PREDICTION MODEL FOR PEAK GROUND ACCELERATION USING DEEP LEARNING

Mona Mohammed
National Research Institute of Astronomy and Geophysics (NRIAG)

Omar M. Saad
National Research Institute of Astronomy and Geophysics (NRIAG)

Arabi Keshk
Computer Science Dept., Faculty of Computers and Information, Menoufia University

Hatem M. Ahmed
Information systems Dept., Faculty of Computers and Information, Menoufia University

Abstract: Over the last decade, several studies have been proposed in the field of earthquake early warning (EEW) systems. Deep learning can be used to determine the magnitude of earthquakes and predict the PGA (peak ground acceleration). Earthquake catalogs are essential for studying fault systems, modeling seismic events, assessing seismic hazards, predicting them, and eventually decreasing seismic risk. In this work, the seismic hazard analysis is given along with the scale of ground vibration in terms of peak ground acceleration (PGA), which would be crucial for constructing earthquake-resistant structures, i.e., the PGA earthquake prediction is crucial. We propose to use artificial neural networks (ANN) and convolutional neural networks (CNN) to predict the PGA using the waveforms of weak motion velocity recorded in Japan. In this study, we use 555 events recorded by 5 seismic stations (velocity data) where the magnitude (Mg) is larger than 3. The selected earthquakes occurred between 2003 and 2022 recorded by the K-NET, Kiki-NET, and Hi-Net networks. As a result, the mean absolute error (MAE) for the test set is 18.23.

Keywords— *peak ground acceleration, earthquake early warning, convolutional neural network, artificial neural network.*

I. INTRODUCTION

A variety of earthquake early warning systems (EEW) are available, and their main function is to alert locals to take action before there is significant shaking. EEW systems might be useful. Divided into categories based on the kinds of information offered as well as the number of sensors employed [1]. However, because current regional warning methods rely on data from a variety of seismic stations and require some computation time to obtain source characteristics, there is sometimes little to no lead time before a damaging wave strikes. Depending on the concept required for determining an earthquake's parameters, the algorithms of EEW systems can be classified into onsite and regional warning ones. Basically, regional EEW systems demand the collection of P-wave data from a variety of seismic network stations nearby the epicenter region. They frequently use the waveform amplitudes and P-wave arrivals of many stations to predict the magnitude and epicenters of the earthquake's distance, and they can to determine the strength of the ground motion, utilize the ground motion prediction equations [2].

Artificial intelligence techniques have recently been used, especially for on-site EEW, to forecast the properties of the source of an impending earthquake at the stage of initial seismic waves, such as distance, magnitude of the epicenters, and seismic intensity PGA and PGV. Several P-wave properties are often collected from stations defined in the Fukushima and Ibaraki regions of Japan in the prediction model constructed using artificial intelligence methodologies for on-site EEW, as in Figure 1. Recently, a variety of cutting-edge techniques based on deep learning methods have been employed to assess the magnitude and amplitude of earthquakes [3]. Many researchers have used artificial intelligence techniques, particularly for on-site EEW, in order to predict the properties of the source of an upcoming earthquake at the stage of early seismic waves, including distance, the magnitude of the epicenters, and seismic intensity PGA and PGV.

II. RELATED WORK

T.Y. Hsu et al. [4] discussed real-time seismic intensity measurement (IM) methods with the goal of examining the current state of the field. In the beginning, we examine various theories regarding the final earthquake magnitude and rupture initiation behavior. Next, we give a brief summary of how the IM predictions have developed in relation to regional and field warnings. The use of simulated seismic wave fields and finite faults in IM forecasts is examined. The approaches for assessing IMs are finally reviewed in terms of the accuracy of the IMs determined by various algorithms and the price of alerts.

Y. Huang et al. [5] Using the first three seconds of P-wave data gathered from a single location, this study is able to estimate the peak ground acceleration (PGA) of the approaching seismic waves using CNN. In order to allow CNN to view the input data at various scales, the multi-scale input of P-wave data is proposed in this study since the amplitude of P-wave data from large and minor earthquakes might vary.

J. Münchmeyer et al. [6] directly separate transient signals from background signals for 3-component seismograms. This work creates a multi-task encoder-decoder network called Wave Decoder Net. Given its abundance of transients (tectonic and volcanic earthquakes) and diffuse ambient noise (strong microseism), we selected the active-volcanic Big Island in Hawaii as a natural laboratory. The method independently predicts the 3-component earthquake and noise waveforms from a noisy 3-component seismogram.
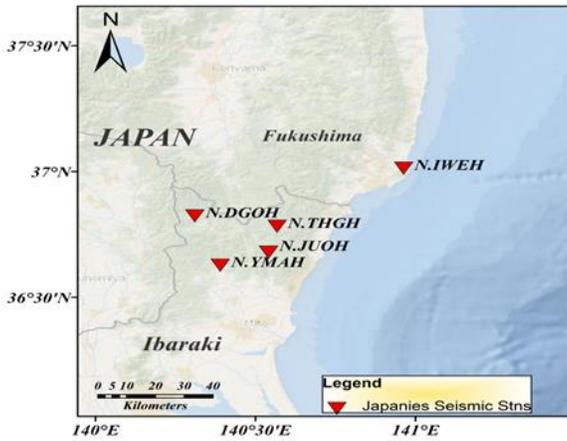
Fig. 1. describes the Fukushima and Ibaraki regions of Japan.

D. Jozinović et al. [7] used long-short-term memory (LSTM) neural networks in a deep learning strategy that may generate a highly nonlinear neural network and calculate an alert probability at each time step. Then, using two significant earthquake occurrences and one minor earthquake event that recently occurred in Taiwan, the proposed LSTM neural network is tested.

Ting-Yu Hsu et al. [8] used machine learning to take many parameters into account at once in order to account for site impacts using site parameters along with P-wave parameters. Different site effect factors are taken into consideration when building a variety of ANN models.

In their study, R. M. Allen et al. [9] outline the seismological architecture and show that a short-term hazard warning system is feasible. We demonstrate that our Earthquake Alarm System (EAS) might, with the HI Net instrumentation in place, send a warning a few to tens of seconds before harmful ground motion by using data from previous earthquakes.

The waveforms of weak motion velocity recorded in Japan will be used to train ANN and CNN to predict the PGA. For this investigation, we used velocity data from 555 seismic events with magnitudes (Mg) greater than 3 that were captured by five seismic sites. The K-NET, Kiki-NET, and Hi-Net networks captured the earthquakes that were chosen between 2003 and 2022. The test set's (MAE) as a result is 18.23.

III. EARTHQUAKE DATA

The fukushima, miyagi, and ibaraki regional regions of the Japan National Strong Motion Network of Stations (N.IWEH,

N.DGOH, N.THGH, N.JUOH, and N.YMAH), built and managed by NIED, K-NET, and KiK-net, served as the data sources for this study [10] [11]. To forecast the PGA of strong ground motion in Japan, full-waveform weak motion recodes were used. The ground-motion records for K-NET and KiK-net have been publicly available since May 1996 and October 1997, respectively, on the corresponding websites. To simplify the prediction problem, we considered an earthquake as a point source in this study and ignored earthquake source finiteness. To create the dataset, we first gathered accessible ground-motion data recorded by the Hi-Net network, The PGA value was obtained from the Strong Motion K-NET and KiK networks, which corresponded to the same earthquakes in the Japan network.

Then, for occurrences that satisfied the following requirements, week-motion data were obtained: (1) magnitude (Mg > 3), (2) epicenter distance less than 200 km, (3) event depth less than 210 km, and (4) ground-motion recordings taken from at least five sites as in Figure. 2. The top limit of Mg was selected since the effect of source quality in large earthquakes (Mg > 3) is predicted to be considerable and the assumption of a point source does not hold in extremely large earthquakes. The top limit of the event depth was selected to remove deep earthquakes from the dataset, which create aberrant intensity distributions of ground motion [12][13]. For both models, 85 percent of the data was used for training and 15 percent for testing. We divided the epicenter distance event PGA and magnitude.
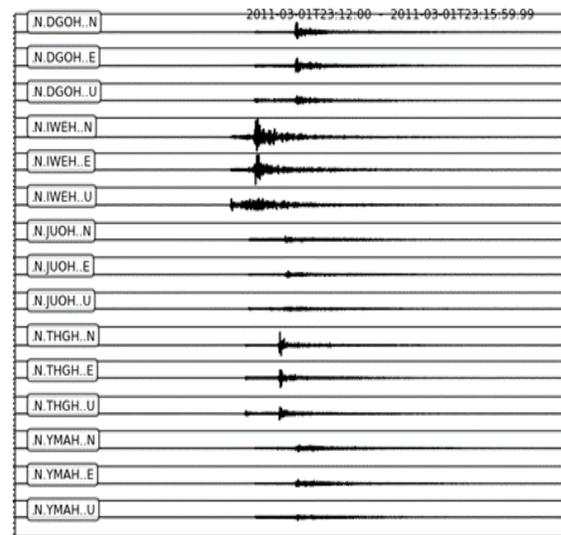

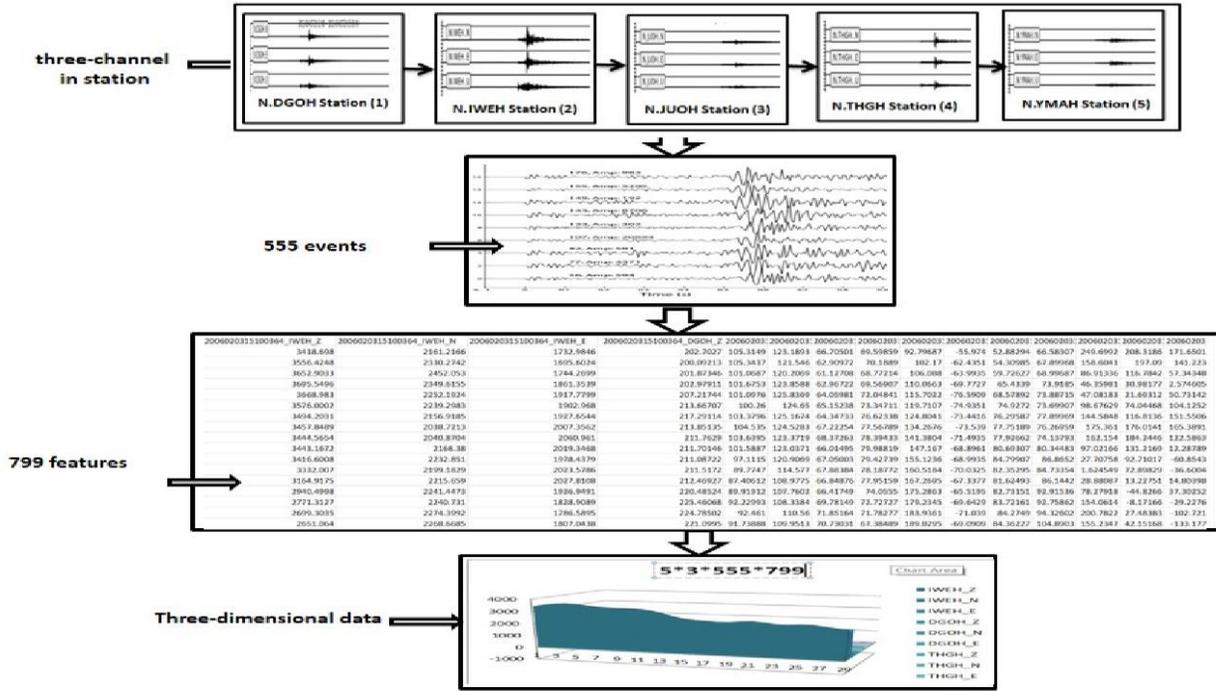Fig. 2. Give an example of an input (Earthquake M = 9).

Fig. 3 shows the data steps prior to the model stage in block diagram format.

dataset into training data with samples (555 events) ranging from 2010 to 2022 and calculated the common of the downloaded data and analyzed it by identifying the epicenters detected in the Hite catalogue, leaving a 2 second gap before and a 6 second gap after the pick [14] [15]. The data used is multi-dimensional, with 3 channels, 5 stations, 799 features, and 555 events getting entered into the model. Data on strong ground motion is being collected, employing full-waveform to predict the PGA. The data is ready for pre-processing and input into the model. Figure 3 depicts the data stages before the Moodle stage.

IV. METHODOLOGY

In this research, deep learning is used to create a prediction model with maximum PGA values using two methodologies: ANN and CNN.

A. Artificial Neural Network

Supervised deep learning using an ANN was employed to determine the maximum earthquake intensity. The full ground motion database was divided into two sections: training and validation. Because the ground motion database is organized by seismic event sequence and data with high PGA is highly regulated, two inputs of ground motion data were used for training and a modified number of inputs of ground motion data for validation. They can employ formulas for predicting ground motion to calculate the intensity of the ground motion based on the magnitude and epicenter distance of the earthquake. As a result, the training and validation data will be modified to contain essentially every seismic event, and the quantity of data with significant PGA training and validation. We used a four-layer, a fully connected, forward-looking neural network as shown in Figure 4, to build the regression model between the relevant parameters and the PGA. The input layer includes 15

neurons, five of which are seismic stations, and each station has three neurons representing channels. The concealed layer has four steps for processing inputs, and the output layer consists of five PGA seismic stations. The activation functions of the hidden and output layers were exponential, curvilinear, and linear, respectively [16]. The earthquake collection is represented by each type of input transaction, which is represented in five stations to correctly train the regression model. Each The hidden neuron and output layers were linked to a nerve cell in the layer before them. The values of neurons in each layer were transmitted to the following layer through a weighting and bias term combination. The following is the general formula for each neuron [17].

$$Zj = U \sum_{I=1}^{Nj} (WijXi + bi) \tag{1}$$

Where is the neuron's output; is the neuron's output in the preceding layer, is the weight able to connect the neuron in the preceding layer to the neuron of previous layers; is the neuron's bias, N^J is the previous layer's number of neurons; and U(x) is the function of activation. The network was trained by minimizing the cost function, defined as the mean-square error of the logarithmic difference between predicted and measured PGAs in Eq. "(1)".

$$er = \frac{1}{N} \sum_{J=1}^{N} (yj - yj^\wedge)^2 \tag{2}$$

Where and are the dataset's output and aim, respectively, and N is the total number of datasets. All regression models have the same cost functions. The Marquardt back propagation algorithm, a mix of gradient descent and Newton's method, was

used to educate the network. If the accuracy of the training subset increased while the accuracy of the validation subset remained constant or declined, the training of the ANN models was terminated to avoid undesirable over fitting. Almost identical models were trained, each with a unique set of randomly generated weights and neuronal biases. The training models' comparable best-performing network was then picked to forecast the likelihood of avoiding local minimums in Eq. "(2)".

The artificial ANN in this study is a prediction approach used to create a mathematical model of an unknown system. [18]. The multi-layer perceptron's' (MLP) most well-known class of is the ANN [19], which typically features feed-forward structures. MLPs are often taught using the back-propagation process and one input and one output layer, as well as at least one hidden layer. The network in this study is trained using a basic linear least-squares optimization approach known as the Marquardt back-propagation algorithm [20]. A regularization approach [21] is used in this procedure to reduce over fitting error, which refers to models that estimate the trained data too well while failing to forecast additional data (e.g., future observations) appropriately.

According to the process proposed by [22], the database of ground motion is divided into two subsets: training 85% and testing 15% of the database. The bias variables and connection weights are computed by the algorithm using the training subset. To avoid over fitting, the testing subset assesses the model's prediction power for data on which it has not been trained for future data.

B. convolutional neural networks

CNN are a type of neural network with a strong ability to extract features from raw data and have been effectively applied to solve numerous real-world issues. Convolution, pooling, activation, and levels that are completely interconnected are common components of a CNN. The convolution layer collects features from input data using various kernels, allowing for the extraction of a huge number of features. Users can construct specific stride widths during convolution to sweep through the given data and generate feature maps with varied weights. The pooling layer subsamples the feature maps and removes the most important information from them, resulting in a reduction in the dimensionality of the feature maps while retaining their critical information. Several convolution and pooling layers can be placed together to handle increasingly complicated issues as the CNN becomes deeper. Lastly, using the flattened feature maps, fully connected layers with activation functions are employed to perform regressions.

The PGA, the largest absolute value of the total acceleration time history in three components, was anticipated using the observed acceleration time in the first 6 seconds after triggering. It is worth mentioning that the amplitude variations between big and mild earthquakes may be rather large. While training the CNN, the amplitude of input data with relatively small values

may have less influence on the loss function than one with greater values. We tried using values of the acceleration time selection as the input of the CNN to obtain better regression results for data with different amplitudes, for example, predicting the PGA more accurately for earthquakes of varying intensities [23], but the prediction results were quite poor because the time history does not follow the lognormal distribution.

The deep learning model makes use of a CNN built using Keras (see Data and Resources). There are five convolutional layers in the network, with filters of size 64, and one fully connected layer, with VGG19 implemented with filters of size 64, and one fully connected layer, with VGG19 implemented as in Figure 5. It helps in more accurately interpreting the picture to produce the best model. For an input array size of (555, 5, 3, 799), where 5 is the number of stations, N is the number of samples (at a sampling rate of 100 samples per second), 555 is the feature, 799 is the fetcher, and 3 is the channel, the input to the model is a combination of all the waveform representations (all 5 stations) for the selected earthquake. [24]. the modified linear unit (Relu) activation function was utilised in this work to start all waveform data for each earthquake at the instant of the event's origin. [25], and the dropout operation was used to minimize over-diffing issues [26]. Normalization of waveforms (mean removal and amplitude scaling), filtering to extract seismic phases, temporal splitting, and windowing are often applied to raw seismic traces. The preprocessed data is then used for network learning and validation, which is most likely preceded by data selection to enable supervised learning [27]. The CNN network's output consists of five neurons indicating the regression value of five stations and the maximum value of the pick-ground acceleration.

V. RESULTS AND DISCUSSION

The algorithms ANN and CNN are modified, including five PGA output parameters and two input parameters, magnitude and hypocentral distance. The trial-and-error approach is used to find the best network for the given set of training data. The residuals given below are computed using all 555 data points, including training and validation.

A. Training and Validation

The goal of CNN was to predict PGA for earthquakes as accurately as possible. Some of the differences among these PGA were quite significant. To be more explicit, the PGA of huge earthquakes might be over 1,000 times greater than those of mild earthquakes. When the principle mean absolute errors were used to calculate the CNN's loss, only the PGA of bigger earthquakes were predicted with high accuracy since the errors of these earthquakes contributed significantly more to huge earthquakes, which might be over 1,000 times greater than those of mild earthquakes.
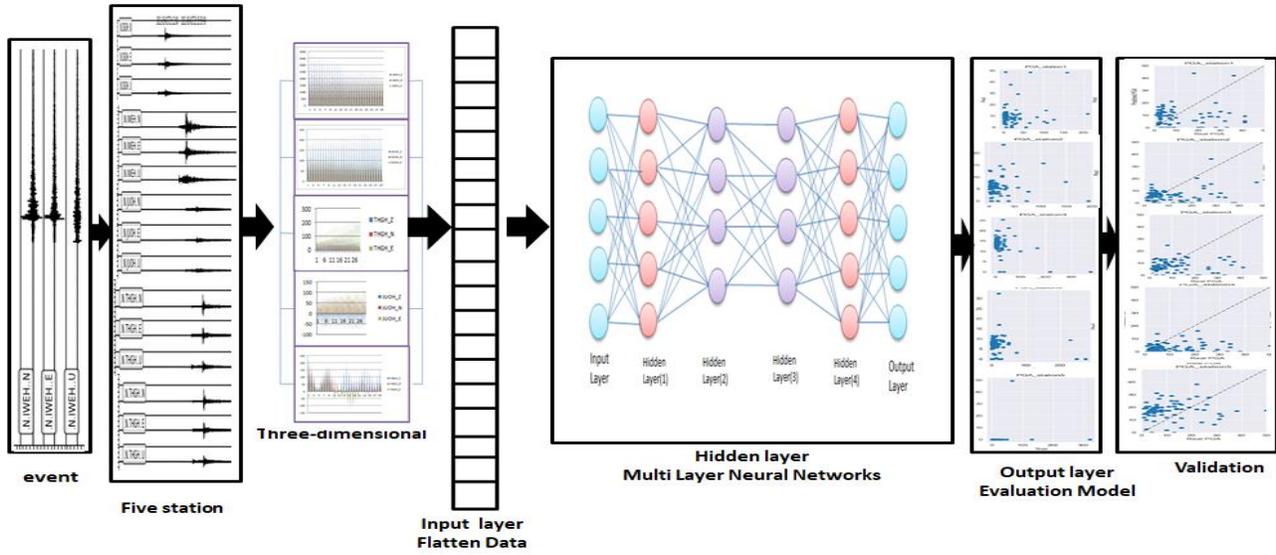
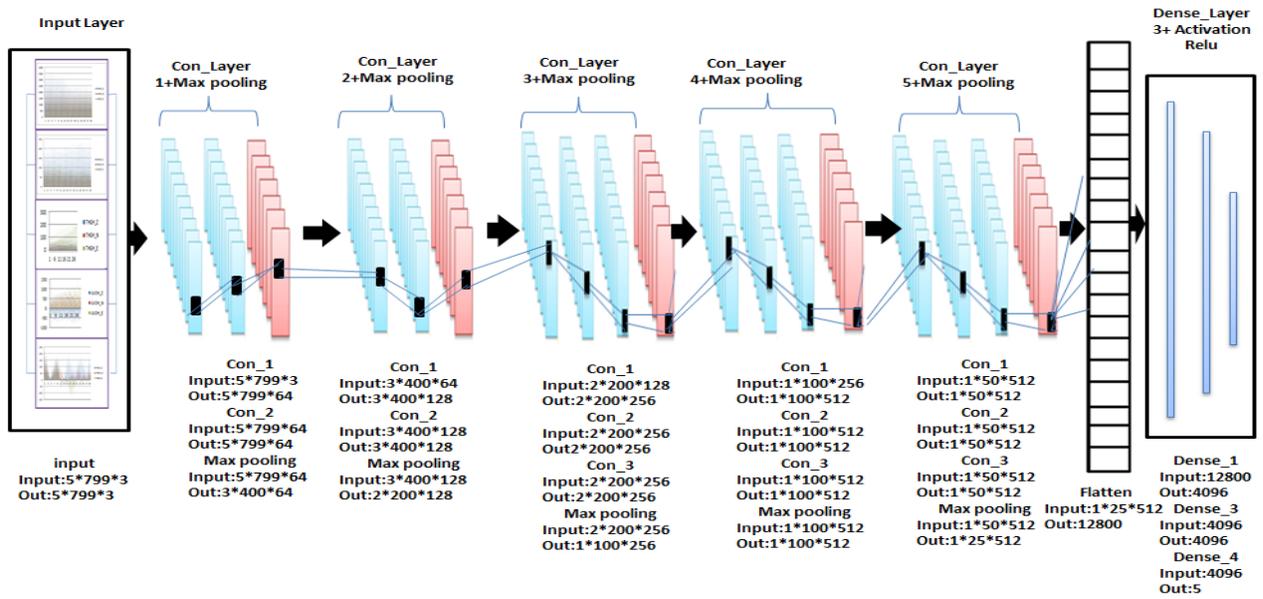Fig. 4. ANN architecture was used in this research.

Fig. 5. The proposed CNN Architecture in this Study.

The principle mean absolute errors are used to calculate the CNN loss; only the PGA of bigger earthquakes were predicted with high accuracy since the errors of these earthquakes contributed significantly more to inaccuracies in the root mean absolute than the errors of tiny earthquakes. absolute. After the processing of the CNN module, the outputs of the module Is concatenated in the merge layer of the feature-fusion module. The final prediction is generated after a fully connected layer. In this work, the principle of MAE was used to calculate the CNN loss, represented as Eq. "(3)".

The predicted value PGA of the earthquake, where N is the size of training or test samples, and predicted value and actual value, respectively [28].

$$MAE = \frac{1}{N} \sum_{L=1}^{N} |(Yi - Yj^{\wedge})| \qquad (3)$$

There are a total of N earthquakes. Each PGA range's T-data in Table 2 was selected at some point and divided into training (85%) and test (15%) sets.
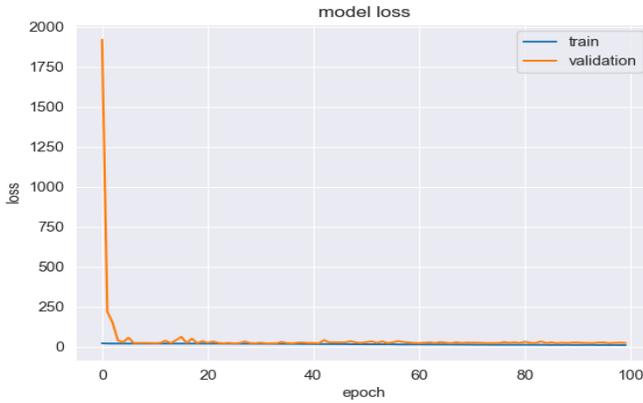
Fig. 6.   Training and Validation in CNN

Fig. 7.   Training and Validation in ANN

We used the Adam optimizer to update the CNN parameters: learning rate = 0.001, epochs = 100, validation split = 0.1. The CNN was updated during the train process by analysis, lowering the loss on a batch-by-batch basis with a batch size of 64. The transformation of feature maps into the output PGA, the training procedure was terminated when the loss of the validation dataset exceeded the loss of the training dataset for five consecutive epochs, as in Figure 6. This measure is present in total params: 83,982,149, Trainable parameters: 83,965,765. Non-trainable parameters: 16,384.

The actual dataset on which we train the model (weights and biases in the case of a neural network) for the learning model, the back-propagation network model [29] is used here, which aims at minimizing iteratively the observed target and prediction MAE output at the kth node of the pth pattern, respectively. The total variety of training patterns considered here is that at each iteration, the global MAE is minimized by adjusting the weights in each layer of the network until reaching convergence. This can be represented in the following formula [30] .

$$MAE = \frac{1}{N} \sum_{I}^{N}(Yi - Yi) \qquad (4)$$

The symbols N is the number of observations in the dataset. Yi is the true value, and Yi is the predicted value PGA of the earthquake. The learning epoch is defined as each step in the learning phase. Here, for the learning phase, the algorithm has been used that minimises E and is expressed as: where J is the error function E, I is the identity matrix, and marks the iteration step value in E. Here, an adaptable learning current is used that changes dynamically during the training stage from 0 to 1. We increase the learning rate by the factor learning increment if performance on the objective declines for an epoch. Otherwise, we adjust the learning rate by the factor of learning decline when performance increases for an epoch. Otherwise, we adjust the learning rate by the factor of learning decline when performance increases for an epoch. Throughout all FFBP simulations, we use a value of 0.0001 as the performance target. After successfully completing the Network's training phase,
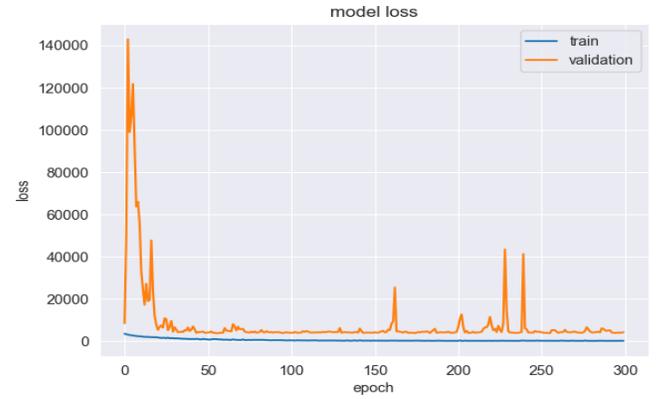
A testing dataset 15% of the total data points was used to evaluate the trained model performance in Eq. "(4)". The validation set is used to evaluate a particular model, but only on a regular basis. This information is used to fine-tune the model parameters: optimizer = 'Adam', patience = 20, batch size = 32, epochs = 300, and validation split = 0.1. As a result, the model encounters this data on occasion but never "learns" from it. The validation set results are used to update higher-level hyper parameters. As a final result, the validation set has an indirect effect on a model. This measure is present. Total params: 27,318,277, Trainable parameters: 27,310,597, Non-trainable parameters: 7,680, Figure 7.

B.   Model performance

The current work uses ANN and CNN to try to build an attenuation connection. Based on waveform ground motion, PGA has been predicted as a function of earthquake magnitude and hypocentral distance. Using raw data, multi-station waveforms, and a 6-second time window starting at the earthquake genesis time, we demonstrated that a CNN model can reliably forecast earthquakes at five stations. The proposed CNN appears to be capable of accurately forecasting the PGAs and predicting separate severe earthquake occurrences. Figure 8 Learn about the performance of PGA prediction using the CNN approach during a typical earthquake of smaller magnitude (between 3 and 9). The model has three steps: The first stage comprises the data form 5 * 3 * 799, which indicates the number of stations, channels, and features inside the waveform, and filter 64 is applied to the model. In the second stage, the hidden layer consists of five stages, and each layer is a convolutional block 3; a filter of 64 is applied, and flattening is done for the model. The dense is divided into 64, 128, 256, 512, and 512 layers, epoch 100, and a batch normalize is made for each layer with activation to be Relu. Figure 9 shows all stations, and error data were used to predict the model.
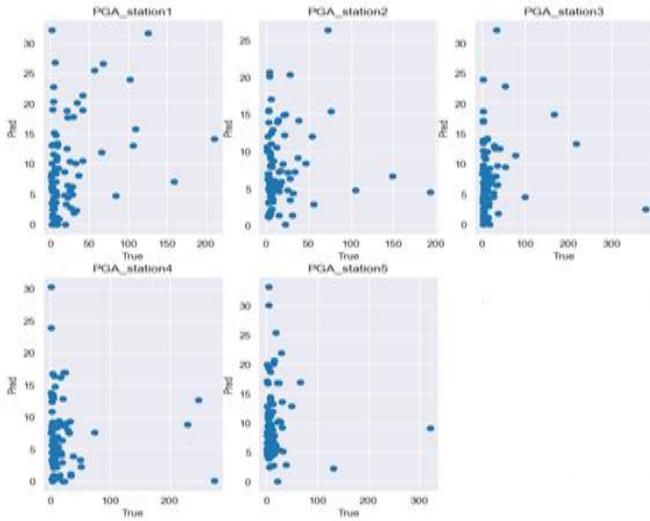
Fig. 8.   CNN predicted model and true PGA data.
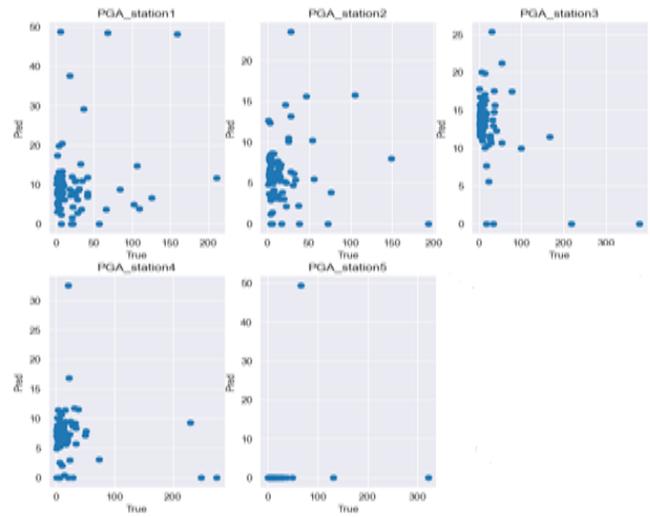

Fig. 10.   ANN predicted model and true PGA data.
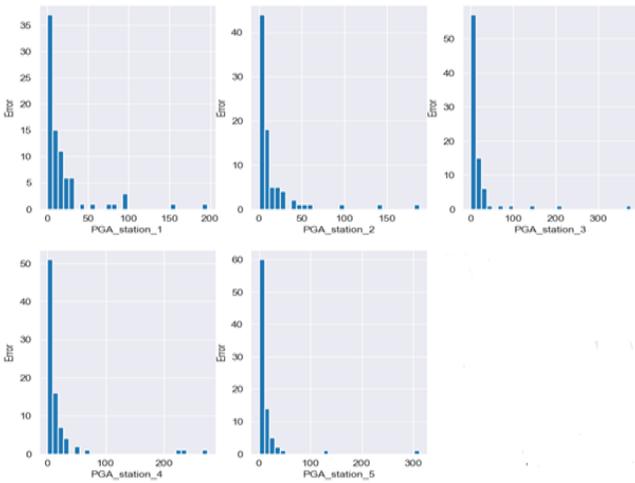

Fig. 9.   CNN predicted model in all stations and data.
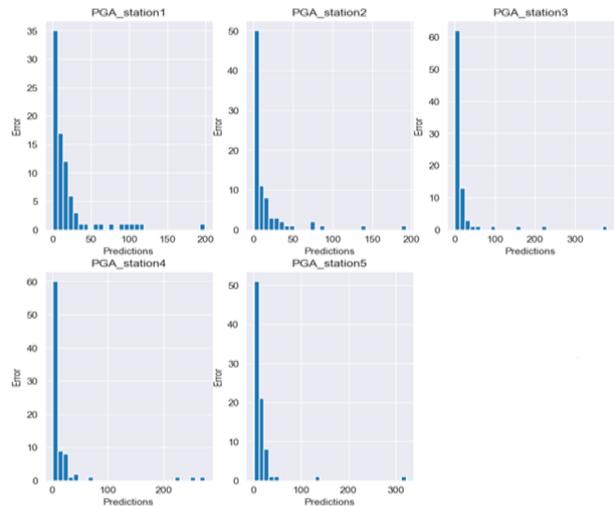

Fig. 11.   ANN Predicted model in all stations and error.

The method used in this work has two input parameters, magnitude and hypocentral distance, and one output parameter, while the output is peak ground acceleration Figure 10. The try-and-error approach is used to find the best network for a given set of training data. The model was developed as a sequential model, and it consists of three major stages: the first stage is an input to the model, as data is entered in the form of 3 * 5 * 799; the second stage is a hidden layer; and the model was designed in four stages, with a dense layer formed in each step. And it will be (2048, 1024, 512, and 256) and build a batch normalization. In the third stage, and the output will be in the form of 5 stations, reflecting the values that will occur for each station. Figure 11 presents a predicted model for all stations and true data.

C. Evaluation model

CNN to evaluate the difference in value between the predicted PGA and the actual PGA for each station, the model evaluation is shown in Figure 12. This table displays for CNN the average error value across five stations as well as the main absolute error for each station. For each station, a comparison of the ANN model's performance using the real and predicted PGA is shown in Figure 13. The average error value between the five stations is shown in this table, along with the main absolute error for each.
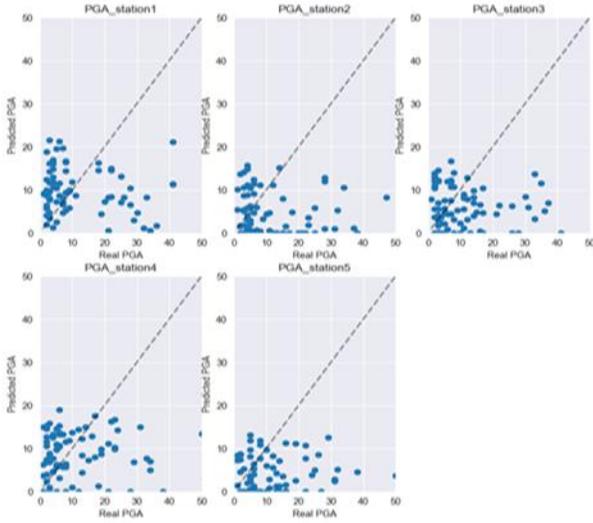
Fig. 12. PGA prediction for the 5 stations using CNN.

I. CONCLUSION

Two methods were examined in this study. The execution of the various methods was trained and tested using a set of waveform data. PGA are predicted in this work using the full waveform of data velocity and the time histories of varied window lengths of P-wave data after the trigger recodes in locations. This uses the ANN and CNN models of the 555-earthquake data set, which performed admirably, with five sites recording modest motion velocity. The K-NET, KiK-NET, and Hi-Net networks recorded the Mg > 3 earthquakes that occurred between 2003 and 2022 and had varied degrees of effect on humans and the built environment. The execution of the CNN and ANN models increased when larger time windows were used; however, the 6-s window Figure 14 used in our configuration appears to be a decent balance between accuracy and time lines.

The CNN model is split into 85% training and 15% testing, and it has a strong performance for predicting the PGAs of independent full-waveform earthquake occurrences. This model has five PGA outputs: a malty layer, a magnitude layer, an epicentral layer hidden with five stages, and these layers. The CNN model can accurately predict the PGAs of independent, entire waveform earthquake occurrences. Layers include things like convolutional layers, pooling layers, and layers that are entirely connected. These are only a few of the many components CNN uses to automatically and adaptively obtain spatial information hierarchies through back propagation. While preserving as much information in seismic waveforms as feasible, CNN is employed to automatically extract relevant characteristics from earlier P-wave data. Table 1 in the output model shows MAE, which is an acronym for all stations (18.2362).
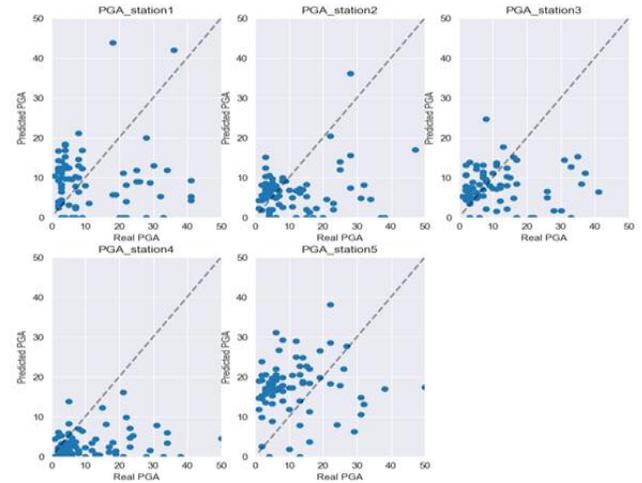

Fig. 13. PGA prediction for the 5 stations using ANN.

TABLE 1. THE MEAN ABSOLUTE ERROR (MAE) OF THE TEST SET FOR DIFFERENT STATIONS USING CNN.

Station #	MAE Proposed Algorithm	MAE Ref. [5]
1	20.77867195790722	22.54646541
2	17.10536413533347	19.89621314
3	20.520488281778636	21.4789130
4	18.02315130758853	19.5764132
5	14.753509095027333	16.8974163
Average	18.2	20.1

The ANN model was used to forecast the PGA values, utilizing the magnitude and epicentral distance of the velocity waveform as input criteria. This study shows that PGA models for restart locations in Japan may be efficiently estimated using ANN approaches. The created model has a hidden layer with four stages, a hidden layer with a malty layer, a magnitude layer, an epicentral layer, and five PGA outputs. The performance data for the model's implementation is divided into 15% testing and 85% training. The dataset used to train the weights and biases of a neural network is the actual dataset. Here, learning is done using the back-propagation network model, with the aim of lowering the MAE in Table 2 for all stations (18.3434).
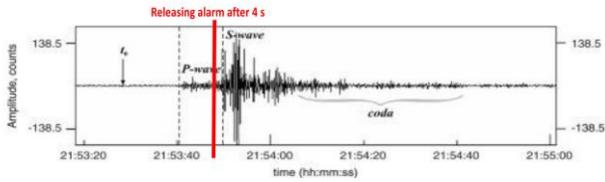
Fig. 14. Releasing alarm before the arrival of destructive wave (4s after the arrival time).

TABLE 2.The mean absolute error (MAE) of the test set for different stations using ANN.

	Station	Mean absolute error
1	Station 1	22.06624750154359
2	Station 2	15.497577278386979
3	Station 3	19.248799257335207
4	Station 4	18.291703581809998
5	Station 5	16.61311399085181
Mean for all stations		18.3

References

[1] H. Kanamori, E. Hauksson, and Heaton, "real-time seismology and earthquake damage mitigation," . annu. rev. earth planet. sci., vol. A247, pp. 529–551, december 2, 2004.

[2] R. M .Allen, , and H. Kanamori , "The potential for earthquake early warning in southern California," . Science., vol. 2. Oxford Clarendon, 1892, pp.68–73, 2003.

[3] M. Böse, T. Heaton, and Hauksson, "Rapid Estimation of Earthquake Source and Ground-Motion Parameters for Earthquake Early Warning Using Data from a Single Three-Component Broadband or Strong-Motion Sensor ," . Bull. Seismol. Soc ., vol .2 , 738–750, E, 2012.

[4] T.Y. Hsu , and A. Pratomo, "Early Peak Ground Acceleration Prediction for On-Site Earthquake Early Warning Using LSTM Neural Network ,". ORIGINAL RESEARCH ., vol. 10, 08 July, 2022.

[5] T . Hsu, and Y . Huang, " Onsite Early Prediction of PGA Using CNN with Multi-Scale and Multi-Domain P-Waves as Input," . Front. Earth Sci ., vol. 626908. doi:10.3389/feart.626908 , C.-W.,2021.

[6] J. Münchmeyer, J. Bindi, D. Leser, and U. ilmann, F "The Transformer Earthquake Alerting Model," . A New Versatile Approach to Earthquake Early Warning. Geophys., vol. 225, 646–656. , J. Int, 2021.

[7] D. Jozinović, A. Lomax, I. Štajduhar, and Michelini, "A, Rapid Prediction of Earthquake Ground Shaking Intensity Using Raw Waveform Data and a Convolutional Neural Network," . Geophys. J. Seismic Hazard for Strong Earthquakes ., vol 37. Taiwan. Nat. Hazards , 39–53.Int. 222, 1379–1389, 2020.

[8] Ting-Yu Hsu, Rih-Teng Wu , Chia-Wei Liang1 , Chun-Hsiang Kuo , and Che-Min Lin, "Peak ground acceleration estimation using P-wave parameters and horizontal-to-vertical spectral ratios," . Ocean. Sci., Vol. 31, No. 1, 1-8, February, 2020.

[9] R. M. Allen,. and H. Kanamori, "The potential for earthquake early warning in southern California," .Science., Vol. 300, 786-789, doi: 10.1126/science.1080912. 2003.

[10] S . Aoi,T. Kunugi,H. Nakamura. ,and Fujiwara, H , "Deployment of new strong motion seismographs of K-NET and KiK-net, ". In Earthquake Data in Engineering Seismology Geotechnical, Geological, and Earthquake Engineering Springer, Dordrecht. Vol. 14 (eds. Akkar, S. et al.) 167–186 2011.

[11] National Research Institute for Earth Science and Disaster Resilience. NIED K-NET, KiK-net. National Research Institute for Earth Science and Disaster Resilience. https://doi.org/10.17598/nied.0004 ,2019.

[12] T. Furumura, and B. L. N . Kennett , " Subduction zone guided waves and the heterogeneity structure of the subducted plate," .intensity anomalies in northern Japan, vol 70, 2020.

[13] H. Kubo, W Suzuki, T.Kunugi, S. Aoi , " Attenuation relationship characteristics of ground motions for deep-focus earthquakes around the Ogasawara Islands," . Japan. J. Jpn Assoc. Earthq. J. Geophys. vol. 110, B10302 ,2005. 39. Eng. 17(4), 13–29 , 2017.

[14] national strong motion data base in Japan , https://www.kyoshin.bosai.go.jp/.

[15] national week motion data base in Japan, https://www.hinet.bosai.go.jp/.

[16] K. Gurney, "An Introduction to Neural Networks," .UCL Press, London, 31 January 2017

[17] Norhisham Bakhary∗, Hong Hao, and Andrew J. Deeks, "Damage detection using artificial neural network with consideration ofuncertainties," scinice direct . Volume 29, Issue 11, Pages 2806-2815 November 2007.

[18] RYAN H. ERWIN A. B. MULDER, AND JONATHAN B. HOPKINS , "SCIENCE ROBOTICS.Vol 7, Issue 71, 19 Oct 2022

[19] J. Cybenko, "Approximation properties of some two-layer feedforward neural networks," . Opuscula Mathematica., Vol. 27, no. 1, 2007.

[20] K. Madsen, N.B. Nielsen, and O. Tingleff , "ff. Methods for nonlinear least squares problems ," . . Informatics and Mathematical Modeling,. vol. 7 , 411–412., 2004..

[21] Emtiyaz Khan, "Disentangling the Gauss-Newton Method and Approximate Inference for Neural Networks," . RIKEN Center for Advanced Intelligence Project Tokyo, Japan, July 24, 2020.

[22] Ting-Yu Hsu , and Chao-Wen Huang, " Onsite Early Prediction of PGA Using CNN With Multi-Scale and Multi-Domain P-Waves as Input,". Frontiers in Earth Science., Vol .9 . Article 626908 , 2021.

[23] Y. Yu, C. Wang ,X. Gu , and Li. J " A novel deep learning-based method for damage identification of smart building structures," .Struct. Health Monit,.vol. 18 , 1–21. , J, 2018.

[24] Kriegerowski, Marius, and et al , " A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms,". Seismological Research Letters ., vol. 90, 510-516 , 2018.

[25] V. Nair, and G.E. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, (Madison, WI: Omnipress), 807–814 , 2010.

[26] N. Srivastava,G. Hinton, A. Krizhevsky, L. Sutskever, I., and R. Salakhutdinov, " Dropout: a simple way to prevent neural networks from overfitting," . J. Mach. Learn. Res. Vol. 5, 1929–1958, 2014.

[27] Otilio Rojas , Beatriz Otero , Leonardo Alvarado, Sergi Mus3 , Ruben Tous , " Artificial Neural Networks as Emerging Tools for Earthquake Detection," Computación y Sistemas, Vol. 23, No. 2, pp. 335–350 doi: 10.13053/CyS-23-2-3197, 2019.

[28] D.P. Kingma, and J. Ba , "Adam: a method for stochastic optimization," in Proceedings of the 3rd International Conference on Learning Representations (ICLR), Conference Track Proceedings, San Diego, CA.2015.

[29] Ryoungwoo Jang, Coreline Soft Seoul, and South Korea, Learning representations by forward-propagating errors," arXiv.22, 2023.

[30] Luigi Grippo , and Marco Sciandrone, "Introduction to Methods for Nonlinear Optimization," springer ,volume 152. 2023.