Nat Hazards (2017) 85:471–486
DOI 10.1007/s11069-016-2579-3

ORIGINAL PAPER

Earthquake magnitude prediction in Hindukush region using machine learning techniques

K. M. Asim[1] · F. Martínez-Álvarez[2] · A. Basit[3] · T. Iqbal[1]

Received: 28 January 2016 / Accepted: 28 August 2016 / Published online: 8 September 2016

Abstract Earthquake magnitude prediction for Hindukush region has been carried out in this research using the temporal sequence of historic seismic activities in combination with the machine learning classifiers. Prediction has been made on the basis of mathematically calculated eight seismic indicators using the earthquake catalog of the region. These parameters are based on the well-known geophysical facts of Gutenberg–Richter's inverse law, distribution of characteristic earthquake magnitudes and seismic quiescence. In this research, four machine learning techniques including pattern recognition neural network, recurrent neural network, random forest and linear programming boost ensemble classifier are separately applied to model relationships between calculated seismic parameters and future earthquake occurrences. The problem is formulated as a binary classification task and predictions are made for earthquakes of magnitude greater than or equal to 5.5 ($M \geq 5.5$), for the duration of 1 month. Furthermore, the analysis of earthquake prediction results is carried out for every machine learning classifier in terms of sensitivity, specificity, true and false predictive values. Accuracy is another performance measure considered for analyzing the results. Earthquake magnitude prediction for the Hindukush using these aforementioned techniques show significant and encouraging results, thus constituting a step forward toward the final robust prediction mechanism which is not available so far.

✉ K. M. Asim
asim.khawaja@ncp.edu.pk

F. Martínez-Álvarez
fmaralv@upo.es

A. Basit
abdulbasit1975@gmail.com

T. Iqbal
talat@ncp.edu.pk

[1] Centre for Earthquake Studies, National Centre for Physics, Islamabad, Pakistan

[2] Department of Computer Science, Pablo de Olavide University of Seville, Seville, Spain

[3] TPD, Pakistan Institute of Nuclear Science and Technology, Islamabad 45650, Pakistan

Keywords Earthquake prediction · Artificial neural networks · Time series · Machine learning

1 Introduction

Natural disasters cause massive casualties, damages and leave many injure. Human beings cannot stop them but timely prediction and due safety measures can prevent human life losses and many precious objects can be saved. Earthquake is one of the major catastrophes. Unlike other disasters, there is no particular mechanism for earthquake prediction, which makes it even more destructive. Some scientists concluded that earthquake cannot be predicted (Geller et al. 1997), while many others have suggested that it is a predictable phenomenon (Brehm and Braile 1998; Ellsworth et al. 1999; Kirschvink 2000; Knopoff 2000). According to them, several procedures can be carried out for earthquake prediction that includes the analysis of precursory phenomenon like variations of electric fields, magnetic fields and total electron content of ionosphere (Pulinets and Ouzounov 2011), animal behavior analysis (Grant et al. 2015) and historic earthquake records (Panakkat and Adeli 2007), that are well maintained in the form of catalogs.

Hindukush region is one of the most earthquake prone regions due to subducting Indian plate under the Eurasian plate. Earthquakes of significant magnitude keep hitting this region regularly at intermediate depth. Occasional deep earthquakes are also recorded in this region. A polygon-shaped Hindukush area is selected for calculation of seismic parameters and earthquake prediction. A system capable of predicting earthquake must predict about its exact location, specific magnitude range and precise time span of occurrence and probability of occurrence (Allen 1976). Such complete earthquake prediction system does not exist until now. However, different researchers have addressed these issues separately, and studies have been carried out to predict any one of the mentioned qualities. An earthquake prediction mechanism that can give confident predictions is, indeed, the need of the hour. A trigger generated by such a system can enable authorities to mobilize resources, shutdown major damage causing systems like nuclear power plants and electricity supply in order to prevent casualties.

This research of earthquake magnitude prediction encompasses a set of input parameters extracted from temporal distribution of past earthquakes. Such temporal distributions illustrate the frequency of occurrence of seismic events as function of their magnitudes (Panakkat and Adeli 2007). These parameters show the underlying relations of geophysical facts of seismic quiescence (Hainzl et al. 2000), Gutenberg–Richter law (Christensen and Olami 1992) and frequency of foreshocks (McGuire et al. 2005). This relationship between seismic activity and geophysical facts needs to be modeled, irrespective of the degree of the nonlinearity that exists among them. Seismic quiescence is break in the normal seismic energy release from the fault region. This accumulation of energy in the faults may lead to the occurrence of an earthquake, and the amount of energy stored is related to the magnitude of upcoming earthquake (Wiemer and Wyss 1994). Similarly, foreshock frequency is considered to be a sign of a major earthquake (Boore 2001). Foreshocks are the series of earthquakes of magnitude slightly higher than the background seismic activity. The Gutenberg–Richter inverse power law shows relation between the earthquake magnitudes and the cumulative frequency of events less than and equal to the corresponding magnitudes (Rundle 1989).

Machine Learning (ML) and Artificial Neural Networks (ANN) have been used in a variety of fields for prediction and classification purposes, like computer vision (Murtza et al. 2015), object recognition (Liang and Thorpe 2000), genetics (Zahur et al. 2014), bioinformatics (Larsen et al. 2012) and weather forecasting (Partal 2015). Researchers have considered using ANN for modeling of highly nonlinear and complex underlying relationship between geophysical facts and earthquakes (Adeli and Panakkat 2009; Morales-Esteban et al. 2013; Panakkat and Adeli 2007; Reyes et al. 2013) with quite meaningful results.

The core idea of this work is to predict earthquakes of magnitude 5.5 and above in Hindukush on monthly basis using ML approaches in combination with eight seismicity indicators (Panakkat and Adeli 2007). The mathematically calculated seismicity indicators from the previously occurred seismic events show the seismic behavior of the region, which are used as input to the different ML approaches. These include Pattern Recognition Neural Network (PRNN), Recurrent Neural Network (RNN), random forest and Linear Programming Boost (LPBoost) ensemble of decision trees for earthquake prediction. The prediction results of all the mentioned techniques are discussed and compared in this paper for the Hindukush.

2 Literature review

Earthquake occurrence is considered to be a random or highly nonlinear phenomenon, and there is no such existing model capable of predicting exact time, location and magnitude of earthquake. Researchers have carried out various studies over earthquake occurrences and predictions, which lead to various conclusions regarding the aspects under consideration. Famous Gutenberg and Richter mathematical model (Dahmen et al. 1998) proposed a relationship between earthquake magnitude and frequency of occurrences; this earthquake probability distribution model is useful for structural designing. Petersen et al. (2007) carried out research under the umbrella of California Geological Survey (CGS) and proposed a time-independent model showing that probability of earthquake occurrence follows poison's distribution model.

Kagan et al. (2007) considered features based on irregularities in earthquake occurrences for Southern California and predicted future seismic events of magnitude 5.0 and above for subsequent 5 years. A probabilistic earthquake forecasting model was proposed by Shen et al. (2007) on the basis of the strain examined between the tectonic plates. According to this model, higher strain observed leads to the higher probability of earthquake. Ebel et al. (2007) presented a long-term prediction model based on extrapolation of past earthquakes of magnitude greater than and equal to 5.0 for predicting future seismic events.

A number of techniques have been presented in the literature that uses ANN in combination with some seismic precursor to predict earthquakes. Negarestani et al. (2002) used Back Propagation Neural Network (BPNN) to detect anomalous behavior in radon concentration induced by earthquakes. A concentration of radon gas is present in soil measured continuously which varies due to environmental changes. Seismic activity also causes the increase in radon concentration of soil, which is differentiated successfully from ordinary changes caused by environment, using neural network. Liu et al. (2004) predicted earthquakes in China by using ensemble of Radial Basis Function (RBF) neural networks. The past earthquake magnitude data are used as an input for the network. Ikram and Qamar

(2015) presented an expert system-based automated methodology for earthquake prediction. The system considers the historic record of earthquakes as input, after dividing the whole globe into four quadrants, predicate logic and association rules are devised. The expert system is capable of predicting earthquake for time span of 24 h in each quadrant of globe.

Panakkat and Adeli (2007) introduced a remarkable approach for earthquake prediction using mathematically calculated seismic indicators from temporal distribution of recorded seismic events for Southern California and San Francisco bay regions. The model makes prediction on monthly basis, and the association between earthquake occurrence and the parameters are modeled using different ANNs. The calculation of these parameters assumed completeness of earthquake catalog, and fixed number of events are used to calculate the seismic parameters before the month under consideration. Following this research, Adeli and Panakkat (2009) used the same seismic parameters in combination with the Probablistic Neural Network (PNN) for earthquake prediction.

Morales-Esteban et al. (2013); Reyes et al. (2013) proposed different seismic parameters using mathematical calculations for earthquake prediction in Chile and Iberian Peninsula for time span of 5–7 days, respectively. These parameters are calculated considering Båth's law (1965) and Omori's law (Utsu and Ogata 1995) followed by ANN for modeling relationship between earthquake occurrences and parameters. A combination of neural networks and fuzzy logic is proposed by Zamani et al. (2013) to predict earthquakes in Iran. This work incorporates a data normalization and subsequent feature selection based on principal component analysis, for a particular set of seismicity indicators. Another model for earthquake prediction in Iran is given by Mirrashid (2014), which includes fuzzy logic along with grid partition, subtractive clustering and fuzzy C-means. This model considers earthquake prediction as a binary classification problem and predicts earthquakes of magnitude 5.5 and greater.

3 Tectonics of Hindukush region

The Hindukush is one of the most seismically active regions in the world especially for seismic events occurring at intermediate depth (70–300 km) (Pavlis and Das 2000). The region is geologically formed due to the collision of Eurasian and Indian plates during Eocene (Farah et al. 1984). Different models have been suggested about the seismicity pattern of the region. These models are mainly divided into two main categories (Searle et al. 2001). The first model suggested by the earlier studies (Billington et al. 1977) that the seismicity of the Pamir-Hindukush region is due to a single, highly twisted slab. The second and existing model states that there are two separate subducting plates, i.e., the Indian plate, on the eastern side is subducting northward under Hindukush region while Asian lithosphere under the Pamir region (Chatelain et al. 1980).

There have been continuous debates about the seismicity of this region. Earlier studies call this region as contorted 'Benioff zone' (Billington et al. 1977) because of the seismic events occurring along a 30-km-wide and slab-like deep region which is similar to oceanic subduction. However, the geology of Pamir, Hindukush and surrounding regions suggest that there is no oceanic crust in the whole region including Karakoram and Tibet (Hamburger et al. 1992). This subducting slab is strongly deformed between 71°–73°E and weakens to approximately 250 km depth. It is probably going through a continuous break off, which makes this region an exceptional seismic region in the world.

The Hindukush region is an intermediate depth seismic region; however, deep seismic events are also recorded in this region (i.e. >300 km). Different contradictory reasons have been proposed by researchers for the deep seismic activity of this region, but the concluding statement is that there is no appropriate and robust explanation available for the fascinating deep seismicity of this region so far (Koulakov and Sobolev 2006).

In this study, the polygon area is selected as shown in Fig. 1, mainly enclosing the Hindukush and some part of Pamir region is also enclosed because of subducting slab region. The latitudes and longitudes of the polygon are given in Table 1. The main idea of selecting a particular region of interest is to make sure that the geological structure and properties remain same throughout the region of interest, so that the modeling of relationship between mathematically calculated seismic parameters and earthquake occurrences shows meaningful results.

4 Calculation of seismic parameters

The seismic parameters are mathematically calculated from a catalog, so the catalog should be complete above the cutoff magnitude. Here, the cutoff magnitude refers to the earthquake magnitude, below which seismic events are not considered for parameter calculations.

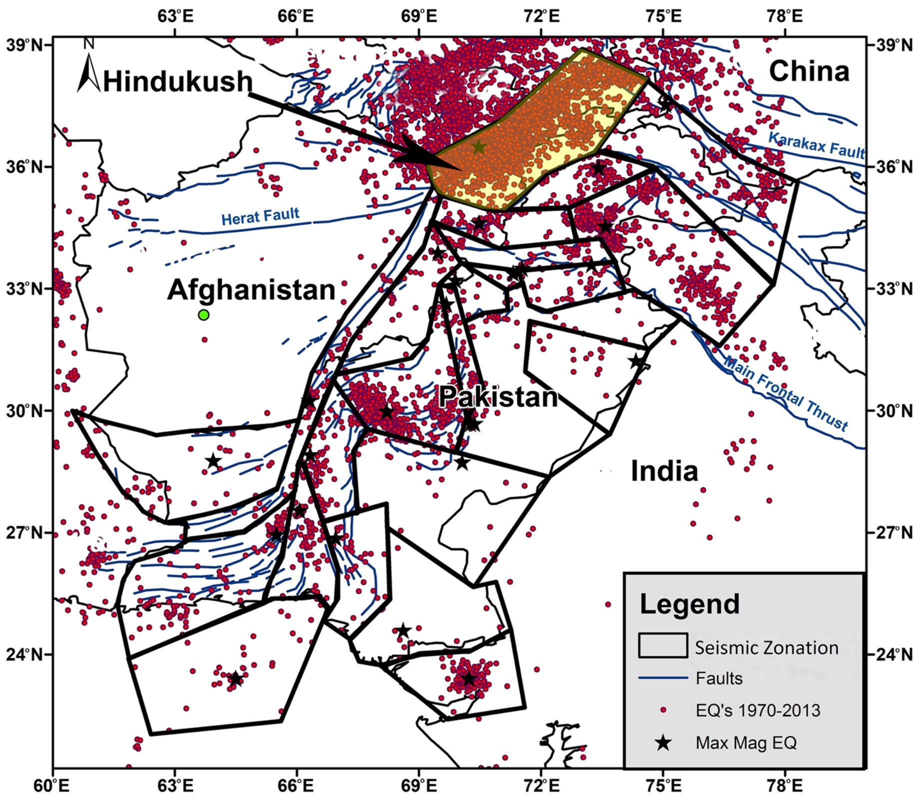

Fig. 1 Seismic zones of Pakistan along with polygon-shaped Hindukush region

Table 1 Coordinates of polygon region enclosing Hindukush region

Longitude °E	Longitude °N
69.19	36.15
71.08	37.24
73.02	38.90
74.62	38.15
73.31	36.37
72.07	35.81
71.13	34.97
69.56	35.29

4.1 Earthquake catalog

There are two main sources of earthquake catalog for Hindukush region, Center for Earthquake Studies, which maintains an internal catalog of all the seismic events happening in Pakistan along with the Hindukush and United States Geological Survey (Survey, Accessed on June, 2015). In Fig. 2, exponential rise in the occurrence of earthquakes with decreasing magnitudes shows that it follows Gutenberg–Richter's relationship, hence the catalog is complete from magnitude 4.0 and onward. Therefore for mathematical parameters calculation, seismic events of magnitude greater than and equal to 4.0 are considered in this study. There are a total of 11137 seismic events recorded from January, 1976 to December, 2013 and all of these are considered for this study. In this research, analysis is carried out on monthly basis and seismic parameters are calculated for every month on the basis of 100 seismic events before that month, as suggested by Panakkat and Adeli (2007).

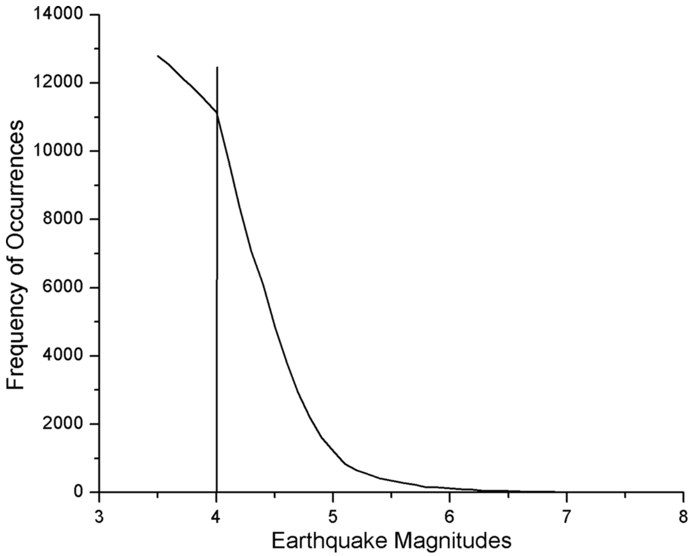

Fig. 2 Inverse log relationship between earthquake magnitudes and frequency of occurrence

4.2 Seismic parameters

This study is carried out by using the eight seismic indicators, which are basically meant to represent the seismic state and potential of the ground. This section contains the overview of all the parameters and their calculation. One of the parameters is the Time T, which is the time span over the last n number of events and n in our case is 100 and t represents the time of earthquake occurrence.

$$T = t_n - t_1 \tag{1}$$

Time T represents the frequency of foreshocks before the month under consideration. The second seismic indicator considered is the mean magnitude of the last n events. It relates to the magnitudes of foreshocks, since the magnitude M of seismic activity increases before a larger earthquake.

$$M_{\text{mean}} = \frac{\sum_i M}{n} \tag{2}$$

The rate of square root of seismic energy release dE is another seismic indicator that can be related to seismic activity through the phenomenon of seismic quiescence. Seismic energy releases gradually from fault lines through low-magnitude seismic events but if this phenomenon gets disturbed, it may lead to a major seismic event. The equation for square root of seismic energy released is given below:

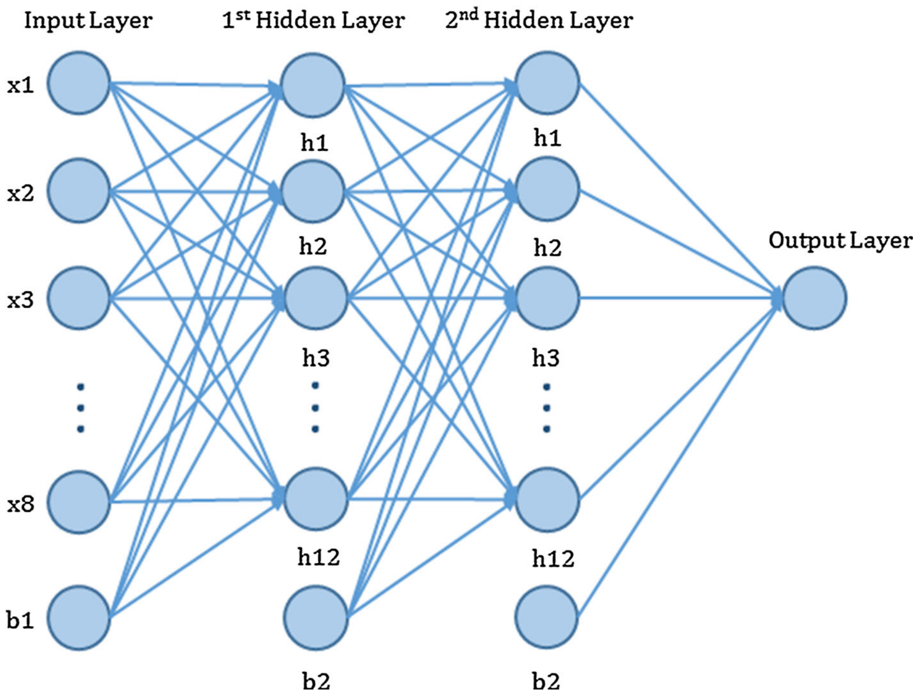

Fig. 3 Structure of pattern recognition neural network for earthquake prediction

$$dE^{\frac{1}{2}} = \frac{\sum(10^{11.8+1.5M})^{\frac{1}{2}}}{T} \tag{3}$$

The energy released obtained through Eq. 3 is represented in ergs. b value is another important seismic parameter obtained from the famous Gutenberg–Richter inverse power law Rundle (1989). It is defined as the slope of the curve between earthquake magnitudes and the log of frequency of earthquakes occurred. The equation of Gutenberg–Richter curve is given in Eq. 4, and the parameters a,b are obtained by using Least Squares method.

$$\log N = a - bM \tag{4}$$

Deviation of actual data from the Gutenberg–Richter inverse power law (Eq. 4) is also considered as a seismic indicator. Summation of mean square deviation is shown in Eq. 5. Lower is the mean square deviation ; greater is the conformance of the data and more it is likely to be predicted by inverse power law.

$$\eta = \frac{\sum(\log N - a - bM)^2}{n - 1} \tag{5}$$

The difference between the maximum observed and the maximum occurred earthquake magnitude is also considered as a seismic indicator. The maximum observed event is listed in the catalog, while maximum expected event is obtained using Eq. 6.

$$M_{\text{expected}} = \frac{a}{b} \tag{6}$$

where a is the y-intercept in inverse power law obtained from Eq. 4. Mean time between characteristic events (μ) among the last n events is also considered as a seismicity precursor. The magnitude of different events grouped together as a single magnitude is considered as characteristic magnitude. For example, events of magnitudes 4.5–5 are considered as ofsingle magnitude known as characteristic magnitude, and the mean time between them is calculated using Eq. 7.

$$\mu = \frac{t_{i_{\text{characteristic}}}}{n_{\text{characteristic}}} \tag{7}$$

Mean time between characteristic events should be equal ideally. Similarly, deviation from this mean time (c) is also taken as a seismic parameter obtained from Eq. 8.

$$c = \frac{\sigma_{\mathrm{t}}}{\mu} \tag{8}$$

where σ_{t} is the standard deviation of observed time.

5 Machine learning techniques for earthquake prediction

Different machine learning approaches have been applied to the eight seismic parameters explained in Sect. 4. The prediction task is treated as a binary classification problem with earthquakes of magnitudes 5.5, and larger are considered as *Yes* or 1 and below as *No* or 0. After the training of these techniques, output is obtained on unseen data parameters, and then performance is evaluated in Sect. 6.

5.1 Pattern recognition neural network

Neural networks are created to work like human nervous system. The network contains sensory units called neuron which are interconnected through weighted connections. The purpose of weights is either to attenuate or amplify the signal coming from neurons. The ability of neural networks is to model highly nonlinear relationships. The revival of neural networks in the fields of Artificial Intelligence (AI) and Pattern Recognition (PR) is of great importance, although PR is not the only application of neural network. It is about making machines to learn and distinguish different patterns and classes automatically (Martin 1995). The purpose of the training the network is to associate outputs with the given inputs and to learn the pattern or underlying relationship that exists between inputs and corresponding outputs. After training, network is provided with unseen inputs without mentioning the outputs, so results predicted by the network on the basis of training are evaluated to check the power of network. In this case, we have used PR topology for binary classification purposes by applying a suitable threshold on the output of network. The training methodology used for training of neural network is Levenberg–Marquardt Backpropagation (LMBP) (Bishop 1995). It is used in combination with an optimization technique such as gradient descent. It first calculates the error of calculated output with respect to the actual one and propagates the error backwards and updates weights. LMBP algorithm is faster as compared to other algorithms and reaches convergence anywhere between 10 and 100 times quicker than the standard BP algorithm, but it requires more memory (Bishop 1995).

Eight seismic parameters are passed as an input to eight neurons and one set of input features corresponds to 1 month. In this case, two hidden layers are used, each with 12 neurons as shown in Fig. 3. This is selected manually on the basis of experimentation to get the best results. The transfer function used for layer one is tan-sigmoid and for layer two is log-sigmoid as shown in Eqs. 9 and 10, respectively.

$$T_1(n) = \tanh(cn) = \frac{e^{cn} - e^{-cn}}{e^{cn} + e^{-cn}} \tag{9}$$

$$T_2(n) = \frac{1}{1 + e^{-cn}} \tag{10}$$

The total number of synaptic connections are calculated using Eq. 11, where H1 and H2 represent the hidden number of neurons in layer 1 and layer 2, while I and O are the number of inputs and outputs, respectively. According to specifications of the neural network mentioned above, there are 277 synaptic connections in the network. Matlab 2013a has been used to train and test the algorithm.

$$w = (I + 1)H_1 + (H_1 + 1)H_2 + (H_2 + 1)O \tag{11}$$

5.2 Recurrent neural network

Recurrent Neural Network (RNN) is another topology of ANN, developed in 1980s. The structural difference between RNN and other topologies of ANN is that directed cycle exists between the units along with the presence of recurrent layer. The internal cycle enables this network to store internal state unlike other neural networks. RNN is suitable for sequence processing problems, like handwriting recognition, speech recognition (Graves et al. 2013) because of its ability to store internal temporal state.

The leaning paradigm used to train RNN is also chosen to be LMBP algorithm, because it outperformed other learning methodologies during experimentation. Input is passed to eight input neurons, which are intern carried to hidden layers. On the basis of exploration, the number of hidden layers are chosen to be two and the neurons in each layer are selected to be six and seven, respectively, as shown in Fig. 4. The transfer function for both the hidden layers is same as PRNN as mentioned in Eqs. 9 and 10.

The performance and accuracy of training a LMBP algorithm is determined by the number of iterations carried out before stopping, the number of epochs and the value of the mean square error function at which training stops, and the minimum magnitude of the gradient below which training stops.

5.3 Random forest

Decision trees are supervised machine learning techniques and have been used for classification and regression purposes. The performance of decision trees can be boosted by using ensemble learning methodologies. Random forest refers to the large number of decision trees, merged through bootstrap aggregating or bagging. The idea of random forests was proposed by Breiman (2001). Simple decision trees usually over train or

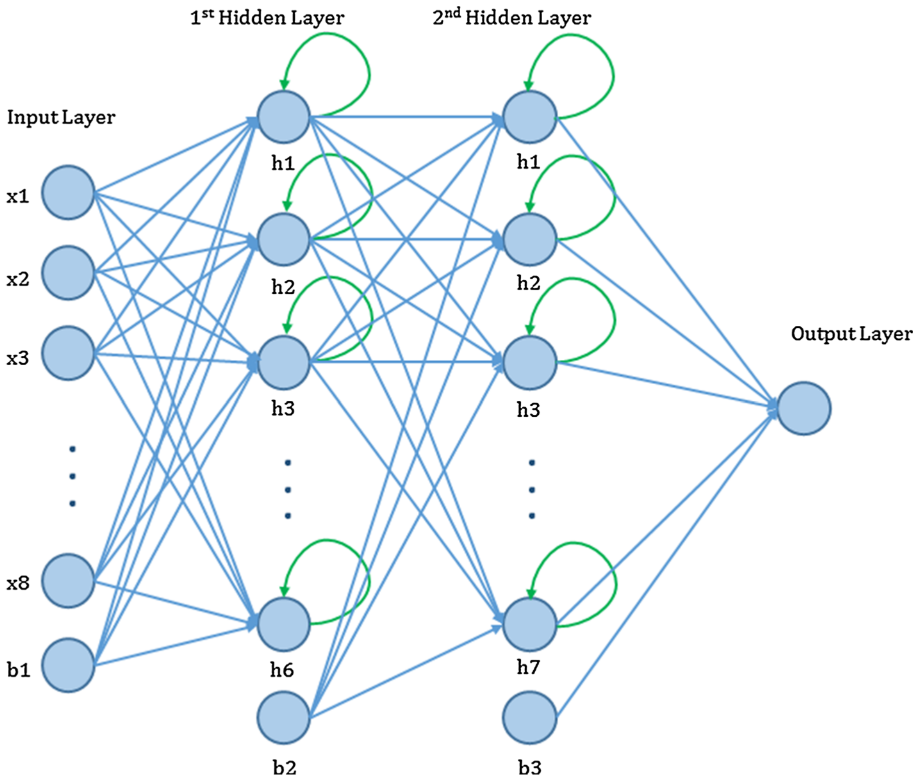

Fig. 4 Structure of recurrent neural network for earthquake prediction

memorize the data and random forest is one of the way of preventing this attribute thus reducing variance of the final predictive model. It includes the averaging of different uncorrelated decision trees trained on sampled dataset (Hastie et al. 2005). The choice of number of decision trees is decided on the basis of experimentation on dataset. In our case, 50 numbers of trees have been chosen for creating ensemble on the basis of experimentation.

5.4 Ensemble of trees using LPBoost

Boosting is another ensemble method used to enhance the performance of weak learners, i.e., trees. Robert E. Schapire first gave the idea of enhancing the performance of weak learners using a methodology, which was later developed into boosting (Schapire 1990). There are different types of boosting based upon the weighting methodologies. LPBoost is one of the boosting types. The idea of LPBoost is to maximize a margin between training instances of different classes and hence belongs to the class of margin maximizing supervised classification algorithms (Ganatra and Kosta 2010). It is a linear combination of many tree classifiers. Each classifier is iteratively added to the set of selected classifiers until no other tree needs to be added.

6 Results and discussion

All the ML classification techniques presented in Sect. 5 are supervised and need to be trained before real-time prediction. The eight parameter dataset is separated into train and test sets in the ratio of 70 and 30 %, respectively. This section contains the details about the dataset, its training and testing.

6.1 Dataset description

The dataset consists of 441 data vectors, with each vector corresponding to every month from April, 1977 to December, 2013. Training of all four above-mentioned techniques is performed on 309 data vectors, which constitutes almost 70 % of the dataset. Evaluation is performed on the rest of 132 data vectors which is approximately 30 % of the dataset. Corresponding to each data vector, there exists a label with a binary value of either 1 or 0, because the authors are treating it as a binary classification problem. Label '1' shows the occurrence of an earthquake of magnitude greater than or equal to 5.5, while label '0' shows either absence of seismic activity or earthquake <5.5 magnitude. The prediction using these parameters is made over a time span of 1 month. This monthly based prediction is decided after rigorous earthquake catalog analysis. Low-level seismic events (Earthquake Magnitude <5.5 in our case) keep hitting the region on weekly basis which are not hazardous, but do participate in modification of seismic parameters for next month. The major earthquakes ($M \geq 5.5$ in our case) strikes the region in months, so prediction time span of 1 month is feasible in this study. However, there are certain cases in which more than one major earthquake struck the region in 1 month, but the algorithm considers it only one. The total percentage of the months in which major earthquake occurred is 38 % approximately, while in rest of the months only low-level seismic activity was recorded.

6.2 Performance evaluation

Several performance measures exist for binary classification problems. The earthquake prediction performance is assessed in terms of following measures:

1. True Positives (TP): This indicates the number of times algorithm predicted an earthquake and it actually occurred.
2. True Negative (TN): The number of times algorithm predicted no earthquake and there was no earthquake in actual.
3. False Positive (FP): The number of times algorithm predicted earthquake but no seismic activity happened in actual.
4. False Negative (FN): The number of times algorithm predicted no earthquake but there was an earthquake in actual.

There are some other evaluation criteria derived from the above-mentioned four measures. Two of the most commonly used statistical measures are Sensitivity and Specificity. Sensitivity S_n relates to rate of actual positives predicted while specificity S_p is the rate of actual negatives predicted as shown in Eqs. 12 and 13, respectively.

$$S_n = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$S_p = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{13}$$

Two other evaluation criteria specifically useful for earthquake prediction are positive predicative value P_1 and negative predictive value P_0. P_1 is the percentage of true positives among all the positive predictions of the classifier as shown by Eq. 14, while P_0 is the percentage of true negatives among all the negative predictions of classifier as mentioned in Eq. 15. These two performance measures indicate the rate of false alarms produced by the classifier.

$$P_1 = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{14}$$

$$P_0 = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \tag{15}$$

Similarly, accuracy is another criterion that indicates the percentage of total number of accurate predictions out of all the predictions made by the classifiers, irrespective of positive or negative predictions. It is calculated as follows:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{16}$$

All these evaluation criteria are calculated from four elements of contingency table, i.e., TP, FP, TN and FN. Each criterion highlights the certain aspect of performance of the classifier, so the purpose of employing all these criteria is to mention every aspect of classifier's performance.

6.3 Training of ML techniques

The training of the classifiers introduced in Sect. 5 has been performed by using the dataset described in Sect. 6.1. The training results of every classifier in terms of performance evaluation parameters discussed above are given in Table 2.

6.4 Testing ML techniques on Hindukush data

This section discusses the results yielded by the above discussed ML classifiers over unseen data of 132 months in terms of performance evaluation parameters discussed in Sect. 6.2. The technique predicts an earthquake of magnitude 5.5 and above for the duration of 1 month. Out of 132 months, earthquake ($M \geq 5.5$) struck the region in 68 months, which leads to the hit rate of approximately 52 %.

Table 3 shows the results of earthquake prediction on test dataset. The results produced by these ML techniques are different from each other in terms of discussed evaluation criteria yet quite encouraging over unseen data, considering the fact that there is no such robust system available for earthquake prediction.

For earthquake prediction, false alarms can cause huge panic and financial loss, thus considering the sensitivity of the issue, 71 % of positive predictive value yielded by PRNN is encouraging. But it is showing lesser sensitivity toward the earthquake concurrences. It also takes lead in specificity with the value of 86 %. LP Boost Ensemble classifier is highlighted with the highest accuracy and sensitivity of 65 and 91 %, respectively. But it lacks in specificity S_p and P_1 with 36 and 61 %, respectively. The significant value of sensitivity is an indication that the classifier is highly sensitive toward earthquake occurrences and rarely misses any seismic event.

A trade-off seems to exists between S_n and P_1, which indicates that more sensitive is the classifier toward earthquake hits, more likely it can catch false positives. A similar relation exists between S_p and P_0, with the exception of RF classifier. In this case, rather than declaring a classifier superior to others, it can be said that every classifier is showing useful results in one way or other and can be employed subject to the requirements.

These results obtained for earthquake prediction in Hindukush region is not the ultimate goal, yet this study is a promising step with motivating results, toward the final and robust earthquake prediction system.

Table 2 Training results of ML techniques

Parameter	PRNN	RNN	Random forest	LPBoost ensemble
TP	57	61	39	82
FP	24	26	7	46
TN	186	159	199	162
FN	42	63	64	19
S_n	58 %	49 %	38 %	81 %
S_p	89 %	86 %	96 %	78 %
P_1	70 %	70 %	84 %	64 %
P_0	81 %	71 %	76 %	89 %
Accuracy	79 %	71 %	77 %	79 %

Table 3 Results of earthquake prediction over unseen dataset

Parameter	PRNN	RNN	Random forest	LPBoost ensemble
TP	22	37	46	63
FP	09	17	27	40
TN	55	48	36	23
FN	46	30	23	06
S_n	32 %	55 %	67 %	91 %
S_p	86 %	74 %	57 %	36 %
P_1	71 %	68 %	63 %	61 %
P_0	54 %	62 %	61 %	79 %
Accuracy	58 %	64 %	62 %	65 %

7 Conclusions

Four machine learning techniques have been used to predict earthquakes in the subducting Hindukush region, which is one of the most active seismic regions of the world. Eight mathematically calculated seismicity parameters have been considered to be used as classifiers inputs. Every applied classifier shows slightly different results from each other. Linear Programming Boost ensemble classifier shows better results in terms of sensitivity, while Pattern Recognition Neural Network tends to produce the least false alarms as compared to the other classifiers. The study shows, although earthquake occurrence is supposed to be decidedly nonlinear and appears to be a random phenomenon, yet it can be modeled on the basis of geophysical facts of the seismic region along with highly sophisticated modeling and learning approaches of machine learning.

Acknowledgments The authors would like to thank Centre for Earthquake Studies, for their continuous support and for providing a platform for carrying out this research. Spanish Ministry of Science and Technology, Junta de Andaluca and University Pablo de Olavide under Projects TIN2011-28956-C02, P12-TIC-1728 and APPB813097 are also acknowledged. Finally, authors would also like to thank friends and colleagues for the useful discussion and ideas.

References

Adeli H, Panakkat A (2009) A probabilistic neural network for earthquake magnitude prediction. Neural Netw 22(7):1018–1024

Allen CR (1976) Responsibilities in earthquake prediction: to the seismological society of America, delivered in Edmonton, Alberta, may 12, 1976. Bull Seismol Soc Am 66(6):2069–2074

Båth M (1965) Lateral inhomogeneities of the upper mantle. Tectonophysics 2(6):483–514

Billington S, Isacks BL, Barazangi M (1977) Spatial distribution and focal mechanisms of mantle earthquakes in the Hindu Kush–Pamir region: a contorted Benioff zone. Geology 5(11):699–704

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Boore DM (2001) Comparisons of ground motions from the 1999 Chi-Chi earthquake with empirical predictions largely based on data from California. Bull Seismol Soc Am 91(5):1212–1217

Brehm DJ, Braile LW (1998) Intermediate-term earthquake prediction using precursory events in the New Madrid seismic zone. Bull Seismol Soc Am 88(2):564–580

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chatelain J-L, Roecker SW, Hatzfeld D, Molnar P (1980) Microearthquake seismicity and fault plane solutions in the Hindu Kush region and their tectonic implications. J Geophys Res Solid Earth (1978–2012) 85(B3):1365–1387

Christensen K, Olami Z (1992) Variation of the Gutenberg–Richter b values and nontrivial temporal correlations in a Spring-Block Model for earthquakes. J Geophys Res Solid Earth (1978–2012) 97(B6):8729–8735
Dahmen K, Erta D, Ben-Zion Y (1998) Gutenberg–Richter and characteristic earthquake behavior in simple mean-field models of heterogeneous faults. Phys Rev E 58(2):1494
Ebel JE, Chambers DW, Kafka AL, Baglivo JA (2007) Non-poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California. Seismol Res Lett 78(1):57–65
Ellsworth WL, Matthews MV, Nadeau RM, Nishenko SP, Reasenberg PA, Simpson RW (1999) A physically based earthquake recurrence model for estimation of long-term earthquake probabilities. US Geol Surv Open-File Rept 99(522):23
Farah A, Abbas G, De Jong KA, Lawrence RD (1984) Evolution of the lithosphere in Pakistan. Tectonophysics 105(14):207–227
Ganatra AP, Kosta YP (2010) Comprehensive evolution and evaluation of boosting. Int J Comput Theory Eng 2(6):931
Geller RJ, Jackson DD, Kagan YY, Mulargia F (1997) Enhanced: earthquakes cannot be predicted. Science 275(5306):1616–1620
Grant RA, Raulin JP, Freund FT (2015) Changes in animal activity prior to a major (m = 7) earthquake in the Peruvian Andes. Phys Chem Earth Parts A/B/C 85:69–77
Graves G, Mohamed A-r, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 6645–6649
Hainzl S, Zller G, Kurths J, Zschau J (2000) Seismic quiescence as an indicator for large earthquakes in a system of self-organized criticality. Geophys Res Lett 27(5):597–600
Hamburger MW, Sarewitz DR, Pavlis TL, Popandopulo GA (1992) Structural and seismic evidence for intracontinental subduction in the Peter the First Range, central Asia. Geol Soc Am Bull 104(4):397–408
Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. Math Intell 27(2):83–85
Ikram A, Qamar U (2015) Developing an expert system based on association rules and predicate logic for earthquake prediction. Knowl Based Syst 75:87–103
Kagan YY, Jackson DD, Rong Y (2007) A testable five-year forecast of moderate and large earthquakes in southern California based on smoothed seismicity. Seismol Res Lett 78(1):94–98
Kirschvink JL (2000) Earthquake prediction by animals: evolution and sensory perception. Bull Seismol Soc Am 90(2):312–323
Knopoff L (2000) The magnitude distribution of declustered earthquakes in Southern California. Proc Natl Acad Sci 97(22):11880–11884
Koulakov I, Sobolev SV (2006) A tomographic image of Indian lithosphere break-off beneath the Pamir–Hindukush region. Geophys J Int 164(2):425–440
Larsen PE, Field D, Gilbert JA (2012) Predicting bacterial community assemblages using an artificial neural network approach. Nat Methods 9(6):621–625
Liang Z, Thorpe CE (2000) Stereo- and neural network-based pedestrian detection. IEEE Trans Intell Transp Syst 1(3):148–154
Liu Y, Wang Y, Li Y, Zhang B, Wu G (2004) Earthquake prediction by RBF neural network ensemble. Springer, Berlin
Martin GL (1995) Inventor; computer technology corp., assignee. Pattern recognition neural network. United States patent US 5,440,651
McGuire JJ, Boettcher MS, Jordan TH (2005) Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults. Nature 434(7032):457–461
Mirrashid M (2014) Earthquake magnitude prediction by adaptive neuro-fuzzy inference system (ANFIS) based on fuzzy C-means algorithm. Nat Hazards 74(3):1577–1593
Morales-Esteban A, Martínez-Álvarez F, Reyes J (2013) Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence. Tectonophysics 593:121–134
Murtza I, Abdullah D, Khan A, Arif M, Mirza SM (2015) Cortex-inspired multilayer hierarchy based object detection system using phog descriptors and ensemble classification. V Comput 1–14
Negarestani A, Setayeshi S, Mohammad G-M, Akashe B (2002) Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction. J Environ Radioact 62(3):225–233
Panakkat A, Adeli H (2007) Neural network models for earthquake magnitude prediction using multiple seismicity indicators. Int J Neural Syst 17(01):13–33
Partal T (2015) Comparison of wavelet based hybrid models for daily evapotranspiration estimation using meteorological data. KSCE J Civil Eng 2015:1–9

Pavlis GL, Das S (2000) The Pamir–Hindu Kush seismic zone as a strain marker for flow in the upper mantle. Tectonics 19(1):103–115

Petersen MD, Cao T, Campbell KW, Frankel AD (2007) Time-independent and time-dependent seismic hazard assessment for the State of California: uniform California Earthquake Rupture Forecast Model 1.0. Seismol Res Lett 78(1):99–109

Pulinets S, Ouzounov D (2011) Lithosphere-atmosphere-ionosphere coupling (LAIC) model-an unified concept for earthquake precursors validation. J Asian Earth Sci 41(4):371–382

Reyes J, Morales-Esteban A, Martínez-Álvarez F (2013) Neural networks to predict earthquakes in Chile. Appl Soft Comput 13(2):1314–1328

Rundle JB (1989) Derivation of the complete Gutenberg–Richter magnitude-frequency relation using the principle of scale invariance. J Geophys Res Solid Earth (1978–2012) 94(B9):12337–12342

Schapire RE (1990) The strength of weak learnability. Mach Learn 5(2):197–227

Searle M, Hacker BR, Bilham R (2001) The Hindu Kush seismic zone as a paradigm for the creation of ultrahigh-pressure diamond-and coesite-bearing continental rocks. J Geol 109(2):143–153

Shen Z-K, Jackson DD, Kagan YY (2007) Implications of geodetic strain rate for future earthquakes, with a five-year forecast of m5 earthquakes in southern California. Seismol Res Lett 78(1):116–120

U.S. Geological Survey (2015) Quaternary fault and fold database for the United States

Utsu T, Ogata Y (1995) The centenary of the Omori formula for a decay law of aftershock activity. J Phys Earth 43(1):1–33

Wiemer S, Wyss M (1994) Seismic quiescence before the Landers (M = 7.5) and big bear (M = 6.5) 1992 earthquakes. Bull Seismol Soc Am 84(3):900–916

Zahur A, Majid A, Kausar N (2014) Novel ensemble predictor for gram-positive bacterial protein sequences. In: Proceedings of the IEEE international conference on frontiers of information technology, pp 319–324

Zamani A, Sorbi MR, Safavi AA (2013) Application of neural network and ANFIS model for earthquake occurrence in Iran. Earth Sci Inf 6(2):71–85