Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: http://www.elsevier.com/locate/soildyn

Spatiotemporally explicit earthquake prediction using deep neural network

Mohsen Yousefzadeh [a], Seyyed Ahmad Hosseini [b], Mahdi Farnaghi [c],[*]

[a] Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran, Iran
[b] Faculty of Geography and Environmental Planning, University of Sistan and Baluchestan, Zahedan, Iran
[c] Faculty of Geo-Information Science and Earth Observation, University of Twente, the Netherlands

ABSTRACT

Due to the complexity of predicting future earthquakes, machine learning algorithms have been used by several researchers to increase the Accuracy of the forecast. However, the concentration of previous studies has chiefly been on the temporal rather than spatial parameters. Additionally, the less correlated variables were typically eliminated in the feature analysis and did not enter the model. This study introduces and investigates the effect of spatial parameters on four ML algorithms' performance for predicting the magnitude of future earthquakes in Iran as one of the most earthquake-prone countries in the world. We compared the performances of conventional methods of Support Vector Machine (SVM), Decision Tree (DT), and a Shallow Neural Network (SNN) with the contemporary Deep Neural Network (DNN) method for predicting the magnitude of the biggest upcoming earthquake in the next week. Information Gain analysis, Accuracy, Sensitivity, Positive Predictive Value, Negative Predictive Value, and Specificity measures were exploited to investigate the outcome of using a new parameter, called Fault Density, calculated using Kernel Density Estimation and Bivariate Moran's I, on the performance of the earthquake prediction, in comparison to other commonly used parameters. We discussed the behavior of the four models while dealing with different combinations of parameters and different classes of earthquake magnitudes. The results showed promising performance of the proposed parameter for the earthquakes of high magnitudes, especially using SVM and DNN models.

1. Introduction

Earthquake is a destructive natural disaster that occurs almost without any warning in advance. It inflicts plenty of casualties and financial loss to human societies. Besides, it can impose several environmental side effects such as surface fault rupture [1] and soil liquefication [2] or initiates other types of disasters like tsunamis [3], landslide [4], and fires [5]. Due to the high potential of destruction and death [6,7] as well as the direct and indirect effects of earthquakes [8], researchers have been vigorously working on the idea of proposing different approaches for earthquake prediction [9–11]. Timely and reliable forecasting can provide the possibility to consider preventive measures for mitigating the devastating effects of powerful earthquakes. Besides, such a forecast would be able to increase the level of public preparedness. A successful forecast determines the geographical location, the time, and the magnitude of an earthquake [12]. Such predictions can save many lives and vast amounts of resources. However, despite proposing various methods using different input parameters, such successful forecasts are rare amongst the past research [13].

Various methods, including mathematical modeling [14,15], hydrological [16], ionospheric analysis [17], and even procedures based on the observation of the animal behaviors [18,19], have been proposed to predict earthquakes. In another direction, a class of methods falls in the ambit of extracting useful information from the pressure wave, *P*, measured by seismographs, to predict the magnitude of an upcoming earthquake, only a few seconds to the event [20–23]. This class of methods is useful for implementing early warning systems [24], which their effectiveness is highly dependent on the accurate detection of the *P* waves and the rejection of false-positive ground vibrations caused by local activities [25]. Most of the mentioned techniques depend on the occurrence of specific precursors [26]. Nevertheless, in practice, such precursors usually either occur without any subsequent seismic events or are hard to detect, and thus those methods do not typically lead to satisfactory results [27]. Therefore, researchers have suggested that new approaches need to be considered for earthquake forecasting [28].

Meanwhile, machine learning (ML) techniques have emerged as a potent tool with undeniable advantages in dealing with data-intensive, nonlinear, and complex problems. These methods are often data-

driven, non-parametric, and less constrained by inductive assumptions [29]. Several researchers have started applying ML algorithms to solve the earthquake prediction problem [30–35]. E.g. Ref. [30], presented a probabilistic neural network model that yielded good prediction accuracies for a range of magnitudes between 4.5 and 6 [31]. also introduced a new scheme for the estimation of significant earthquake events based upon Radial Basis Function ANN, where the model was trained using leave-one-out cross-validation. In another study, researchers utilized two different quantitative association rules (QAR) and M5P to discover the temporal patterns of seismic data beneficial in earthquake prediction [33]. [35] examined the spatial-temporal variations of seismicity parameters for the Qeshm earthquake in southern Iran. After calculating seismicity parameters and normalization, Principal Component Analysis (PCA) was applied to make the data ready for being fed into the model, which was comprised of Radial Basis Function (RBF) and ANFIS [32]. devised a methodology in which the validity of the seismicity indicators could be tested using Nearest Neighbors, Naïve Bayes, Support Vector Machines (SVM), Decision Tree (DT), and Artificial Neural Networks (ANN) algorithms.

From a temporal standpoint, earthquake prediction is categorized into two general categories of forecast (months or years in advance) and short-term predictions (hours or days in advance). Earthquake forecasting is very useful for identifying the seismic gap and portions of the plate boundaries that have not ruptured in a significant earthquake for a long time [13]. However, this study focuses on short-term forecasts, directly dealing with protecting human lives and social infrastructures [36]. Short-term prediction of earthquakes is considered a challenging problem [13,37,38] due to the complex nature of earthquake phenomenon [39], the complexities of the Earth's lithosphere and its crustal blocks-and-faults structure [40] so that no specific method is yet regarded as a reliable method for such predictions [13,41].

For short-term earthquake prediction, the effective seismic parameters utilized in previous studies are often the seven parameters of $x_1, x_2, x_3, x_4, x_5, x_6, and\ x_7$, introduced by Ref. [42]; expressing the seismic facts of Bath, Gutenberg–Richter and Omori/Otsu's law and the nine parameters of $b, a, \eta, \ \Delta M, T, \mu, C, dE^{1/2}\ and\ M_{mean}$, introduced by Ref. [43]; that represent the seismic potential of the ground. Besides, Depth, latitude, and longitude of the seismic events extracted directly from the catalog data were also considered as input variables in some studies [26, 44]. Some research investigated the effective parameters for earthquake prediction [32,45]. However, these studies have primarily focused on the extent to which the dependent variables are affected by the independent variables, and in feature analysis, they sought to use the parameters that were more correlated with the output variable. The less correlated parameters were often omitted in the feature analysis process. It is noteworthy that the input parameters' influence on the results is profoundly affected by the capability of method for extracting useful information from the input parameters. Moreover, a review of the past research in the realm of earthquake prediction using machine learning methods reveals that most of these studies only consider temporal rather than spatial correlations between the dependent and independent variables [27].

Many destructive earthquakes have occurred along active fault zones or in their proximity. This observation reinforces the hypothesis that future damaging earthquakes occur mostly along active faults or within the areas where the density of the active faults is rather high [46–48]. Thus, there is a need to devise a methodology that leverages fault location data, converts it to information, examines its usefulness as an input variable to predict future earthquakes, and evaluates its impact on the prediction accuracy. Hence, the main goal of this study is to introduce and investigate the role of a spatial parameter, called Fault Density (FD), on the Accuracy of short-term earthquake prediction models that work based on ML algorithms. In particular, the performance of three well-known ML algorithms of SVM, DT, and Shallow Neural Network with one hidden layer (SNN) are compared to those of the DNN (Deep

Neural Network) algorithm for short-term prediction in a spatio-temporal setting. The proposed FD parameter is calculated by applying the Kernel Density Estimation (KDE) function on the active faults data, while the radius of the KDE is calculated through Bivariate Moran's I [49] to account for spatial correlation. The models receive effective parameters proposed by previous research [42,43,50], along with FD and predict the magnitude of the largest earthquake over the next week. Information Gain analysis (IGA), Accuracy and Sensitivity measures were exploited to assess each input parameters' explanatory power, including the proposed FD, as well as the performances of the models.

Iran, as one of the most earthquake-prone countries in the world [51, 52], was selected as the study area. The country has already experienced many large and destructive earthquakes such as Tabas (1978), Rudbar (1990), Bam (2003), and Varzaqan (2012), with the death toll of about 126000 attributed to 14 earthquakes with magnitudes of 7.0 Richter and 51 earthquakes of 6.0–6.9 Richter since 1900 [53–56]. Therefore, the need for accurate and reliable forecasting for mitigation measures is greatly sensed in the study area.

The rest of the article is organized as follows. Section 2 summarizes the theory of the four machine learning algorithms used in this study. Section 3 describes the methodology. Results and discussion are presented in section 4. Finally, section 5 concludes the study and proposes future works.

2. Machine learning algorithms

SVM is a supervised learning method based on statistical learning theory and the structural risk minimization principle [57]. As a binary classifier, SVM constructs optimal hyperplanes to separate the members of two classes while maximizing the distance between the closest samples of the classes in the training data [58]. However, in most real-world cases, the problem is not linearly separable. To handle the nonlinear cases, a kernel maps the input data to a high dimensional space, known as feature space, where the data would supposedly be linearly separable. The training points that are closest to the optimal hyperplane are called support vectors [59]. The performance of SVM highly depends on the selection of a proper kernel and the regularization constant C. Linear, polynomial, RBF (a.k.a. Gaussian), and sigmoid are four widely applied SVM kernels in the literature [59,60].

DT is a hierarchical model made up of decision rules that recursively divides the independent variables into homogeneous regions [61,62]. The purpose of a DT is to find a set of decision rules so that they can be used to predict the output from a set of input parameters. During the training process, the DT strives to obtain the maximum amount of information along with the minimum entropy generated in the tree subgroups [63]. Initially, all data is aggregated in a root node, and then it is divided into subgroups with higher purity and homogeneity using parameter values. These subsets are called internal nodes [64]. Labels are assigned to leave (terminal) nodes by an allocation strategy like majority voting [65]. In this study, the C5.0 algorithm with a boosting approach introduced by Ref. [66] is used for short-term earthquake modeling to enhance the predictive ability of the C5.0 algorithm. The core idea of the boosting approach is to create multiple classifiers rather than just one. When a new case is to be classified, each classifier votes for its predicted class. The votes are counted afterward to determine the final class [67].

ANNs have been one of the most powerful machine learning methods for predicting and modeling [68]. ANNs can learn complicated and nonlinear relationships; they do not need prior assumptions about the distribution of input data; they have proved their feasibility in dealing with noisy and incomplete data [69,70]. MLP, as a feed-forward neural network, is a well-known ANN method that has been used by several researchers for earthquake prediction [42,71,72]. An MLP model is composed of at least three layers of input, hidden, and output. The neurons are fully connected, meaning that every node in one layer is

connected to every node in the next layer [73]. MLP networks can be built with an arbitrary number of layers. However, it has been proved [74–77] and tested [78,79] that a three-layered MLP network (one input layer, one hidden layer, and one output layer) can simulate any nonlinear function up to a desired degree of Accuracy. In this study, we refer to the MLP network with three layers of input, hidden, and output as Shallow Neural Network (SNN).

DNN is a particular type of ANN with a deep structure of multiple hidden layers, attempting to model hierarchical representation beneath data and comprehend the patterns by stacking multiple layers of information processing modules in hierarchical architectures [80]. Increasing the number of hidden layers and hence adequate data transformations in deep neural network structures result in extracting the most appropriate hierarchical representation of the data [81]. In addition to their significant improvements in a variety of domains including image classification, object detection, and speech recognition [82], their generality, availability of open-source code and computer hardware for accelerating their process, mainly when the task at hand deals with abundant data, are amongst the reasons augmented the prominence of these models [83]. Different architectures of DNN have been proposed and used in different domains, e.g., Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory Neural Network. In this study, we used a deep neural network feed-forward architecture for earthquake prediction purposes.

3. Methodology

3.1. Case study

The case study of this study is Iran (the longitude between 24.5 and 40 and latitude between 43.5 and 64), a high land in the northern hemisphere, situated in the central part of the Alpine–Himalayan orogenic belt. The seismic activities of the Iranian plateau result from its position as a 1000-km-wide zone of compression between the colliding Arabian and Eurasian plates [84]. Fig. 1 shows the abundance of earthquakes in Iran during 1973 and 2019.

3.2. Data

After collecting and storing raw catalog data, from January 1973 to July 2019, from USGS[1] and IIEES,[2] the data were integrated, and the duplicate rows were identified and removed. Amongst the columns in the catalog data, only latitude, longitude, and Depth were directly taken as input variables for the prediction. In order to deal with more critical earthquakes, catalog data were filtered based on their magnitude so that events with magnitudes less than 3 Richter were eliminated. Such a filtering approach has been adopted priorly by previous studies [85,86].

Fig. 1 shows the location of the earthquake events after filtering. Events with larger magnitudes are shown in red. As can be seen, seismic events ranging from 3 to 7.7 Richter are covering the whole country. Fig. 2 also shows the frequency of seismic events by year, where we are witnessing a significant increase in the number of incidents in recent years.

After data collection, a 1×1 degree grid was constructed in the study area. In order to analyze the regions that are more prone to earthquakes, this study only considered pixels that contain at least 500 seismic events (Criterion 1: C1), and there is at least one event with a magnitude of greater than 5 Richter (Criterion 2: C2).

There were only three pixels that satisfied C1 and C2. These pixels were selected as the input pixels for the analysis. The locations of these pixels and some information about the earthquake incidences in each pixel are presented in Fig. 1 and Table 1, respectively.

---

[1] U.S. Geological Survey.
[2] International Institute of Earthquake Engineering and Seismicity.

3.3. Dependent and independent variables

The input data need to be converted into well-structured records so that we can feed them to the prediction models. Each record of data is composed of a dependent variable and several independent variables.

The output (dependent) variable represents the maximum magnitude of the next seismic event occurring in the next seven days. In this study, the problem of earthquake prediction is considered as a classification problem. The magnitude of the most massive earthquake happening in the next week is predicted as one of the four classes specified in Table 2.

Notably, previous studies have shown that if the classification of the dependent variable results in an imbalanced dataset, the performance of machine learning-based models for earthquake prediction might diminish significantly [87]. Therefore, we used the frequency distribution of the dependent variable to specify the intervals so that the class boundaries were determined by the Natural Breaks classification method [88].

The independent variables are composed of 19 parameters borrowed from previous studies, including 16 seismic parameters, latitude, longitude, and Depth accompanied by the proposed FD parameter. Overall, they constitute our 20 input variables that all had been normalized (between 0 and 1) before being used by the models. Table 3 lists the sixteen seismic input parameters proposed along with their definition.

The first parameter, named *b value*, is related to the famous Gutenberg Richter geophysical law [89]. [43] proposed this parameter and used the least-squares method to calculate it. However, due to the lack of robustness in dealing with infrequent earthquakes [42], suggested that *b* value should be calculated through maximum likelihood via Equation (1).

$$b = \log(e) \Big/ (1/n) \sum_{j=0}^{n-1} M_{i-j} - M_0 \tag{1}$$

In Equation (1), *n* is the number of events considered before the event e_i, M_{i-j} is the magnitude of e_i, *e* in the numerator is the *Euler's number* (approximately 2.718), and the cutoff magnitude is also indicated by M_0. In this study *n* was set to fifty, as suggested by previous studies [30,42, 45,90]. Having the parameter *b* calculated, the other parameters were calculated based on the description in Table 3.

In addition to the above-mentioned parameters, this study proposes a new parameter called FD to be used in short-term earthquake prediction procedures. The initial assumption is that short distances to the active faults can increase the chance of large earthquakes in the area [46]. To convey the effect of the surrounding faults, we calculated the FD by applying Kernel Density Estimation (KDE) analysis [91,92] on the faults data layer. The cardinal parameter of KDE analysis is the search radius. The proper radius of the KDE analysis is the distance that maximizes the correlation between the dependent variable and the neighborhood faults. To determine this distance, Bivariate Moran's I [49] was employed as proposed by Ref. [93]. The distance that maximizes Moran's I index between the independent variable (distance from the faults) and the dependent variable (the magnitude of the largest earthquake in the following week) is considered the proper distance of the KDE analysis. The KDE was calculated for the study area, and its value in each cell was considered as the FD parameter.

3.4. Prediction model

Fig. 3 demonstrates the overall process of the proposed short-term earthquake prediction procedure. The ultimate goal was to estimate the dependent variable, which classifies *the magnitude of the most massive earthquake happening in the next seven days*. The process started by receiving the data related to the three selected pixels. At first, the independent variables were calculated for each record of the data. Then, the data was divided into three chunks of train, validation, and test. Fifty
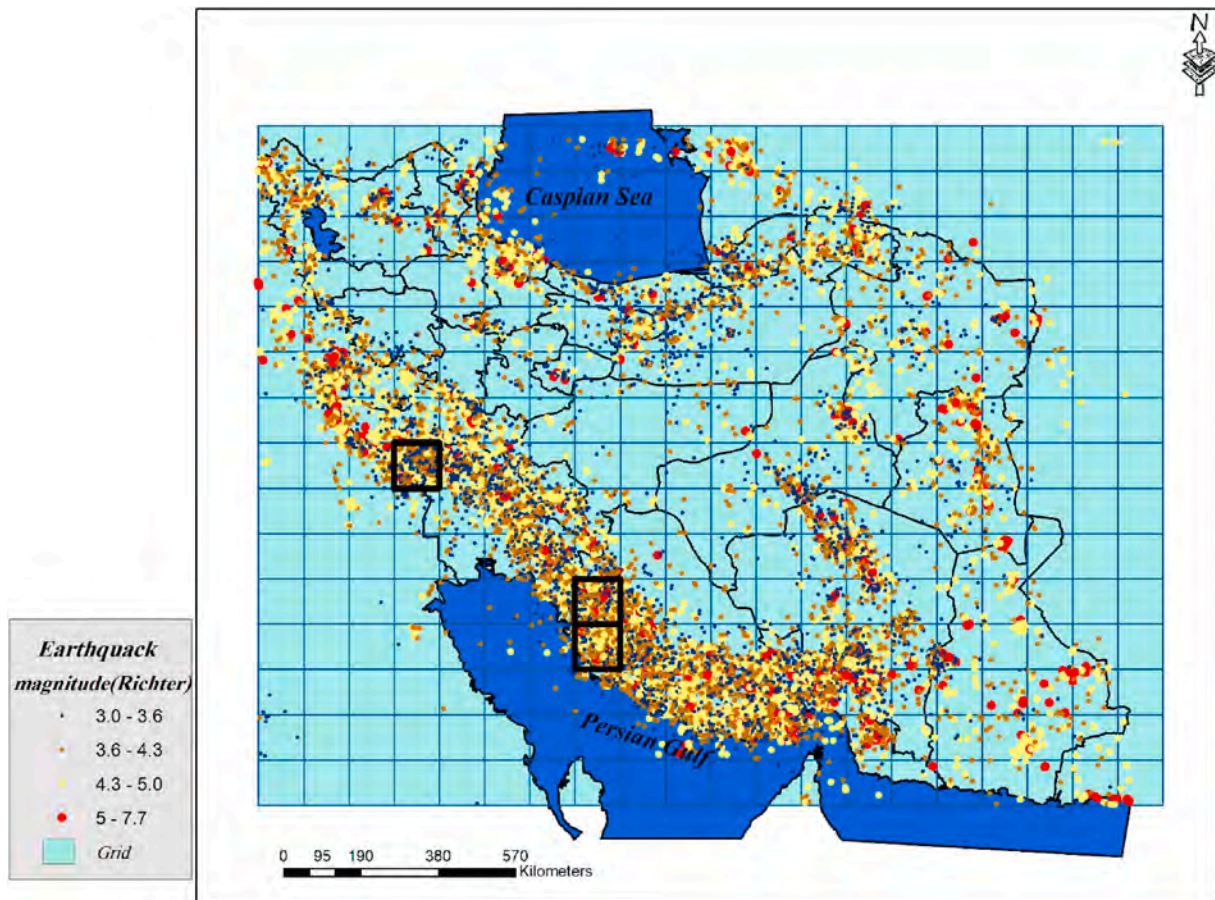
Fig. 1. Location and magnitude of earthquakes, greater than 3 Richter during 1973–2019, with grids and the selected earthquake-prone areas.
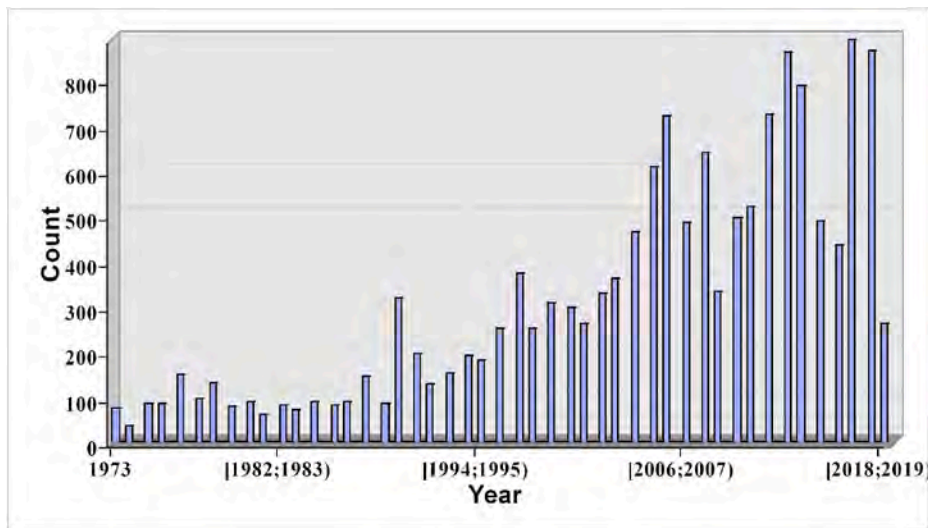
Fig. 2. Earthquake frequency histogram per year, 1973–2019.

Table 1
Statistical Information of the events within the 3 selected pixels.

Pixel (Row - Column)	Number of Events	Average of Magnitude	Variance of Magnitude	Standard Deviation of Magnitude	Max Magnitude	Min Magnitude
(11–8)	526	4.140569	0.374002	0.611557	6.2	3
(8–4)	528	3.994318	0.422207	0.649774	6.2	3
(12–8)	510	4.039216	0.319441	0.565191	6.4	3

Table 2
The boundaries of output classes.

Range (Dependent Variable)	Class	Number of Events
3–3.7	1	125
3.7–4.5	2	345
4.5–5.1	3	294
5.1–6.4	4	235

Table 3
Seismic parameters, adopted from Refs. [42,43]].

#	Feature	Description
1	b value	Gutenberg–Richter (GR) law's b value
2	X1	Increment of b between the events i and i-4
3	X2	Increment of b between the events i-4 and i-8
4	X3	Increment of b between the events i-8 and i-12
5	X4	Increment of b between the events i-12 and i-16
6	X5	Increment of b between the events i-16 and i-20
7	X6	The maximum magnitude from the events recorded during the last week (OU's law)
8	X7	Probability of events with magnitude larger or equal to 6.0, calculated as $P(M_s \geq 6) = e^{-3b \ /\log(e)} = 10^{-3b}$
9	a	Gutenberg–Richter law's a value
10	η	Sum of the mean square deviation from the regression line based on GR's law
11	ΔM	Difference between the largest observed magnitude and largest expected based on GR's law
12	T	Elapsed time which is the period between the last n events, calculated from $T = t_n - t_1$
13	μ	Average time between major seismic events (also known as characteristic events) amongst the last n events
14	C	Coefficient of variation
15	$dE^{1/2}$	Rate of the square root of seismic energy
16	M_{mean}	Mean magnitude of the last n events

percent of the data was devoted to training, twenty-five percent to validating, and the last twenty-five percent to testing. Utilization of Natural Breaks for determining the class boundaries and shuffling of the records resulted in the uniform distribution of the classes in all three training, validation, and testing subsets. In other words, all classes were uniformly represented in training, validation, and testing datasets.

The four ML algorithms of SNN, SVM, DT, and DNN were trained and calibrated using the train and validation data chunks. We used the

trained models afterward to estimate the class of the earthquakes happening in the next seven days for the test data. Finally, using the predicted and expected classes, the confusion matrix was calculated for the test dataset.

The calibration process encompasses determining the best combination of the hyper-parameters of each method. As for both SNN and DNN neural networks, the models were calibrated to achieve high generalization while mitigating overfitting. We used the Weight Decay parameter [94] for the SNN model and Dropout [95] for the DNN model to lessen the effect of overfitting. The number of layers and nodes, dropout rate, activation function, and weight decay were tuned for DNN and SNN, respectively. To achieve ideal DNN and SNN models with high performance, which neither overfit nor underfit, the models were repeatedly modified, trained, and validated on the validation data. We iteratively changed different hyperparameters of the models, including the number of layers, number of units per layer, learning rate, dropouts, and regularization. The combination that resulted in the best model performances were selected as the optimal hyperparameters. It is worth mentioning that some researchers have used metaheuristics approaches, e.g., particle swarm optimization [96], genetic algorithm [97], coronavirus optimization [98], and artificial bee colony [99], to tackle the problem of hyperparameter tuning.

Regarding SVM, the RBF kernel [59] exposed the best performance in the calibration process. The *C* parameter and the kernel width (gamma parameter) were calculated by iterating over ranges of possible values. For DT, the Trials parameter, controlling the number of boosting iterations [67], was optimized in the calibration process.

The calibration process was conducted using 4-fold-cross validation. Specifically, after separating 25% of the data for the test, the rest was divided into four equal parts. As demonstrated in Fig. 4, the training and validation were performed in four iterations so that in each iteration, three parts were used for training, and the remaining one part was used for validation. The final validation score was obtained and calculated from the average of the four validation scores.

Ultimately, after training and determining the optimal hyperparameters for the four models based on SNN, SVM, DT, and DNN using the validation score, each trained model predicted the test data that had not been fed to the models during training and validation.
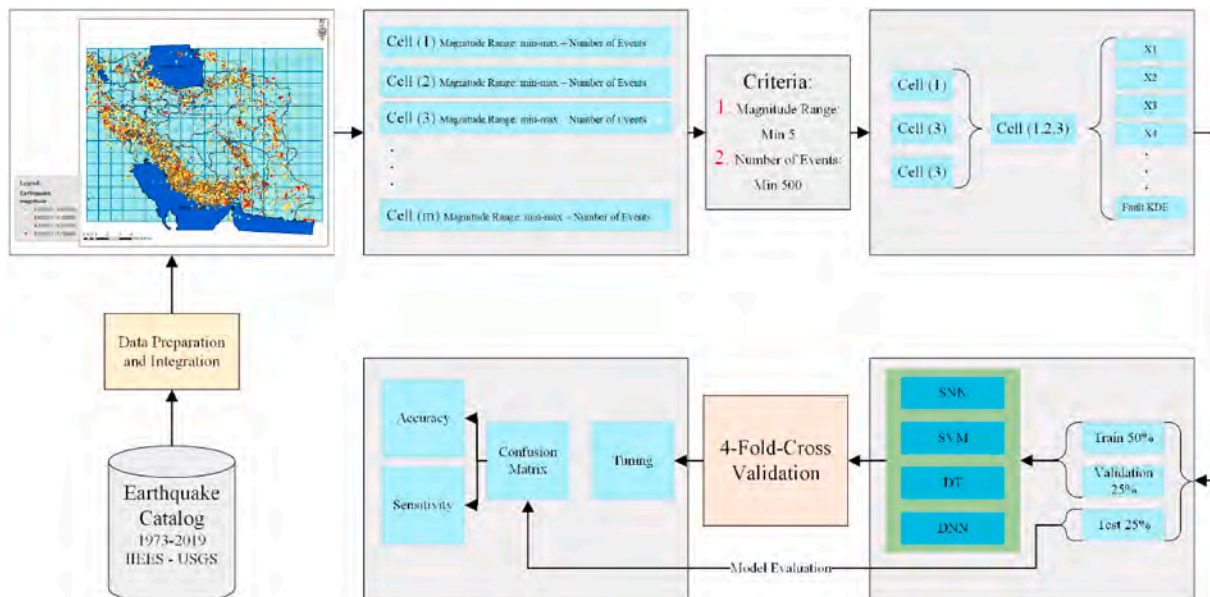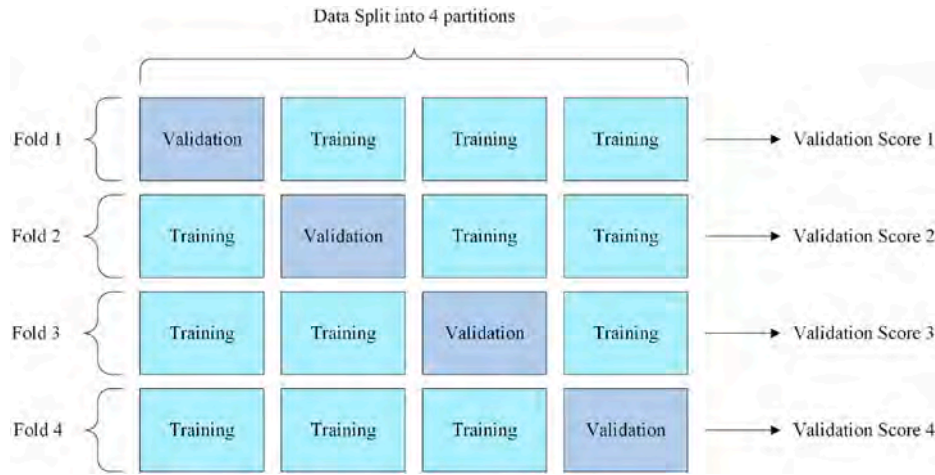

Fig. 3. Overall process.

Data Split into 4 partitions

Fig. 4. 4-fold cross validation technique.

3.5. Evaluation

IGA, Accuracy, and Sensitivity measures have been exploited in this study to evaluate the outputs. Firstly, IGA was used to 1) measure the explanatory power of each input parameter and 2) to gauge the degree to which each machine learning algorithm could take advantage of these parameters. Based on IGA, the attribute that reduces the entropy by the largest amount is considered the most significant attribute for the classification [100]. The information gain of an attribute A over the dataset S is defined as Equation (2) [101].

$$Gain(S, A) = Entropy(S) - \sum_{v \ \varepsilon \ values(A)} \frac{S_v}{S} Entropy(S_v) \qquad (2)$$

In Equation (2), $Entropy(S)$ is the entropy of the entire dataset, S_v is the subset of S for which the attribute A has the value v and $Entropy(S_v)$ is the entropy of this subset. More precisely, the entropy of S, as a measure of impurity, is calculated via Equation (3) [101].

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i \qquad (3)$$

where p_i is the probability that a particular instance belongs to the class i and c is the number of classes.

In addition to IGA, after running the models, the observation and expected values resulted from the test data were used to form the confusion matrix. Using the confusion matrix, the following parameters were calculated.

- Accuracy, as the number of events that the model has successfully predicted (Equation (4)).
- Sensitivity, as the indicator of how correctly the model has predicted the earthquakes that happened (positive class) (Equation (5)).

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \qquad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (5)$$

Furthermore, to understand the DNN model's behavior, we calculated its Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV).

- Specificity represents the rate of actual negative predictions of models (Equation (6)).

- PPV (Equation (7)) represents the ratio of actual positives (true predictions) out of all the generated earthquake predictions (positive predictions).
- NPV (Equation (8)) denotes the ratio of actual negatives amongst all the negative predictions.

$$Specificity = \frac{TN}{TN + FP} \qquad (6)$$

$$PPV = \frac{TP}{TP + FP} \qquad (7)$$

$$NPV = \frac{TN}{TN + FN} \qquad (8)$$

In Equations (4)–(8), TP, TN, FP, and FN are defined based on the confusion table as follows [102]:

- TP (true positive): An earthquake occurred and predicted by the model.
- FP (false positive): No earthquake occurred but falsely predicted by the model.
- TN (true negative): No earthquake occurred, and the model made no prediction.
- FN (false negative): An earthquake occurred, but the model was unable to predict it.

Table 4
Information Gain Values for the input parameters.

ID	Variables	Attribute Importance
1	X6	0.250
2	T	0.138
3	Latitude	0.088
4	B Value	0.079
5	X7	0.079
6	A Value	0.079
7	M_{mean}	0.079
8	Etta	0.078
9	C	0.077
10	longitude	0.072
11	dE1/2	0.068
12	M Deficit	0.048
13	**FD**	0.047
14	μ	0.042
15	**depth**	0.037
16–20	X1, X2, X3, X4, X5	0

4. Results and discussion

This section presents and discusses the result of the proposed short-term earthquake prediction models. The outputs of IGA, presented in Table 4, revealed that the FD variable, introduced in this study, has a higher value in predicting earthquakes than some other features, including X1, X2, X3, X4, X5, and Depth.

To further investigate the FD variable's role along with Depth, recognized by IGA as the spatial variables of moderate importance, we ran the four ML algorithms with different combinations of input parameters presented in Table 5. It is worth noting that in contrast to the widespread practice of excluding variables with low information gain value, we did not remove the variables X1, X2, X3, X4, and X5 from the input vector. The rationale behind not removing those variables proceeds from the idea that a variable's usefulness is proportionally dependent on the ability of the underlying model. A potent model would take advantage of the little amount of useful information coming from less significant variables and provide better predictions.

Table 6 shows the optimal hyper-parameters for the three ML techniques of SNN, SVM, and DT, while the structure of the optimal DNN together with the output shape and the number of parameters is shown in Table 7. It is worth mentioning that to find the ideal DNN model, we tested several architectures with different hyperparameters, and the model with the highest validation accuracy was selected as the best model. Some of the tested structures and the corresponding validation accuracies during 500 epochs of training were presented in Table 8 and Fig. 5. Finally, the DNN structure with 1 input layer, 6 hidden layers, and 1 output layer was selected as the best DNN structure. The output layer had 4 nodes along with the SoftMax activation function to predict the 4 classes (Table 7).

The train and test accuracy obtained by different models on the three parameter-sets are presented in Table 9 and Table 10. As shown in Table 9, the best overall test accuracy was obtained by DT, followed by DNN, SVM, and SNN, for the three parameter-sets. Considering the two parameters of Depth and FD, it seems that the two models of SNN and SVM were not able to use the latent information carried by these parameters. However, the two models based on DNN and DT were more successful in exploiting these two parameters. Meanwhile, DNN was the most successful ML algorithm in terms of utilizing the information in the FD and Depth parameters. Such an improvement by DNN could be rooted back to the deep neural structure of DNN that can extract useful information from less correlated independent input parameters.

To examine the performance of models from various aspects, in addition to Accuracy, the Sensitivity measure was calculated. Accuracy was chosen as a general metric, assessing the overall performance of the models. In contrast, we went into more detail using Sensitivity to understand better how each model performed for each class. In other words, Sensitivity signifies the capabilities of the models to correctly sense the earthquakes that occurred while Accuracy summarizes the overall performance of the classifiers. Sensitivities obtained for different classes are displayed in Table 11. Low values of the Sensitivity measure for class one (earthquakes between 3 and 3.7 Richter) and class two (earthquakes between 3.7 and 4.5) means that almost every model performed weakly in estimating these classes compared to the third and fourth classes. A reason for the deterioration of the sensitivities when it

Table 5
Examined parameter-sets.

	16 seismicity parameters (Table 3), Longitude and Latitude	FD	Depth
Parameter-set 1	*		
Parameter-set 2	*	*	
Parameter-set 3	*	*	*

Table 6
Optimal parameters of the shallow methods.

	SNN	SVM	DT
Parameter-Set1	Activation: Logistic Structure: 1 hidden, 9 neurons Decay: 1e-04 Total parameters: 211	Kernel: RBF Gamma: 0.23 Cost: 16	Trials: 40
Parameter-Set2	Activation: Logistic Structure: 1 hidden, 9 neurons Decay: 4e-04 Total parameters: 220	Kernel: RBF Gamma: 0.3 Cost: 16	Trials: 30
Parameter-Set3	Activation: Logistic Structure: 1 hidden, 9 neurons Decay: 0 Total parameters: 229	Kernel: RBF Gamma: 0.14 Cost: 256	Trials: 40

Table 7
Optimal structure of the DNN for parameter-set 3.

Layer (Type)	Output Shape	Param #
Dense (Units: 256, Activation: Tanh)	(None, 256)	5376
Dropout (0.4)	(None, 256)	0
Dense (Units: 512, Activation: ReLU)	(None, 512)	131584
Dropout (0.4)	(None, 512)	0
Dense (Units: 512, Activation: ReLU)	(None, 512)	262656
Dropout (0.4)	(None, 512)	0
Dense (Units: 256, Activation: Tanh)	(None, 256)	131328
Dropout (0.4)	(None, 256)	0
Dense (Units: 256, Activation: ReLU)	(None, 256)	65792
Dropout (0.4)	(None, 256)	0
Dense (Units: 128, Activation: ReLU)	(None, 128)	32896
Dropout (0.4)	(None, 128)	0
Dense (Units: 4, Activation: SoftMax)	(None, 4)	516

Total params: 630,148.
Trainable params: 630,148.
Optimizer: RMSprop.
Loss function: Categorical Crossentropy
Metric: Accuracy.

comes to class one and two compared to class three and four, for all models, would be a great deal of noise in the low magnitude data enfolding these classes. It is worth mentioning that some studies [102, 103] recommended that the cutoff magnitude based on the Gutenberg-Richter law should be calculated beforehand and then all the events that come below the calculated cutoff magnitude should be filtered out. The reasoning behind such a suggestion is to ensure that incomplete and misleading information is not considered in the model [102]. However, this way of calculating and applying the cutoff magnitude resulted in losing the dataset's main chunk, which was not appropriate for running the ML models. To examine the effect of cutoff magnitude on the performance, we ran the DNN model with three cutoff magnitudes of 3, 4, and 5 Richter and calculated the Accuracy. As shown in Fig. 6, the Accuracy of DNN deteriorates as we increase the cutoff magnitude. Another contributing factor could be the lower number of instances recorded for the first class (Table 2), which might have exacerbated the situation even further. Perhaps, that is why the results of the predictions for the second class are generally better than class one for all models.

The sensitivities obtained for classes three and four have been higher compared to the first two classes. A closer look reveals that these classes' highest sensitivities (three and four) came about while the models were using the second parameter-set. The underlying reason could stem from the idea that higher magnitudes are more correlated with the FD parameter since high-magnitude earthquakes are more likely to occur in areas that are closer to active faults.

Although in terms of overall Accuracy (Table 9) DT performed better than the other methods, it did not score the highest Sensitivity. The best methods for predicting classes 1, 2, 3, and 4 were SNN on parameter-set 3, SVM on parameter-set 1, DNN on parameter-set 2, and SVM on

Table 8
The structures for the DNN architectures.

Blue architecture	Green architecture	Black architecture	Yellow architecture	Purple architecture	Orange architecture	Red architecture (Best)
L1:Dense 256	L1:Dense 256	L1:Dense 256	L1:Dense 256	L1:Dense 512	L1:Dense 256	L1:Dense 256
L2: Dropout	L2:Dense 128	L2: Dropout	L2: Dropout	L2:Dense 512	L2: Dropout	L2: Dropout
L3: Dense 128	L3:Dense 4	L3:Dense 256	L3:Dense 512	L3:Dense 512	L3:Dense 512	L3:Dense 512
L4: Dropout		L4: Dropout	L4: Dropout	L4:Dense 256	L4: Dropout	L4: Dropout
L5:Dense 4		L5:Dense 256	L5: Dense 512	L5:Dense 256	L5:Dense 512	L5:Dense 512
		L6: Dropout	L6: Dropout	L6:Dense 4	L6: Dropout	L6: Dropout
		L7:Dense 4	L7:Dense 256		L7:Dense 512	L7:Dense 256
			L8: Dropout		L8: Dropout	L8: Dropout
			L9: Dense 256		L9:Dense 256	L9:Dense 256
			L10: Dropout		L10: Dropout	L10: Dropout
			L11:Dense 4		L11:Dense 256	L11:Dense 128
					L12: Dropout	L12: Dropout
					L13:Dense 4	L13:Dense 4
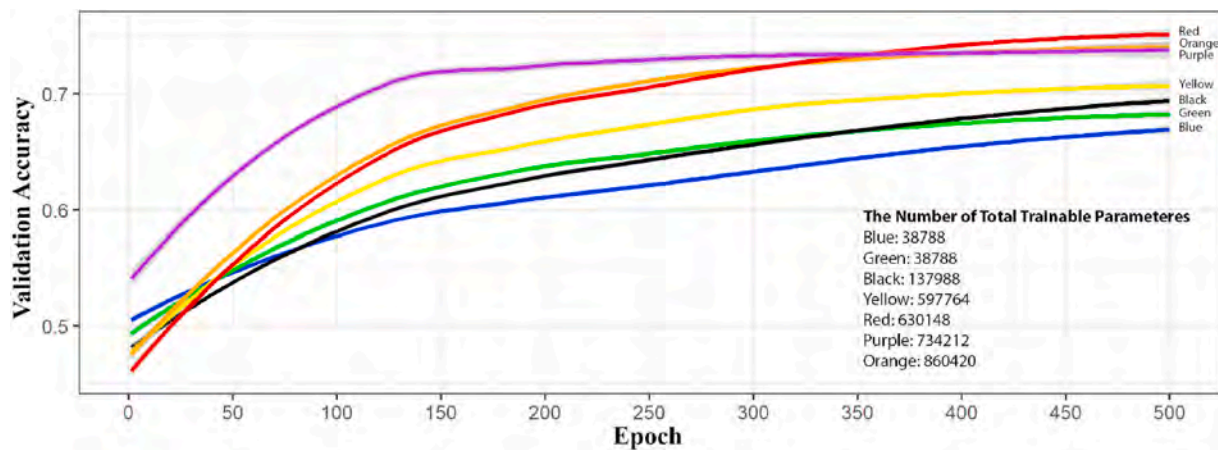
Fig. 5. Validation accuracies for different DNN architectures.

Table 9
Test data Accuracy.

	SNN	SVM	DT	DNN
Parameter-Set 1	70.4%	78%	82%	78%
Parameter-Set 2	70.0%	78%	80%	78.4%
Parameter-Set 3	61.2%	74.8%	81.2%	79.6%

Table 10
Train data Accuracy.

	SNN	SVM	DT	DNN
Parameter-Set 1	76.5%	99.7%	100%	93.4%
Parameter-Set 2	79.4%	99.8%	100%	92.1%
Parameter-Set 3	78.2%	100%	100%	93.2%

Table 11
Sensitivity.

	Class1	Class2	Class3	Class4
Parameter-Set1/SNN	32%	72.2%	68.6%	85.9%
Parameter-Set2/SNN	20.0%	81.1%	61.1%	82.4%
Parameter-Set3/SNN	**76%**	42.5%	68.6%	78.9%
Parameter-Set1/SVM	66.6%	**81.7%**	68.1%	93.7%
Parameter-Set2/SVM	66.6%	76.4%	72.8%	**97.7%**
Parameter-Set3/SVM	59%	77.3%	67%	88.4%
Parameter-Set1/DT	60%	78.2%	88%	91.2%
Parameter-Set2/DT	60%	76.2%	86.5%	87.7%
Parameter-Set3/DT	52%	79.2%	86.5%	91.2%
Parameter-Set1/DNN	56%	67.3%	88%	94.7%
Parameter-Set2/DNN	48%	69.3%	**95.5%**	87.7%
Parameter-Set3/DNN	56%	74.2%	88%	89.4%

parameter-set 2. A closer look at the results (Table 11) discloses that the best prediction of classes 1, 3, and 4 occurred in the models using the FD parameter, which indicates the suitability and usefulness of the parameter, especially for predicting earthquakes of larger magnitudes. Classes 3 and 4 can be predicted with the likelihood of more than 95% using the new FD parameter.

In other circumstances, other methods outperformed DNN. DNN is the model with the highest complexity amongst the implemented ones. Thus, in some cases, its lower accuracy and sensitivity may be due to its higher parametrization, as has already been seen before, in a study done by Ref. [30].

Remarkably, a recent literature review [104] suggested that neural network models with shallow structures can compete with DNNs in terms of their predictive power for earthquake prediction because of the

structured, tabular nature of catalog data and the limited number of calculated features. Some other studies also noted such an observation about the predictive power of SNNs [105,106]. Decision ensembles like Boosting and Random Forest, on the other hand, have attracted some attention and grown in popularity [107], where researchers compared their performances with different machine learning algorithms [30, 108]. Meanwhile, SVM has shown higher generalization ability for earthquake forecasting [109,110]. Having known the superiority of these four models, we assessed their prediction powers per class in the study area. Our results showed that when the goal is to use a general classifier to forecast earthquakes entailing both low and high magnitudes, DT would be a proper choice. However, considering the sensitivity analysis of the third and fourth classes, DNN and SVM could sense and detect moderate and high magnitude earthquakes better than other methods. Despite the network size and the considerable number of
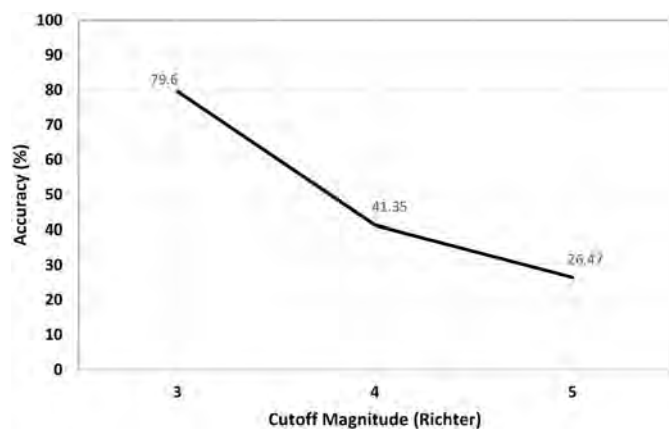
Fig. 6. The relationship between Accuracy and cutoff magnitude for DNN on parameter-set 3.

parameters needed to be trained for the DNN models, the results demonstrated that these complex models were the most successful in utilizing the information underneath the FD and Depth parameters. Moreover, DNNs outperformed other methods in predicting moderate magnitudes, though the best model in predicting low magnitude earthquakes was SNN. This behavior of SNN was expected since its structure is relatively simple, and the relationships between the input variables and the tremors of higher magnitudes are quite complex. In fact, the introduction of multiple hidden layers in DNN provides the possibility to learn features at different levels of abstraction [111].

From the disaster management organization's perspective, an earthquake prediction model should generate a few false alarms because false alarms can result in a big panic and financial loss [112]. Based on that, Specificity, PPV, and NPV were calculated per class for the DNN model (Table 12).

As shown in Table 12, the PPV value of 88% for the fourth class predicted by the DNN model is quite encouraging. There seems to be a trade-off between Specificity and NPV, indicating that when the Specificity is high, it is more likely that the classifier predicts false positives.

5. Conclusion

In this study, conventional machine learning algorithms of SNN, SVM, and DT, as well as the contemporary DNN method, were exploited to predict earthquakes in Iran. In addition to the commonly used seismic parameters described in the previous research, a new parameter named FD was also introduced, which ameliorated the Accuracy of the deep learning earthquake prediction model. The results showed satisfactory performances of DNN and SVM in predicting the classes of high magnitudes. However, the performance of DT was more promising in coping with events of both high and low magnitudes.

In the future, we will examine the usability and suitability of other deep neural network architectures, e.g., Convolutional and Recurrent Neural Networks, for earthquake prediction and compare their performance with the four algorithms of this study. Furthermore, the effect of the FD parameter on the performance of those methods will be evaluated.

Availability of data and material

The datasets are published by USGS and IIEES and publicly available through the following links.

Table 12
Specificity, PPV, and NPV for the DNN model on parameter-set 3.

	Class1	Class2	Class3	Class4
Specificity	92.4%	90.6%	92.9%	**96.3%**
PPV	45.1%	84.2%	81.8%	**87.9%**
NPV	94.9%	83.8%	95.5%	**96.8%**

- https://earthquake.usgs.gov/earthquakes/search
- http://www.iiees.ac.ir/fa/eqcatalog/

CRediT authorship contribution statement

Mohsen Yousefzadeh: Conceptualization, Methodology, Investigation, Programming, Writing – original draft, Writing – review & editing. **Seyyed Ahmad Hosseini:** Supervision, Writing – original draft, Writing – review & editing. **Mahdi Farnaghi:** Supervision, Conceptualization, Methodology, Critical commenting, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

[1] Bray JD. Developing mitigation measures for the hazards associated with earthquake surface fault rupture. In: Workshop on seismic fault-induced failures—possible remedies for damage to urban facilities. University of Tokyo Press; 2001. p. 55–79.

[2] Verdugo R, González J. Liquefaction-induced ground damages during the 2010 Chile earthquake. Soil Dynam Earthq Eng 2015;79:280–95.

[3] Jain N, Virmani D, Abraham A. Proficient 3-class classification model for confident overlap value based fuzzified aquatic information extracted tsunami prediction Intelligent Decision Technologies. 2004. p. 1–9.

[4] Keefer DK. Landslides caused by earthquakes. Geol Soc Am Bull 1984;95:406–21.

[5] Cassidy JF. Earthquake. In: Bobrowsky PT, editor. Encyclopedia of natural hazards. Dordrecht: Springer Netherlands; 2013. p. 208–23. https://doi.org/10.1007/978-1-4020-4399-4_104.

[6] Ambraseys NN, Melville CP. A history of Persian earthquakes. Cambridge university press; 2005.

[7] Bilham R. The seismic future of cities. Bull Earthq Eng 2009;7:839.

[8] Jia J. Earthquake damages. In: Modern earthquake engineering : offshore and land-based structures. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017. p. 413–31. https://doi.org/10.1007/978-3-642-31854-2_13.

[9] Florido E, Martínez-Álvarez F, Morales-Esteban A, Reyes J, Aznarte-Mellado JL. Detecting precursory patterns to enhance earthquake prediction in Chile. Comput Geosci 2015;76:112–20.

[10] Saba S, Ahsan F, Mohsin S. BAT-ANN based earthquake prediction for Pakistan region. Soft Computing 2017;21:5805–13.

[11] Tucker BE. Reducing earthquake risk Science 2013;341:1070–2.

[12] Allen C, et al. Predicting earthquakes: a scientific and technical evaluation—with implications for society panel on earthquake prediction of the committee on seismology, assembly of mathematical and physical sciences. Washington, DC: National Research Council, US National Academy of Sciences; 1976. p. 1–4.

[13] Otari G, Kulkarni R. A review of application of data mining in earthquake prediction. Int J Comput Sci Inf Technol 2012;3:3570–4.

[14] Şen Z. Point cumulative semivariogram for identification of heterogeneities in regional seismicity of Turkey. Math Geol 1998;30:767–87.

[15] Şen Z, Al-Suba'i K. Seismic hazard assessment in the Tihamat Asir region. southwestern Saudi Arabia Mathematical geology 2001;33:967–91.

[16] Hartmann J, Levy JK. Hydrogeological and gasgeochemical earthquake precursors–. A review for application Natural Hazards 2005;34:279–304.

[17] Pulinets S. Ionospheric precursors of earthquakes. recent advances in theory and practical applications Terrestrial Atmospheric and Oceanic Sciences 2004;15:413–36.

[18] Cao K, Huang Q. Geo-sensor (s) for potential prediction of earthquakes: can earthquake be predicted by abnormal animal phenomena? Spatial Sci 2018;24:125–38.

[19] Fidani C. Biological anomalies around the 2009 L'Aquila earthquake Animals, vol. 3; 2013. p. 693–721.

[20] Kanamori H. Real-time seismology and earthquake damage mitigation. Annu Rev Earth Planet Sci 2005;33:195–214.

[21] Wang Z, Zhao B. Method of accurate-fast magnitude estimation for earthquake early warning———Trial and application for the 2008 Wenchuan earthquake. Soil Dynam Earthq Eng 2018;109:227–34.

[22] Wu YM, Zhao L. Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. Geophys Res Lett 2006;33.

[23] Yamada M, Mori J. Using τc to estimate magnitude for earthquake early warning and effects of near-field terms. J Geophys Res: Solid Earth 2009;114.

[24] Wang W, Ni S, Chen Y, Kanamori H. Magnitude estimation for early warning applications using the initial part of P waves: a case study on the 2008. Wenchuan sequence Geophysical research letters 2009;36.

[25] Reiz R, Purcaru D. Using time-frequency analysis to seismic. Records Processing Journal of Electrical and Electronics Engineering 2010;3:183.

[26] Ikram A, Qamar U. A rule-based expert system for earthquake prediction. J Intell Inf Syst 2014;43:205–30.

[27] Wang Q, Guo Y, Yu L, Li P. Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. IEEE Transactions on Emerging Topics in Computing 2017.

[28] Tiampo KF, Shcherbakov R. Seismicity-based earthquake forecasting techniques. Ten years of progress Tectonophysics 2012;522:89–121.

[29] Sikder IU, Munakata T. Application of rough set and decision tree for characterization of premonitory factors of low seismic activity. Expert Syst Appl 2009;36:102–10.

[30] Adeli H, Panakkat A. A probabilistic neural network for earthquake magnitude prediction Neural networks, vol. 22; 2009. p. 1018–24.

[31] Alexandridis A, Chondrodima E, Efthimiou E, Papadakis G, Vallianatos F, Triantis D. Large earthquake occurrence estimation based on radial basis function neural networks. IEEE Trans Geosci Rem Sens 2013;52:5443–53.

[32] Asencio-Cortés G, Martínez-Álvarez F, Morales-Esteban A, Reyes J. A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. Knowl Base Syst 2016;101:15–30.

[33] Martínez-Álvarez F, Troncoso A, Morales-Esteban A. Riquelme JC Computational intelligence techniques for predicting earthquakes. In: International conference on hybrid artificial intelligence systems. Springer; 2011. p. 287–94.

[34] Martínez-Álvarez F, Morales-Esteban A. Big data and natural disasters: new approaches for spatial and temporal massive data analysis. Elsevier; 2019.

[35] Zamani A, Sorbi MR, Safavi AA. Application of neural network and ANFIS model for earthquake occurrence in Iran. Earth Sci India 2013;6:71–85.

[36] Uyeda S. On earthquake prediction in Japan. Proceedings of the Japan Academy, Series B 2013;89:391–400.

[37] Bakun W, et al. Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. Nature 2005;437:969.

[38] Hayakawa M. Earthquake prediction with radio techniques. John Wiley & Sons; 2015.

[39] Turcotte DL. Fractals and chaos in geology and geophysics. Cambridge university press; 1997.

[40] Kossobokov VG. Earthquake prediction: 20 years of global experiment. Nat Hazards 2013;69:1155–77.

[41] Ghaedi K, Ibrahim Z. Earthquake prediction earthquakes-tectonics, hazard and risk mitigation. 2017. p. 205–27.

[42] Reyes J, Morales-Esteban A, Martínez-Álvarez F. Neural networks to predict earthquakes in Chile. Appl Soft Comput 2013;13:1314–28.

[43] Panakkat A, Adeli H. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. Int J Neural Syst 2007;17:13–33.

[44] Külahcı F, İnceöz M, Doğru M, Aksoy E, Baykara O. Artificial neural network model for earthquake prediction with radon monitoring. Appl Radiat Isot 2009; 67:212–9.

[45] Martínez-Álvarez F, Reyes J, Morales-Esteban A, Rubio-Escudero C. Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula Knowledge-Based Systems 2013;50:198–210.

[46] Hetényi G, et al. Spatial relation of surface faults and crustal seismicity: a first comparison in the region of Switzerland. Acta Geodaetica et Geophysica 2018;53: 439–61.

[47] Matsuda T. Active faults and damaging earthquakes in Japan—macroseismic zoning and precaution fault zones Earthquake prediction. Int Rev 1981;4:279–89.

[48] Matsuda T. Estimation of future destructive earthquakes from active faults on land in Japan. J Phys Earth 1977;25:S251–60.

[49] Hu Z, Rao KR. Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data. Environ Health 2009;8:26.

[50] Alarifi AS, Alarifi NS, Al-Humidan S. Earthquakes magnitude predication using artificial neural network in northern Red Sea area. J King Saud Univ Sci 2012;24: 301–13.

[51] Berberian M, Yeats RS. Contribution of archaeological data to studies of earthquake history in the Iranian Plateau. J Struct Geol 2001;23:563–84.

[52] Ibrion M, Mokhtari M, Nadim F. Earthquake disaster risk reduction in Iran: lessons and "lessons learned" from three large earthquake disasters—Tabas 1978, Rudbar 1990, and Bam 2003. International Journal of Disaster Risk Science 2015; 6:415–27.

[53] Berberian M. Earthquakes and coseismic surface faulting on the Iranian Plateau. Elsevier; 2014.

[54] Haerifard S, Jarahi H, Pourkermani M, Almasian M. Seismic hazard assessment at esfaraen–bojnurd railway. North–East of Iran Geotectonics 2018;52:151–6.

[55] Jarahi H. Probabilistic seismic hazard deaggregation for Karaj City (Iran). Am J Eng Appl Sci 2016;9:520–9.

[56] Zafarani H, Soghrat M. A selected dataset of the Iranian strong motion records. Nat Hazards 2017;86:1307–32.

[57] Vapnik V. Statistical learning theory. New York: Wiley; 1998.

[58] Ghaemi Z, Alimohammadi A, Farnaghi M. LaSVM-based big data learning system for dynamic prediction of air pollution in. Tehran Environmental monitoring and assessment 2018;190:300.

[59] Pradhan B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci 2013;51:350–65.

[60] Yu L, Porwal A, Holden E-J, Dentith MC. Towards automatic lithological classification from remote sensing data using support vector machines. Comput Geosci 2012;45:229–39.

[61] Cho JH, Kurup PU. Decision tree approach for classification and dimensionality reduction of electronic nose data. Sensor Actuator B Chem 2011;160:542–8.

[62] Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. J Chemometr: A Journal of the Chemometrics Society 2004;18: 275–85.

[63] Quinlan JR. C4. 5: programs for machine learning. Elsevier; 2014.

[64] Akkaş E, Akin L, Çubukçu HE, Artuner H. Application of decision tree algorithm for classification and identification of natural minerals using SEM–EDS. Comput Geosci 2015;80:38–48.

[65] Pal M, Mather PM. An assessment of the effectiveness of decision tree methods for land cover classification. Rem Sens Environ 2003;86:554–65.

[66] Freund Y. Schapire RE Experiments with a new boosting algorithm. In: icml. Citeseer; 1996. p. 148–56.

[67] Arditi D, Pulket T. Predicting the outcome of construction litigation using boosted decision trees. J Comput Civ Eng 2005;19:387–93.

[68] Lippmann RP. An introduction to computing with neural nets. In: Artificial neural networks: theoretical concepts. IEEE Computer Society Press; 1988. p. 36–54.

[69] Kalogirou SA, Bojic M. Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy 2000;25:479–91.

[70] Vellido A, Lisboa PJ, Vaughan J. Neural networks in business: a survey of applications (1992–1998) Expert Systems with applications, vol. 17; 1999. p. 51–70.

[71] Alves EI. Earthquake forecasting using neural networks: results and future. work Nonlinear Dynamics 2006;44:341–9.

[72] Moustra M, Avraamides M, Christodoulou C. Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals. Expert Syst Appl 2011;38:15032–9.

[73] Lantz B. Machine learning with R. Packt Publishing Ltd; 2013.

[74] Baheer I. Selection of methodology for modeling hysteresis behavior of soils using neural networks. J Comput Aided Civil Infrastruct Eng 2000;5:445–63.

[75] Funahashi K-I. On the approximate realization of continuous mappings by neural networks Neural networks 2. 1989. p. 183–92.

[76] Hecht-Nielsen R. Kolmogorov's mapping neural network existence theorem. In: Proceedings of the international conference on neural networks. New York: IEEE Press; 1987. p. 11–4.

[77] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators Neural networks, vol. 2; 1989. p. 359–66.

[78] Dewapriya MAN, Rajapakse RKND, Dias WPS. Characterizing fracture stress of defective graphene samples using shallow and deep artificial neural networks Carbon. https://doi.org/10.1016/j.carbon.2020.03.038; 2020.

[79] Gordan B, Armaghani DJ, Hajihassani M, Monjezi M. Prediction of seismic slope stability through combination of particle swarm optimization and neural network Engineering with. Computers 2016;32:85–97.

[80] Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring Mechanical Systems and Signal Processing, vol. 115; 2019. p. 213–37.

[81] Van Dao D, et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. Catena 2020;188:104451.

[82] Zhong G, Ling X, Wang LN. From shallow feature learning to deep learning: benefits from the width and depth of deep architectures. Wiley Interdisciplinary Reviews: Data Min Knowl Discov 2019;9:e1255.

[83] Feng X, Yang J, Lipton ZC, Small SA, Provenzano FA, AsDN Initiative. Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification bioRxiv. 2018. p. 456277.

[84] Engdahl ER, Jackson JA, Myers SC, Bergman EA, Priestley K. Relocation and assessment of seismicity in the Iran region. Geophys J Int 2006;167:761–78.

[85] Asencio-Cortés G, Martínez-Álvarez F, Troncoso A, Morales-Esteban A. Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks Neural Computing and Applications, vol. 28; 2017. p. 1043–55.

[86] Asencio–Cortés G, Morales–Esteban A, Shang X, Martínez–Álvarez F. Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. Comput Geosci 2018;115:198–210.

[87] Bhatia A, Pasari S, Mehta A. EARTHQUAKE FORECASTING USING ARTIFICIAL NEURAL NETWORKS international archives of the photogrammetry. Remote Sensing & Spatial Information Sciences 2018.

[88] Jenks GF. The data model concept in statistical mapping. Int Yearb Cartogr 1967; 7:186–90.

[89] Gutenberg B, Richter CF. Frequency of earthquakes in California. Bull Seismol Soc Am 1944;34:185–8.

[90] Morales-Esteban A, Martínez-Álvarez F, Reyes J. Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence. Tectonophysics 2013;593:121–34.

[91] Bailey TC, Gatrell AC. Interactive spatial data analysis, vol. 413. Longman Scientific & Technical Essex; 1995.

[92] De Smith MJ, Goodchild MF, Longley P. Geospatial analysis: a comprehensive guide to principles, techniques and software tools. Troubador Publishing Ltd; 2007.

[93] Yousefzadeh M, Farnaghi M, Pilesjö P, Mansourian A. Proposing and investigating PCAMARS as a novel model for NO 2 interpolation. Environ Monit Assess 2019; 191:183.

[94] Krogh A, Hertz JA. A simple weight decay can improve generalization. In: Advances in neural information processing systems; 1992. p. 950–7.

[95] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014; 15:1929–58.

[96] Liu Z, Sun X, Wang S, Pan M, Zhang Y, Ji Z. Midterm power load forecasting model based on kernel principal component analysis and back propagation neural network with particle swarm optimization Big data 2019;7:130–8.

[97] Chung H, Shin K-s. Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability 2018;10:3765.

[98] Martínez-Álvarez F, et al. Coronavirus optimization algorithm: a bioinspired metaheuristic based on the COVID-19 propagation model big data, vol. 8; 2020. p. 308–22. https://doi.org/10.1089/big.2020.0051.

[99] Bosire A. Recurrent neural network training using ABC algorithm for traffic. Prediction Informatica 2019;43.

[100] Toolan F. Carthy J Feature selection for spam and phishing detection. In: eCrime researchers summit, 2010. IEEE; 2010. p. 1–12.

[101] Mitchell TM. Machine learning. New York: McGraw-hill; 1997.

[102] Asim KM, Idris A, Iqbal T, Martinez-Alvarez F. Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification. Soil Dynam Earthq Eng 2018;111:1–7.

[103] Asim KM, Idris A, Iqbal T, Martínez-Álvarez F. Earthquake prediction model using support vector regressor and hybrid neural networks. PloS One 2018;13: e0199004.

[104] Mignan A, Broccardo M. Neural network applications in earthquake prediction (1994–2019): meta-analytic and statistical insights on their limitations. Seismol Res Lett 2020;91:2330–42.

[105] Mignan A. Broccardo M A deeper look into 'deep learning of aftershock patterns following large earthquakes': illustrating first principles in neural network physical interpretability. In: International work-conference on artificial neural networks. Springer; 2019. p. 3–14.

[106] Mignan A, Broccardo M. One neuron versus deep learning in aftershock prediction. Nature 2019;574:E1–3.

[107] Rouet-Leduc B, Hulbert C, Lubbers N, Barros K, Humphreys CJ, Johnson PA. Machine learning predicts laboratory earthquakes. Geophys Res Lett 2017;44: 9276–82.

[108] Debnath P, et al. Analysis of earthquake forecasting in India using supervised machine learning classifiers sustainability, vol. 13; 2021. p. 971.

[109] Murwantara IM, Yugopuspito P, Hermawan R. Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years. historical data Telkomnika 2020;18:1331–42.

[110] Wang W, Liu Y, Li G-z, Wu G-f, Ma Q-z, Zhao L-f, Lin M-z. Support vector machine method for forecasting future strong earthquakes in Chinese mainland. Acta Seismologica Sinica 2006;19:30–8.

[111] Bevilacqua V, Brunetti A, Guerriero A, Trotta GF, Telegrafo M, Moschetta M. A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. Cognit Syst Res 2019;53:3–19.

[112] Asim K, Martínez-Álvarez F, Basit A, Iqbal T. Earthquake magnitude prediction in Hindukush region using machine learning techniques. Nat Hazards 2017;85: 471–86.