End-to-End LSTM-Based Earthquake Magnitude Estimation With a Single Station

Aarón Cofré, Marcelo Marín, Oscar Vásquez Pino, Nicolás Galleguillos, Sebastián Riquelme, Sergio Barrientos, and Néstor Becerra Yoma, *Senior Member, IEEE*

***Abstract*— In this letter, a method based on long short-term memory (LSTM) is presented to address the problem of earthquake magnitude estimation for earthquake early warning (EEW) and tsunami early warning (TW) purposes using a seismic station. An end-to-end-based scheme is adopted, and particular attention is paid to compute the magnitude of seismic events larger than M6 and reduce the effective time for TW. More so, these earthquakes are the ones that cause more fear or uncertainty in the population with the provision of the most significant destructive potential. However, the occurrence of large earthquakes is low, but to counteract the drawback of limited training data, engineered features were also proposed. The earthquake magnitude relative error estimation reported here in experiments with Chilean seismic data was 4.01% and 8.04% with earthquakes M4.0 or larger (up to M8.1) and M4.0 or smaller, respectively, by employing seismic traces in the nearest station to the corresponding seismic event. The average earthquake-nearest station distance was 196 km, and in 26% of the data, this distance was greater than 200 km. These results are competitive with those published elsewhere and suggest the possibility to reduce the time required for EEW and especially TW.**

***Index Terms*— Deep learning, earthquake magnitude estimation, engineered features, limited training data.**

I. Introduction

EARTHQUAKE early warning (EEW) and tsunami warning (TW) are relatively new strategies to reduce death tolls and increase the resilience to seismic hazards in urban environments [1], [2]. The purpose of EEW is to provide real-time information about ongoing earthquakes, allowing communities, governments, businesses, and critical infrastructure to take timely action to reduce damage or losses before the arrival of secondary and surface waves that are generally the most destructive ones [1], [3]. EEW aims at characterizing earthquakes (i.e., their location, depth, and magnitude) in the shortest possible time [4]. Of these parameters, the estimation of the magnitude is critical since it is an indication of the energy released and the force with which it will strike locations of interest [5]. In addition, the earthquake magnitude is one of the criteria employed to determine if a potential tsunami is likely to be generated. One strategy to reduce the response time of these systems is to automatize the earthquake characterization, particularly the magnitude computation. Usually, this information is generated after the event has been observed in a minimum number of seismic stations, which in turn requires a few minutes from the earthquake occurrence. However, from the EEW perspective, this procedure should take place as soon as the earthquake is observed for the first time at the nearest station [6]. Although this problem has already been addressed [7], current methods provide estimation errors or restrictions that may limit their applicability.

In the literature on earthquake magnitude estimation, the most frequent use of a deep neural network (DNN) has been utilized to extract features and decrease the dimensionality of seismic signals from station networks with convolutional neural networks (CNNs) [8]–[10] to automate the analysis of seismic traces. This kind of strategy needs large databases, which in turn should also be representative regarding magnitude and location. The first CNN-based method, ConvNetQuake, was presented in [10]. It extracts the signal characteristics without any preprocessing and classifies ranges of distances and magnitudes. In [9], a CNN was employed to extract the characteristics of seismic traces from the time-frequency representation of signals resulting in complex or real only features. The use of data from a single station has been explored for earthquake characterization in [11], [12], and [13]. The databases obtained in this condition are usually smaller and provide a narrower range of magnitudes. In [12], CNN is employed on an end-to-end basis to detect events and classify their magnitude with a single station that was located on a fault. Most of the events were M4 or smaller and the data (more than 65 000 events) was recorded in 29 years, which in turn results in traces obtained with different instruments.

In contrast to CNN, long short-term memory (LSTM) has found a niche in predicting seismic events [14], [15]. LSTM has also been combined with CNNs, which in turn extracts features and decreases the dimensionality of the raw trace before being processed by the recurrent network [8]. Nonetheless, the method requires a large amount of data that should also present a high SNR. Surprisingly, LSTM has not been explored exhaustively to characterize earthquakes even though this neural network is recurrent and was conceived to process the time sequence of data such as seismic traces. One of the

Manuscript received January 20, 2022; revised March 14, 2022 and April 8, 2022; accepted May 10, 2022. Date of publication May 13, 2022; date of current version May 26, 2022. This work was supported by the Agencia Nacional de Investigacion y Desarrollo (ANID)-Fondef under Grant ID20I10212. *(Corresponding author: Néstor Becerra Yoma.)*

Aarón Cofré, Marcelo Marín, Oscar Vásquez Pino, Nicolás Galleguillos, and Néstor Becerra Yoma are with the Speech and Processing Transmission Laboratory, Department of Electrical Engineering, University of Chile, Santiago 8370451, Chile (e-mail: nbecerra@ing.uchile.cl).

Sebastián Riquelme and Sergio Barrientos are with the Centro Sismológico Nacional (CSN), University of Chile, Santiago 8370451, Chile.

Digital Object Identifier 10.1109/LGRS.2022.3175108

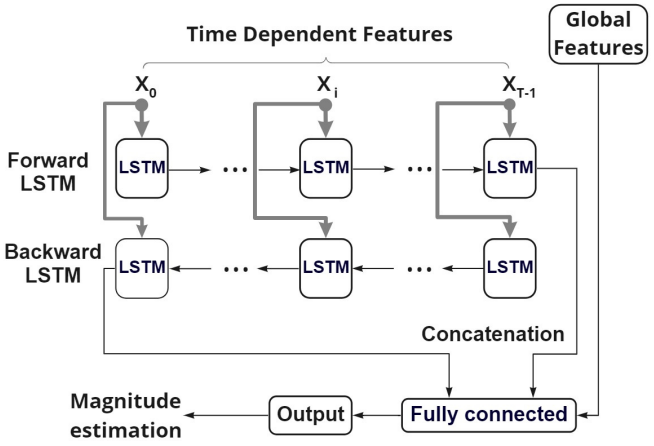

Fig. 1. Block diagram of the proposed LSTM-based architecture.

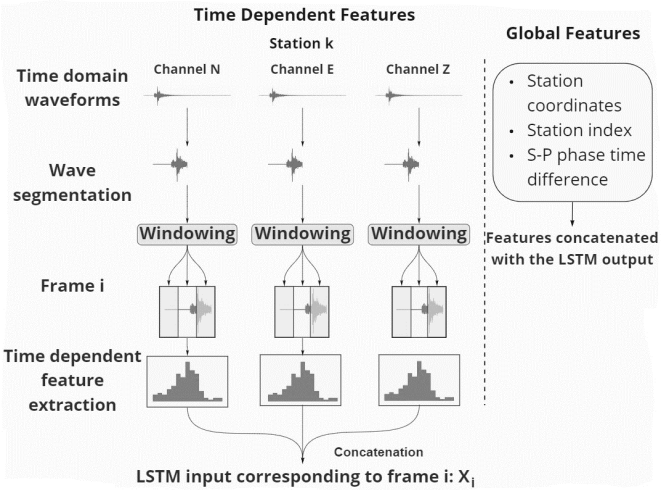

Fig. 2. Feature extraction.

motivations to employ deep learning is the possibility, in principle, to use raw signals. This assumes that the first layers of DNN architectures can generate features automatically out of the input signals [16], [10]. Nevertheless, as aforementioned, this requires a large amount of data. In contrast, engineering the input features allows for improving the DNN performance with a limited amount of training data because its task becomes less difficult. This strategy is especially interesting for M6 or larger earthquakes whose frequency of occurrence can be much lower than smaller seismic events according to the Gutenberg–Richter law. Moreover, the problem of limited training data becomes more acute because a machine learning engine trained with seismic information from a given location decreases its performance when exported to other regions [17].

The contribution of this letter corresponds to an LSTM-based scheme for EEW and TW magnitude estimation that needs only the seismic trace in the first station where the earthquake was observed that is usually the nearest one. One of the motivations is to reduce the effective time for TW, which in turn requires estimating the magnitude of seismic events greater than M6. But the number of this kind of earthquake that is registered is low. To counteract the limitation of limited training data and avoid training the proposed system with raw data, engineered features were also proposed.

II. LSTM-Based Proposed Architecture and Engineered Feature Extraction

A block diagram of the proposed LSTM-based scheme is shown in Fig. 1. The information feeding the deep learning architecture is composed of two inputs: first, the time-dependent features that feed the bidirectional LSTM; and second, the global features that are concatenated and combined with the LSTM output are input to a multilayer perceptron (MLP) or fully connected neural network to deliver earthquake magnitude estimation. LSTM is a recurrent neural network that can model long-term dependencies [18]. In contrast to other neural network architectures, LSTM does not need the sequence of observation vectors to have the same temporal duration, which in turn is very convenient to capture the dynamics of seismic events.

A. Preprocessing

These raw data corresponded to time-domain seismic signals that were downloaded in counts were transformed to velocity and equally had the instrument response removed. Subsequently, the amplitude was multiplied by a constant (i.e., 10^{+10}) to avoid negative numbers after applying log to the FFT absolute value. The sampling rate was also adjusted to 40 Hz. Finally, the signals were processed by applying a high-pass filter with a cut-off frequency of 1 Hz.

B. Engineered Features Extraction

Engineered features extraction aims to obtain a representation of the signal that highlights or preserves the information that is relevant to the final objective, that is, the earthquake magnitude estimation in this article. Engineering features usually require both prior knowledge of the target process and parameter tuning. Notably, feature extraction leads to a reduction in the dimensionality of the input data, which in turn results in fewer ML parameters that need to be trained. Fig. 2 shows the block diagram corresponding to the feature extraction process. Each event signal was segmented as follows: the beginning corresponded to around 20 s before the P wave arrival, and the end of the earthquake was determined within the coda stage when the frame energy was 3% of the maximum frame energy of the seismic event. Then, the segmented signal was divided into overlapped Hamming windows. Time-dependent features correspond to those estimated in each analysis window and result from the short-term FFT power spectrum plus the frame energy (see Fig. 2). The logarithm of the FFT power spectrum can be computed. In contrast, global parameters denote those that are allocated to the whole event and are composed of the station coordinates, station index, and S-P phase time difference. Station index corresponds to an integer assigned to each station aleatorily. In principle, the time-dependent features extracted from the north–south, west–east, and vertical components can be concatenated to feed the LSTM layer. The global characteristics attempt to individualize the seismic stations and preserve the event-seismic station distance by

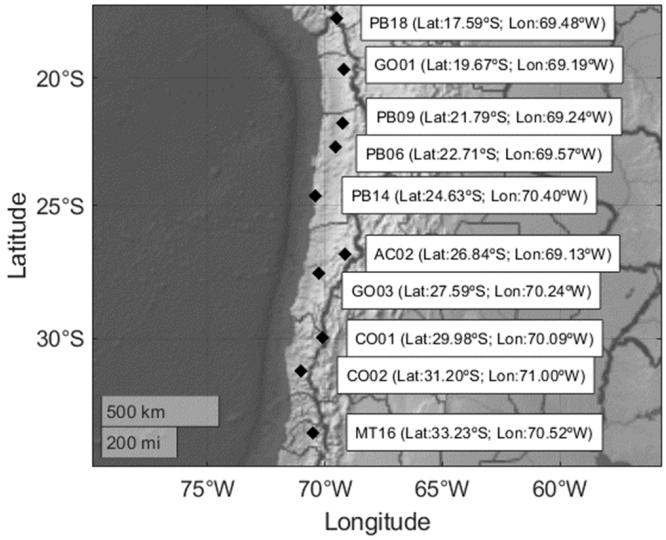

Fig. 3. Reference stations.

means of the S-P phase time difference, which in turn can be estimated with the STA/LTA method [19]. Finally, both sets of features need to be normalized. Both the mean-and-variance (MVN) and min-max normalization procedures were tested. The corresponding means and variances and min-max values were determined with the training data. In the case of the time-dependent parameters, the normalization takes place in each time trajectory.

III. Dataset

The data was provided by the National Seismological Center (CSN, "Centro Sismológico Nacional") in Chile. It was downloaded with the Incorporated Research Institute for Seismology (IRIS) platform. The seismic traces provide the date, time, magnitude, and location of the seismic events. The catalog contains 7580 earthquakes that occurred in the north-central region of Chile between 2014 and 2021, between latitudes 18.23°S and 33.43°S, and longitudes 69.16°W and 74.79°W, that were observed in ten reference stations (see Fig. 3). The stations cover a distance equal to 2.000 km along the country. The reference stations were chosen among those with the highest SNR along the target Chilean region. The maximum hypocenter depth is 70 km. The minimum and maximum earthquake magnitudes are M1.8 and M8.3, respectively. The distance from the hypocenter to the nearest station is important for an early warning system. In the data employed here, the nearest station was on average 196 km from the hypocenter. From the catalog, 500 events were chosen with an SNR equal to or higher than 10 dB, allowing a clear contrast with the background signal.

IV. Results and Discussions

Three error metrics were used to evaluate the performance of the proposed scheme for early earthquake magnitude estimation. First, the mean absolute percentage error (MAPE) which is defined as MAPE $= 1/N \sum abs((M_{\text{Ref}} - M_{\text{Est}})/M_{\text{Ref}})100$, where N is the number of tests performed, M_{Ref} and M_{Est} are the reference and estimated earthquake magnitudes, respectively. The second metric corresponds to the mean absolute error (MAE) described as MAE $= 1/N \sum abs(M_{\text{Ref}} - M_{\text{Est}})$, and finally, the third metric is the mean squared error (MSE) that is calculated as MSE $= 1/N \sum (M_{\text{Ref}} - M_{\text{Est}})^2$.

A. Feature and Hyperparameter Optimization

The following procedure was adopted to optimize the features and hyperparameters. First, the proposed neural network hyperparameters were kept constant and the time-dependent features were optimized. Once these features had been optimized, the neural network hyperparameters were tuned. Finally, the global features were evaluated. The proposed LSTM-MLP network was trained 30 times for each setting configuration. The whole database was divided into training (60%), validation (20%), and testing (20%) making sure that the magnitude distribution was similar in all the three subsets. MAPE was employed as an objective metric and was computed on the validation subset for tuning purposes and not on the testing one. In this optimization procedure, the training and validation subsets were composed of the signals observed in the nearest stations to the seismic events. As a result, the optimal feature setting corresponded to: FFT size equal 256 samples; window width equal to 4 s with a 2 s overlap; and frame log energy. Surprisingly, the highest accuracy was achieved with the Z component only and the north–south and east–west signals were discarded. Consequently, the time-dependent features were composed of 129 FFT log spectrum bins + frame log-energy = 130 parameters. The optimal threshold to determine the end of seismic events was equal to 3% of the maximum frame log-energy in the event. Regarding the neural network configuration and hyperparameter tuning, the forward/backward LSTM with output dimensionality equal to 10 and an MLP with a single 30-node hidden layer were the result of the optimization. Regularization had a positive effect when applied to the LSTM only. The optimal regularization corresponded to L2 (10^{-2}).

After optimizing the time-dependent features and the neural network configuration/hyperparameters, the global features were evaluated. The time-dependent features provided a MAPE equal to 7.44%. Global features were tested individually or in combination with other global features. Interestingly, the incorporation of global features always gave improvements in the error metrics. In these conditions, the lowest MAPE was achieved with the station index and was equal to 6.77%, which is 10% lower than the one with temporal features only. This result suggests that the identification of the station from where the seismic event is observed is important information for magnitude estimation. When the engineered features were replaced with raw data, that is, the time waveform or its FFT, MAPE increased dramatically to around 30%, which corresponds to an increase in MAPE greater than 300%.

B. On Discriminating Event–Station Distance and Event Magnitude

It is common sense in pattern recognition that the higher the SNR is, the higher the recognition accuracy. In addition,

TABLE I
RESULTS OBTAINED WITH THE TESTING SUBSETS

Database	Nearest station			Distant station		
	MAE	MSE	MAPE	MAE	MSE	MAPE
All magnitudes	0.23	0.12	6.77	0.47	0.35	14.34
Magnitude > 4	0.21	0.09	4.01	0.26	0.11	4.92
Magnitude < 4	0.23	0.10	8.04	0.47	0.30	17.03

TABLE II
STATISTICAL SIGNIFICANCE ANALYSIS (DEPENDENT t-TEST)

Comparison with respect to MAPE	Statistical significance level (p)
With (6,77%) v. Without (7,44%) Global Features	$p < 0.05$
Engineered Features (6,77%) v. Raw data (31,65%)	$p < 10^{-6}$
Near (6,77%) v. Distant Stations (14,34%)	$p < 10^{-6}$
Magnitude > M4 (4,01%) v. Magnitude < M4 (8,04%)	$p < 10^{-6}$

it is reasonable to consider that, on average, the greater the event–station distance is, the lower the SNR. As a matter of fact, low-magnitude earthquakes may not be observed at distant stations. Consequently, the proposed earthquake magnitude estimation method was evaluated in two conditions: first, with the signal observed in the station that was the closest one to the seismic event; and second, with the signal observed in a station considered distant to the seismic event. The nearest station to a given seismic event (hypocenter) is at an average distance equal to 196 km. On the other hand, what are considered here as distant stations are those that are at distances greater than 500 km (i.e., an average distance equal to 775 km) from the hypocenter. Identifying and employing the first station where an earthquake is observed is very important: the SNR should be the highest possible one on average and the early earthquake or tsunami warning could be given in a shorter time. Additionally, the earthquakes were grouped in two sets according to their magnitude: those that were greater or equal to and those that were smaller than M4. Each set contained 250 events. Conclusively, six sets of data were generated to train and test six systems (Table I): earthquakes of any magnitude and the nearest station; earthquakes of any magnitude and a distant station to each event; earthquakes greater than M4 and the nearest station; earthquakes greater than M4 and a distant station to each event; earthquakes smaller than M4 and the nearest stations; and earthquakes smaller than M4 and a distant station to each event.

According to Table I, when the proposed system was evaluated with the whole database, that is, employing all the available magnitudes, MAPE was equal to 6.77% and 14.34% with the closest station and a distant one, respectively. This result corroborates the convenience of processing the seismic signal observed at the stations that are the nearest ones to the earthquake to achieve more accurate results. Similar behavior was observed with MAE and MSE: they are at least twice as higher as the signals observed at a distant station when compared to the closest one to the earthquake. It was observed that when the system was evaluated with magnitudes greater than M4, MAPE was 50% of the one achieved with the system trained and tested with magnitudes smaller than M4 when the closest stations to the earthquakes were employed. This was due to the fact that the greater the earthquake magnitude, the higher the SNR of the corresponding seismic signal. Moreover, the difference in MAPE is even higher when a distant station is adopted. This result should be a consequence of the fact that medium and smaller seismic events can hardly be observed when their distances to the stations increase. Moreover, an average relative error equal to 4.01% (average standard deviation equal = 3.27) achieved

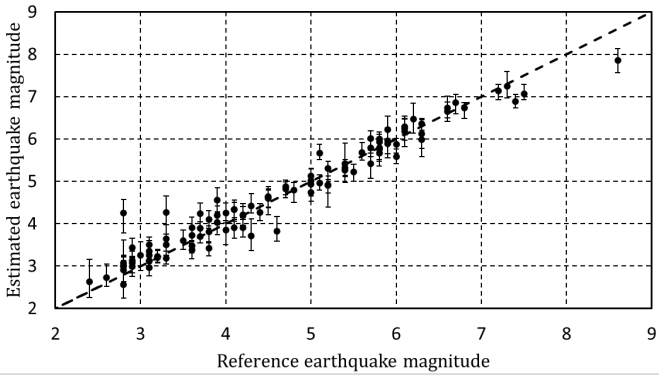

Fig. 4. Error bar graph of the average estimated magnitude versus the reference one when the closest station to each earthquake was employed.

with earthquakes greater than M4 by making use of the closest station suggests that the proposed method is applicable to provide more precious time to the population in the case of EEW and TW. This result includes the more potentially harmful earthquakes as soon as they are detected by the first seismic station, and the method needs only 0.27 s on an i7-10700 PC with GPU. It is worth highlighting that the first magnitude computation may be delivered in around five minutes after observing the earthquake in a given minimum number of stations (e.g., 10 at CSN). Compared to the final magnitude computation, which requires some extra minutes, this preliminary computation provides a MAPE equal to 4.03% (standard deviation = 3.29) in 477 seismic events greater than M4, between January 2019 and August 2021, in the same region where the database employed here was collected. This result suggests a clear motivation to employ the proposed method that can provide an earthquake magnitude estimate with a similar precision but in a much shorter time. Statistical significance analysis with a dependent t-test is provided in Table II.

Fig. 4 shows the graph of the average estimated magnitude versus the reference one when the closest station to each earthquake was employed. The proposed system was evaluated with the full database including all the available magnitudes. The error bar graph in Fig. 4 depicts the average magnitude estimation and the corresponding min-max values obtained with the testing subset. Fig. 4 basically corroborates the precision of the proposed method: all the earthquake magnitude estimations concentrate coherently along the diagonal with low dispersion, particularly with earthquakes greater than M4.

The advantages of the proposed method for earthquake magnitude estimation become clearer after comparison with similar results published elsewhere. In [9], an MAE equal to 0.26 was

achieved with a CNN-based architecture by making use of a database composed of 450.000 seismic events from STEAD, a large-scale global dataset [20], with magnitudes that span between -0.5 and 7.9. The method proposed here achieved an MAE equal to 0.23 with a similar magnitude range. However, the dataset used in this letter is much smaller, which in turn makes feasible the generation of region-dependent systems for earthquake magnitude estimation. A CNN-based solution was also presented in [11] to compute magnitudes greater than 6 obtaining an MAE equal to 0.35. The scheme presented here provides MAEs equal to 0.21 and 0.25 for seismic events greater than M4 and M6, respectively, without retraining the LSTM-MLP neural network with earthquakes greater than M6 only. In [13], an SVM-based method was applied to seismic signals from a single station and achieved an MAE equal to 0.19 with earthquakes smaller than M4 only from a database composed of 863 events. The earthquake–station distance was shorter than 120 km. The scheme proposed here reached an MAE equal to 0.23 with similar earthquake magnitudes (see Table I) but with a smaller database, that is, 250 events smaller than M4. It is worth emphasizing that in our case the average earthquake–nearest station is 196 km and there are 131 events where this distance is greater than 200 km. Moreover, the Chilean geographic extension and seismic characteristics make it impossible to use a single station. Instead, a vast and inherently heterogeneous station network is required. Consequently, the task addressed in this article seems more challenging.

V. Conclusion

An LSTM-based method was proposed to address the problem of earthquake magnitude estimation for EEW and TW purposes using only one seismic station. An end-to-end-based strategy was adopted. To reduce the TW effective time, it is necessary to estimate the magnitude of seismic events greater than M6 accurately. Nevertheless, the number of these kinds of earthquakes is low, and to counteract the drawback of limited training data, engineered features were also proposed and tested. The earthquake magnitude relative error estimation reported here in experiments with Chilean seismic data was 4.01% and 8.04% with earthquakes M4.0 or larger (up to M8.1) and M4.0 or smaller, respectively, by employing seismic traces in the nearest station to the corresponding seismic event. The average earthquake–closest station distance was 196 km, and in 26% of the data, this distance was greater than 200 km. These results are competitive with those published elsewhere and suggest that it is feasible to reduce the time required for EEW and especially TW. Also, the engineered features proposed here were dramatically superior to raw data. Thus, it is worth highlighting that the earthquakes greater than M6 are the ones that cause more fear or uncertainty in the population and provide the most significant destructive potential. Finally, increasing the earthquake database and incorporating traces from further stations are proposed for future research.

References

[1] G. Cremen and C. Galasso, "Earthquake early warning: Recent advances and perspectives," *Earth-Sci. Rev.*, vol. 205, Jun. 2020, Art. no. 103184.

[2] H. Kanamori, "Earthquake hazard mitigation and real-time warnings of tsunamis and earthquakes," *Pure Appl. Geophys.*, vol. 172, no. 9, pp. 2335–2341, Sep. 2015.

[3] G. Festa *et al.*, "Performance of earthquake early warning systems during the 2016–2017 mw 5–6.5 central Italy sequence," *Seismolog. Res. Lett.*, vol. 89, no. 1, pp. 1–12, Nov. 2017.

[4] O. M. Saad, A. G. Hafez, and M. S. Soliman, "Deep learning approach for earthquake parameters classification in earthquake early warning system," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1293–1297, Jul. 2021.

[5] T. Lay, C. Liu, and H. Kanamori, "Enhancing tsunami warning using P wave coda," *J. Geophys. Res., Solid Earth*, vol. 124, no. 10, pp. 10583–10609, Oct. 2019.

[6] S. Colombelli, A. Caruso, A. Zollo, G. Festa, and H. Kanamori, "A P wave-based, on-site method for earthquake early warning," *Geophys. Res. Lett.*, vol. 42, no. 5, pp. 1390–1398, Mar. 2015.

[7] D. Melgar and G. P. Hayes, "Characterizing large earthquakes before rupture is complete," *Sci. Adv.*, vol. 5, no. 5, May 2019, Art. no. eaav2032.

[8] S. M. Mousavi and G. C. Beroza, "A machine-learning approach for earthquake magnitude estimation," *Geophys. Res. Lett.*, vol. 47, no. 1, Jan. 2020, Art. no. e2019GL085976.

[9] N. Ristea and A. Radoi, "Complex neural networks for estimating epicentral distance, depth, and magnitude of seismic waves," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 7502305.

[10] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, Feb. 2018, Art. no. e1700578.

[11] A. Kundu, Y. S. Bhadauria, S. Basu, and S. Mukhopadhyay, "Artificial neural network based estimation of moment magnitude with relevance to earthquake early warning," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2017, pp. 1955–1959.

[12] J. Majstorović, S. Giffard-Roisin, and P. Poli, "Designing convolutional neural network pipeline for near-fault earthquake catalog extension using single-station waveforms," *J. Geophys. Res., Solid Earth*, vol. 126, no. 7, Jul. 2021, Art. no. e2020JB021566.

[13] L. H. Ochoa, L. F. Niño, and C. A. Vargas, "Fast magnitude determination using a single seismological station record implementing machine learning techniques," *Geodesy Geodyn.*, vol. 9, no. 1, pp. 34–41, Jan. 2018.

[14] Q. Wang, Y. Guo, L. Yu, and P. Li, "Earthquake prediction based on spatio-temporal data mining: An LSTM network approach," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 1, pp. 148–158, Jan. 2020.

[15] A. Berhich, F.-Z. Belouadha, and M. I. Kabbaj, "LSTM-based models for earthquake prediction," in *Proc. 3rd Int. Conf. Netw., Inf. Syst. Secur.*, Mar. 2020, pp. 1–7.

[16] V. Carbune *et al.*, "Fast multi-language LSTM-based online handwriting recognition," *Int. J. Document Anal. Recognit. (IJDAR)*, vol. 23, no. 2, pp. 89–102, Feb. 2020.

[17] A. G. Iaccarino, P. Gueguen, M. Picozzi, and S. Ghimire, "Earthquake early warning system for structural drift prediction using machine learning and linear regressors," *Frontiers Earth Sci.*, vol. 9, p. 545, Jul. 2021.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] M. Withers *et al.*, "A comparison of select trigger algorithms for automated global seismic phase and event detection," *Bull. Seismolog. Soc. Amer.*, vol. 88, no. 1, pp. 95–106, Feb. 1998.

[20] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "Stanford earthquake dataset (STEAD): A global data set of seismic signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019.