the number of learning and testing conducted times, starting from $k = 1, 2, ...$ until $k = K$. In each time of learning and testing, one subset of data is set to be testing data, and the other $(k - 1)$ subsets of data are assigned as training data. For the first value of k, the first subset data is assigned as testing data, and the other $(k - 1)$ subsets are training data. For each of the next values of k, each of testing and training data is rotated to the next subset of data. Hence, at the end of the process, when $k = K$, all subsets' data will have been tested and the accuracy of the network can be calculated as follows:

$$acc_K = \frac{1}{K} \cdot \sum_{k=1}^{k=K} acc_k \quad (7)$$

where acc_K is accuracy testing for a certain value of K, where K is number of folds which is set in the experiment, and acc_k is accuracy testing of the k-th iteration of testing.

K-fold cross-validation can also be used to avoid data overfitting, which often happens in ANN. After validation is

done, the best ratio number between the number of testing data and the number of training data can be obtained to minimize the possibility of data overfitting [35].

C. Proposed Architecture of Backpropagation ANN

The architecture proposed here consists of five input neurons $(x_0, x_1, \dots, x_4)$, one hidden layer with a certain number of hidden neurons $(z_0, z_1, \dots, z_n)$, and two output-neurons (Y_0, Y_1). It uses five input neurons because four seismic features are used in the proposed system plus a biased input. The number of output neurons is determined based on the number of classes that should classify data input. For predicting tsunami, features representing the seismic events will be classified into two classes, i.e., earthquake events with tsunami potential and earthquake events with no-tsunami potential. Earthquake events with tsunami potential will lead to tsunami events, while the events with no-tsunami potential are on the contrary. So, the number of output neurons selected for the architecture is two output neurons.

Unlike input and output-neurons, the number of hidden-neurons cannot be determined by calculations. It can be determined by some experiments to obtain the optimal number which results in optimal accuracy. Therefore, one of experiment scenarios undertaken in this study is to set the number of hidden-neurons as a variable. It is assigned iteratively and for each value, the architecture is tested to measure its accuracy. The final architecture can be seen in Fig. 2.

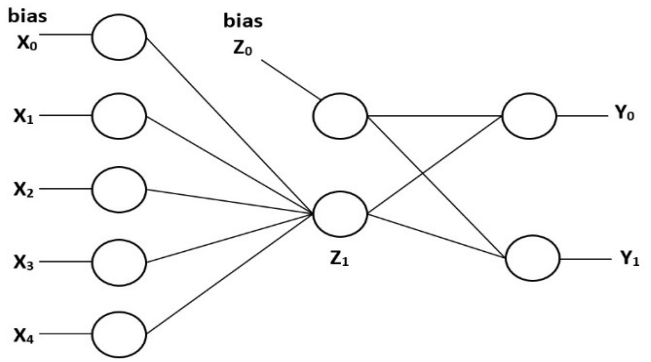

Fig. 2 The final proposed backpropagation ANN architecture with 2 hidden neurons

D. Learning and testing configuration

There are two objectives of learning and testing conducted in this study. The first one is to obtain the optimal backpropagation ANN architecture, which yields the best accuracy of the proposed system. The second is to measure the system performance in some accuracy terms with some validation using K-fold cross-validation.

For those purposes, learning and testing are done using K-fold cross-validation for a various number of K, starting from $K = 1$ to $K = 6$. In each value of K, a various number of hidden neurons were applied in learning and testing experiments to obtain the optimal ANN architecture for the proposed system. The learning rate (α) is set 0.1 for all scenarios of training. Some standard terms in machine learning are used for the accuracy analysis, i.e., true-positive (TP), true-negative (TN), false-positive (FP), false-negative (FN), precision, recall, and accuracy.

The formulas for calculating precision, recall, and accuracy are as follows.

$$Precision = \frac{num_{TP}}{num_{TP}+num_{FP}} \times 100\% \quad (8)$$

$$Recall = \frac{num_{TP}}{num_{TP}+num_{FN}} \times 100\% \quad (9)$$

$$Accuracy = \frac{num_{TP}+num_{TN}}{num_{TP}+num_{TN}+num_{FP}+num_{FN}} \times 100\% \quad (10)$$

where num_{TP} is the number of true positive predictions, num_{TN} is the number of true negative predictions, num_{FP} is the number of false-positive predictions, and num_{FN} is the number of false negative predictions.

III. Result and Discussions

In the first experiment using $K = 1$, the learning and testing process were conducted using 61 data events. Hence, there was no-unseen data involved in the testing. In other words, the system was tested with the same data as the system learned. This scenario of testing measures the ANN ability to memorize data that has been learned before.

The proposed architecture was tested using the variable number of hidden neurons, starting from 2 to 6 hidden-neurons. Architecture with only one hidden-neuron is excluded from the testing scenario because it only consists of one single-bias input in the hidden-layer. The result is shown in Table I.

TABLE I
First scenario testing result without unseen data

Number of Neuron Hidden	Mean Error (ME)	Accuracy (%)
2	0.020889328	88.52459016
3	0.024873797	88.52459016
4	0.012139936	91.80327869
5	3.67E-05	93.44262295
6	0.013155981	93.44262295

The result shows that the architecture with five hidden neurons results in the maximum accuracy (93,443%) and the minimum mean of error (0.0000367). However, we cannot recommend the optimal number of hidden neurons yet only from the result of the first testing scenario. The basic accuracy obtained from the first scenario only shows the ability of the proposed ANN architecture to memorize the already learned pattern of data.

The next experiment involved unseen data testing to measure the proposed ANN architecture's ability to recognize the data pattern that is different from the previous data pattern that the system has learned. In this scenario, the K value is set as a variable, starting from $K = 2$ to $K = 6$. Note that for all K, the total number of tested data is 61 data since the testing was carried out K times, and the testing data is shifted in each testing process. Table II and Fig. 3 show the result of this scenario.

TABLE II
RESULT OF THE TESTING SCENARIO USING UNSEEN DATA WITH K-FOLD CROSS VALIDATION

K	Number of Neuron Hidden	Testing Result						
		TP	TN	FP	FN	Precision (PPV) (%)	Recall (TPR) (%)	Accuracy (%)
2	2	13	37	4	7	76.471	65.00	81.967
	3	13	36	5	7	72.222	65.00	80.328
	4	13	33	8	7	61.905	65.00	75.410
	5	13	34	7	7	65.000	65.00	77.049
	6	13	33	8	7	61.905	65.00	75.410
3	2	13	37	4	7	76.471	65.00	81.967
	3	10	35	6	10	62.500	50.00	73.770
	4	10	34	7	10	58.824	50.00	72.131
	5	10	29	12	10	45.455	50.00	63.934
	6	5	34	7	15	41.667	25.00	63.934
4	2	16	37	4	4	80.000	80.00	86.885
	3	8	37	4	12	66.667	40.00	73.770
	4	14	35	6	6	70.000	70.00	80.328
	5	9	34	7	11	56.250	45.00	70.492
	6	7	33	8	13	46.667	35.00	65.574
5	2	15	37	4	5	78.947	75.00	85.246
	3	15	37	4	5	78.947	75.00	85.246
	4	12	36	5	8	70.588	60.00	78.689
	5	13	36	5	7	72.222	65.00	80.328
	6	9	32	9	11	50.000	45.00	67.213
6	2	14	37	4	6	77.778	70.00	83.607
	3	14	37	4	6	77.778	70.00	83.607
	4	14	37	4	6	77.778	70.00	83.607
	5	13	35	6	7	68.421	65.00	78.689
	6	12	36	5	8	70.588	60.00	78.689

Data in Table II shows that the optimal number of hidden neurons is two, resulting in the best accuracy value. Hence, the ANN architecture with two hidden neurons is selected to be the final proposed architecture, as shown in Fig.2.

Fig. 3 shows a comparison of accuracy for each K in the experiment using the final architecture (2 hidden neurons).

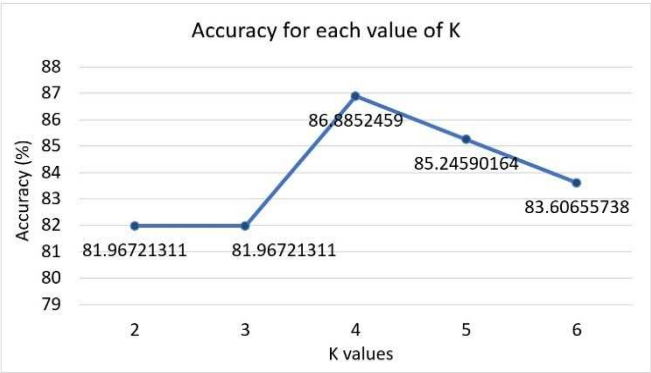

Fig. 3 Accuracy performance of final architecture using two hidden- neurons with some K values

It is worth noting that $K = 1$ was tested in the first testing scenario and it shows the network's ability in memorizing data. From Fig. 3, the final ANN architecture results in better accuracy in memorizing ($K = 1$) than that in generalizing the data pattern. However, for prediction and analysis purposes, the accuracy given from the scenario using unseen data should be considered, i.e., $K = 2, .,$ etc.

Therefore, the graphic data in Fig. 3 shows that the best prediction accuracy is achieved when $K = 4$. It means that the best ratio number between the number of testing data and training data is 1:3. Hence, the final testing of the proposed system will use this ratio number for the testing and training data.

In the final testing scenario, all seismic events data are sorted based on the time of occurrence. According to the ratio number recommended previously, the last quarter events of the 61 data are selected as testing data. Hence, the seismic events for the training data are the events that occurred in 1990-2009, while the testing data consists of seismic events that occurred in 2010-present. The result and its comparison can be seen in Table III.

The architecture network applied in the final testing is the final architecture with two hidden neurons. To present an objective analysis of the proposed system's final accuracy, the final testing result is compared to the tsunami prediction issued by one of the existing TEWSs, namely the Indonesia Tsunami Early Warning System (InaTEWS).

The table shows that the number of overestimating predictions issued by InaTEWS for tsunami warning during 2010-2017 is more than the proposed systems. InaTEWS issued overestimating predictions in seven seismic events. For the same events, the proposed system results in only two overestimating predictions.

The seismic events with overestimated predictions issued by InaTEWS are Aceh earthquake event (6th April 2010 and 9th May 2010), the 16th of June 2010 event and 29th September 2010 event in Irian Jaya region, the 10thJanuary 2012 event, the Molucca Sea earthquake event (15th November 2014), and the Southwest of Sumatra event (2nd March 2016). InaTEWS predictions for the events were false-positive (FP).

TABLE III
COMPARISON OF TSUNAMI PREDICTIONS FOR SEISMIC EVENTS IN INDONESIA REGION (2010-2017)

No.	Origin time of event[a]	Location of event[a]	Magnitude[a]	Fact	Ina TEWS		Proposed System	
					Prediction[s]	State	Prediction	State
1	06/04/2010 22:15	Northern Sumatra	7.8	No-Tsunami	Tsunami [16]	FP	Tsunami	FP
2	09/05/2010 05:59	Northern Sumatra (Aceh)	7.3	No-Tsunami	Tsunami [45]	FP	No-Tsunami	TN
3	16/06/2010 03:16	Irian Jaya Region	7.0	No-Tsunami	Tsunami [10]	FP	No-Tsunami	TN
4	29/09/2010 17:11	Irian Jaya Region	7.0	No-Tsunami	Tsunami [15]	FP	No-Tsunami	TN
5	25/10/2010 14:42	Southern Sumatra	7.8	Tsunami	Tsunami [10]	TP	Tsunami	TP
6	10/01/2012 18:37	Off W Coast of Northern	7.2	No-Tsunami	Tsunami [24]	FP	No-Tsunami	TN
7	11/04/2012 08:38	Off W Coast of Northern	8.6	Tsunami	Tsunami [24]	TP	Tsunami	TP
8	11/04/2012 10:43	Off W Coast of Northern	8.2	Tsunami	Tsunami [24]	TP	Tsunami	TP
9	10/12/2012 16:53	Banda Sea	7.1	No-Tsunami	No-Tsunami [24]	TN	Tsunami	FP
10	06/04/2013 04:42	Irian Jaya	7.0	No-Tsunami	No-Tsunami [7]	TN	No-Tsunami	TN
11	15/11/2014 02:31	Northern Molucca Sea	7.0	No-Tsunami	Tsunami [8]	FP	No-Tsunami	TN
12	27/02/2015 13:45	Flores Sea	7.0	No-Tsunami	No-Tsunami [24]	TN	No-Tsunami	TN
13	27/07/2015 21:41	Irian Jaya	7.0	No-Tsunami	No-Tsunami [6]	TN	No-Tsunami	TN
14	02/03/2016 12:49	Southwest Of Sumatra	7.8	No-Tsunami	Tsunami [46]	FP	No-Tsunami	TN
15	10/01/2017 06:13	Celebes Sea	7.3	No-Tsunami	No-Tsunami [5]	TN	No-Tsunami	TN

[a]Data at column 2, 3 and 4 were taken from IRIS database [28].
[b]Data at column 5 were taken from Global Historical Tsunami Database [36], except for no. 1 and 14 were taken from [37] and [38], respectively.
[c]Most of data at column 6 were taken from official report of Indian Ocean Tsunami Warning and Mitigation System (IOTWS) Unesco [39] and official webpage of Agency for Meteorology, Climatology and Geophysics (BMKG) Indonesia Tsunami Service Provider [40], except for no. 1, 2, 3, 4 and 14 were taken from [41][42][43][44], and [38], respectively.

According to the testing result in Table III, the accuracy performed by the proposed system is 86.67%, with four TP predictions, nine TN predictions, and two FP predictions. In other words, the proposed system results in better predictions for some seismic events that occurred during 2010-2017 compared to InaTEWS.

IV. CONCLUSION

The accuracy of tsunami prediction is a very crucial part of a tsunami early warning system. Either underestimating or overestimating prediction should be avoided to be issued as a tsunami warning. We have proposed a tsunami prediction system using a machine learning approach to provide a more reliable prediction than the comparable existing system (InaTEWS) using a rule-based inference method. From the testing conducted in this study, the accuracy of tsunami prediction issued by the proposed system is better than InaTEWS, at least for the data used in our analysis. However, as this study is still in its early stage, the results need to be improved further in future work.

ACKNOWLEDGMENT

We thank the Ministry of Research, Technology and Higher Education of the Republic of Indonesia, as the institution supports this study through the national strategic research grant. The same gratitude is also for BMKG – the Agency of Meteorology, Climatology, and Geophysics of Indonesia as the partner institutions of the research.

REFERENCES

[1] F. A. Ismail, A. Hakam, and T. Ophiyandri, "Earthquake safe houses training for tsunami preparedness in West Sumatra," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 1, pp. 318–324, 2020, doi: 10.18517/ijaseit.10.1.7850.

[2] S. Koshimura, L. Moya, E. Mas, and Y. Bai, "Tsunami damage detection with remote sensing: A review," *Geosci.*, vol. 10, no. 5, pp. 1–28, 2020, doi: 10.3390/geosciences10050177.

[3] A. Lomax and A. Michelini, "Mwpd: A duration-amplitude procedure for rapid determination of earthquake magnitude and tsunamigenic potential from P waveforms," *Geophys. J. Int.*, vol. 176, no. 1, pp. 200–214, 2009, doi: 10.1111/j.1365-246X.2008.03974. x.

[4] R. Atika, A. E. Raditya, R. N. Marjianto, and H. S. Pramono, "Automatic Tsunami Early Warning System Based on Open Data of Indonesia Agency for Meteorological, Climatological, and Geophysics," *J. Phys. Conf. Ser.*, vol. 1413, no. 1, 2019, doi: 10.1088/1742-6596/1413/1/012012.

[5] S. Harig *et al.*, "The Tsunami Scenario Database of the Indonesia Tsunami Early Warning System (InaTEWS): Evolution of the Coverage and the Involved Modeling Approaches," *Pure Appl. Geophys.*, vol. 177, no. 3, pp. 1379–1401, 2020, doi: 10.1007/s00024-019-02305-1.

[6] A. Fauzi and N. Mizutani, "Machine Learning Algorithms for Real-time Tsunami Inundation Forecasting: A Case Study in Nankai Region," *Pure Appl. Geophys.*, vol. 177, no. 3, pp. 1437–1450, 2020, doi: 10.1007/s00024-019-02364-4.

[7] A. Pratondo, C. K. Chui, and S. H. Ong, "Integrating machine learning with region-based active contour models in medical image segmentation," *J. Vis. Commun. Image Represent.*, vol. 43, no. 1, pp. 1–9, 2017, doi: 10.1016/j.jvcir.2016.11.019.

[8] M. A. Meier *et al.*, "Reliable Real-Time Seismic Signal/Noise Discrimination with Machine Learning," *J. Geophys. Res. Solid Earth*, vol. 124, no. 1, pp. 788–800, 2019, doi: 10.1029/2018JB016661.

[9] N. A. Hitam and A. R. Ismail, "Comparative performance of machine learning algorithms for cryptocurrency forecasting," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 3, pp. 1121–1128, 2018, doi: 10.11591/ijeecs.v11.i3.pp1121-1128.

[10] S. Naseer *et al.*, "Enhanced network anomaly detection based on deep neural networks," *IEEE Access*, vol. 6, pp. 48231–48246, 2018, doi: 10.1109/ACCESS.2018.2863036.

[11] W. Wang *et al.*, "A systematic review of machine learning models for predicting outcomes of stroke with structured data," *PLoS One*, vol. 15, no. 6, pp. 1–16, 2020, doi: 10.1371/journal.pone.0234722.

[12] L. Huang, J. Li, H. Hao, and X. Li, "Micro-seismic event detection and location in underground mines by using Convolutional Neural Networks (CNN) and deep learning," *Tunn. Undergr. Sp. Technol.*, vol. 81, no. June, pp. 265–276, 2018, doi: 10.1016/j.tust.2018.07.006.

[13] M. Malfante, M. D. Mura, J. Métaxian, and J. I. Mars, "Machine Learning for Volcano-Seismic Signals," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, 2018.

[14] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, "Machine learning in seismology: Turning data into

insights," *Seismol. Res. Lett.*, vol. 90, no. 1, pp. 3–14, 2019, doi: 10.1785/0220180259.
[15] W. Li, N. Narvekar, N. Nakshatra, N. Raut, B. Sirkeci, and J. Gao, "Seismic data classification using machine learning," *Proc. - IEEE 4th Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2018*, pp. 56–63, 2018, doi: 10.1109/BigDataService.2018.00017.
[16] Z. E. Ross, M. A. Meier, E. Hauksson, and T. H. Heaton, "Generalized seismic phase detection with deep learning," *Bull. Seismol. Soc. Am.*, vol. 108, no. 5, pp. 2894–2901, 2018, doi: 10.1785/0120180080.
[17] Z. E. Ross, Y. Yue, M. A. Meier, E. Hauksson, and T. H. Heaton, "PhaseLink: A Deep Learning Approach to Seismic Phase Association," *J. Geophys. Res. Solid Earth*, vol. 124, no. 1, pp. 856–869, 2019, doi: 10.1029/2018JB016674.
[18] T. Perol, M. Gharbi, and M. A. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, 2018, doi: 10.1126/sciadv.1700578.
[19] G. Pughazhendhi, A. Raja, P. Ramalingam, and D. K. Elumalai, "Earthosys—tsunami prediction and warning system using machine learning and IoT," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, 2019, vol. 28, pp. 103–113, doi: 10.1007/978-981-13-6459-4_12.
[20] A. Novianty, C. Machbub, S. Widiyantoro, I. Meilano, and H. Irawan, "Tsunami potential identification based on seismic features using KNN algorithm," *Proceeding - 2019 IEEE 7th Conf. Syst. Process Control. ICSPC 2019*, no. December, pp. 155–160, 2019, doi: 10.1109/ICSPC47137.2019.9068095.
[21] F. Martínez–Álvarez and A. Morales–Esteban, "Big data and natural disasters: New approaches for spatial and temporal massive data analysis," *Comput. Geosci.*, vol. 129, no. May, pp. 38–39, 2019, doi: 10.1016/j.cageo.2019.04.012.
[22] H. Jia, J. Lin, and J. Liu, "An earthquake fatalities assessment method based on feature importance with deep learning and random forest models," *Sustain.*, vol. 11, no. 10, 2019, doi: 10.3390/su11102727.
[23] O. I. Abiodun *et al.*, "Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.
[24] A. Kundu, Y. S. Bhadauria, S. Basu, and S. Mukhopadhyay, "Application of ANN and SVM for identification of tsunamigenic earthquakes from 3-component seismic data," *RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc.*, vol. 2018-Janua, no. 2, pp. 10–13, 2017, doi: 10.1109/RTEICT.2017.8256549.
[25] M. Syifa, P. R. Kadavi, and C. W. Lee, "An artificial intelligence application for post-earthquake damage mapping in Palu, central Sulawesi, Indonesia," *Sensors (Switzerland)*, vol. 19, no. 3, 2019, doi: 10.3390/s19030542.
[26] Y. Bai *et al.*, "A framework of rapid regional tsunami damage recognition from post-event terraSAR-X imagery using deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 43–47, 2018, doi: 10.1109/LGRS.2017.2772349.
[27] GFZ German Research Centre for Geosciences, "GEOFON and EIDA Data Archives," 2013. http://eida.gfz-potsdam.de/webdc3/ (accessed Jul. 10, 2018).
[28] Incorporated Research Institutions for Seismology (IRIS), "Wilber 3: Select Event," 2010. http://ds.iris.edu/wilber3/find_event (accessed Jul. 10, 2018).
[29] I. Dumke and C. Berndt, "Prediction of seismic p-wave velocity using machine learning," *Solid Earth*, vol. 10, no. 6, pp. 1989–2000, 2019, doi: 10.5194/se-10-1989-2019.
[30] Y. M. Wu and H. Kanamori, "Experiment on an onsite early warning method for the Taiwan early warning system," *Bull. Seismol. Soc. Am.*, vol. 95, no. 1, pp. 347–353, 2005, doi: 10.1785/0120040097.
[31] M. Picozzi, D. Bindi, P. Brondi, D. Di Giacomo, S. Parolai, and A. Zollo, "Rapid determination of P wave-based energy magnitude: Insights on source parameter scaling of the 2016 Central Italy earthquake sequence," *Geophys. Res. Lett.*, vol. 44, no. 9, pp. 4036–4045, 2017, doi: 10.1002/2017GL073228.
[32] M. S. Ghanim and K. Shaaban, "Estimating Turning Movements at Signalized," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1828–1836, 2019.
[33] N. Salleh, S. S. Yuhaniz, S. F. Sabri, and N. F. M. Azmi, "Enhancing prediction method of ionosphere for space weather monitoring using machine learning approaches: A review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 1, pp. 9–15, 2020, doi: 10.18517/ijaseit.10.1.10163.
[34] T. T. Wong and N. Y. Yang, "Dependency Analysis of Accuracy Estimates in k-Fold Cross Validation," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 11, pp. 2417–2427, 2017, doi: 10.1109/TKDE.2017.2740926.
[35] T. T. Wong and P. Y. Yeh, "Reliable Accuracy Estimates from k-Fold Cross Validation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, 2020, doi: 10.1109/TKDE.2019.2912815.
[36] National Geophysical Data Center / World Data Service (NGDC/WDS), "NGDC/WDS Global Historical Tsunami Database," 2010. https://www.ngdc.noaa.gov/hazard/tsu_db.shtml (accessed Sep. 01, 2018).
[37] Madlazim, "Assessment of tsunami generation potential through rapid analysis of seismic parameters: Case study: Comparison of the Sumatra Earthquakes of 6 April and 25 October 2010," *Sci. Tsunami Hazards*, vol. 32, no. 1, pp. 29–38, 2013.
[38] UNESCO Intergovernmental Oceanographic Comission, "2nd March 2016 Southwest of Sumatra Earthquake and Tsunami Event: Post-event Assesment of the Performance of the Indian Ocean Tsunami Warning and Mitigation System," Paris, 2017.
[39] UNESCO Indian Ocean Tsunami Warning and Mitigation System (IOTWS), "Tsunami Warning and Mitigation Systems to protect Coastal Communities," Paris, 2015.
[40] Agency for Meteorology Climatology and Geophysics Indonesia (BMKG), "Indonesia Tsunami Service Provider." http://rtsp.bmkg.go.id/ (accessed Feb. 07, 2018).
[41] D. Hartanto and T. Yatinamantoro, "Indonesia Tsunami Early Warning System (InaTEWS)." http://jexp.main.jp/h24soukai/Indonesia.pdf (accessed Feb. 07, 2018).
[42] "Magnitude 7.4 earthquake rattles western Indonesia, 9 May 2010," *The Independent News*, 2010. http://www.independent.co.uk/news/world/asia/magnitude-74-earthquake-rattles-western-indonesia-1969787.html (accessed Feb. 10, 2018).
[43] "Earthquakes kill three in Indonesia, 16 June 2010," *Brisbane Times*, 2010. https://www.brisbanetimes.com.au/world/earthquakes-kill-three-in-indonesia-20100616-yfqw.html (accessed Feb. 12, 2018).
[44] Community Tsunami Early-Warning Center, "September 29th 2010." http://www.communitytsunamiwarning.com/page20.htm (accessed Feb. 12, 2018).