Neural Comput & Applic (2017) 28:1043–1055
DOI 10.1007/s00521-015-2121-7

ORIGINAL ARTICLE

Medium–large earthquake magnitude prediction in Tokyo with artificial neural networks

G. Asencio-Cortés[1] · F. Martínez-Álvarez[1] · A. Troncoso[1] · A. Morales-Esteban[2]

Received: 15 July 2015 / Accepted: 4 November 2015 / Published online: 21 November 2015
© The Natural Computing Applications Forum 2015

Abstract This work evaluates artificial neural networks' accuracy when used to predict earthquakes magnitude in Tokyo. Several seismicity indicators have been retrieved from the literature and used as input for the networks. Some of them have been improved and parameterized in order to extract more valuable knowledge from datasets. The experimental set-up includes predictions for five consecutive datasets referring to year 2015, earthquakes with magnitude larger than 5.0 and for a temporal horizon of seven days. Results have been compared to four well-known machine learning algorithms, reporting very promising results in terms of all quality parameters evaluated. The statistical tests applied conclude that differences between the proposed artificial neural network and the other methods are significant.

Keywords Earthquake prediction · Artificial neural networks · Time series

✉ F. Martínez-Álvarez
fmaralv@upo.es

G. Asencio-Cortés
guaasecor@upo.es

A. Troncoso
ali@upo.es

A. Morales-Esteban
ame@us.es

[1] Division of Computer Science, Universidad Pablo de Olavide, Seville, Spain

[2] Department of Building Structures and Geotechnical Engineering, University of Seville, Seville, Spain

1 Introduction

Earthquakes occur, apparently, after no patterns and can cause huge human and material losses. For this reason, the issue of earthquake prediction has been widely addressed by means of many different strategies. Nevertheless, results are not very convincing to date because no developed method can be used worldwide.

Japan emerges as one of the countries with larger seismic activity, with more than 5000 quakes per year, being 1000 of them felt by the population. Destructive earthquakes, often resulting in tsunamis, occur several times a century. The most recent major quakes include the 2011 Tohoku earthquake and tsunami, the 2004 Chuetsu earthquake and the Great Hanshin Earthquake of 1995.

The important seismic and volcanic activity of Japan is due to the tectonic plate movement. It is also responsible for the shape and contents of the Japan archipelago. Fifteen million years ago Japan formed part of Eurasia. The subduction pulled Japan to the east, creating the Sea of Japan. Between 19 and 15 million years ago (Ma) it rotated to occupy its current position [8].

The oldest rocks in Japan are Precambrian and they were part of the several supercontinents before the formation of Rodinia.

Between 750 and 600 Ma, the supercontinent Rodinia was broken by the Pacific superplume. Next, around 450 Ma, the subduction process that formed the Japanese territory began. 300–200 Ma a cold superplume-Laurasia was formed by the addition of different continental blocks. It included the Sino-Korean and the Yangtze cratons. The Japanese landmass was the eastern continental margin of the Yangtze block.

Japan has 15 major geological belts lying NE–SW. These have been formed by accretionary and igneous

processes during the successive subduction processed it has suffered.

Given this sharp geology, it is easy to understand that Japan is, perhaps, the country with the biggest seismic activity across the world. For this reason, the main objective of this study is to evaluate the accuracy of artificial neural networks (ANN) when predicting medium–large earthquakes in Tokyo, and surroundings. The idea underlying this approach is to focus on a very particular location with little spatial resolution. That is, the area under study is a circle with radius 200 km where Tokyo is just in the centre. Additionally, a horizon prediction of seven days is set, thus leading to a reduced temporal resolution as well. Finally, earthquakes with magnitude larger than 5.0 are only considered. Note that this experimental set-up meets the requirements established by Allen [5], whose work claims that an earthquake prediction must provide (a) a specific location or area (Tokyo and 200 km around it), (b) a specific span of time (within the next seven days) and (c) a specific magnitude range ($M > 5.0$).

For this purpose, several seismicity indicators described in previous works [32, 38] have been chosen to be used as input for the designed ANN. Additionally three of these indicators have been modified and improved in this work in order to extract more valuable knowledge from the datasets. Their parametrization is based on a magnitude threshold.

To show the effectiveness of ANN in this context, several well-known and powerful classifiers (nearest neighbours, support vector machines, Bayesian networks and decision trees) have been used to compare its performance. Reported comparisons confirm that the ANN exhibits a very stable and successful behaviour when dealing with seismic time series.

The remainder of this work is structured as follows. Section 2 provides an overview on the latest and most relevant related works found in the literature. A brief description of the Gutenberg-Richter's law can be found in Sect. 3. The methodology proposed is fully described in Sect. 4. Section 5 shows all the results reported from the application of ANN to Tokyo and surroundings. Finally, the conclusions are drawn in Sect. 6.

2 Related work

This section explores the different approaches recently published in the literature related to the earthquakes prediction [7, 31, 42]. Most of methods proposed to address the prediction of earthquakes can be divided into probabilistic approaches, which intend to discover the seismicity distribution, and the artificial intelligence techniques, which obtain a learning model from time series data.

Among the first ones, the method presented in [35] and developed by the U. S. Geological Survey and the California Geological Survey is worthy of mention. Namely, this method assumed that occurrence probability of earthquakes follows a Poisson distribution. Kagan et al. [19] presented a model based on a smoothed seismicity to predict earthquakes of magnitude equal or larger than 5.0 in southern California. Another probabilistic forecast based on strain rate was proposed by Shen et al. in [40]. In particular, the strain rate that is presumably proportional to occurrence rate was geodetically observed and averaged over approximately ten years. In 2014, the estimation of large earthquake event occurrence based on radial basis function neural network models was also applied to data from California [4]. The authors claimed that their model is able to estimate elapsed times between significant seismic events.

A probabilistic neural network to classify earthquakes in California and San Francisco was presented in [1]. The earthquakes were classified in seven different classes with magnitude ranging from 4.0 to 6.0 and results reported showed that two of four earthquakes were predicted correctly. The main contribution of this work is the set of the eight seismic indicators, including the latitude and longitude of the epicentre of the earthquake and the occurrence time of the following earthquake, used as input data of the neural network. Nonetheless, this kind of neural network was not able to predict earthquakes with larger magnitudes. For that, the same authors used a recurrent neural network in [33] to predict time and location of earthquakes with magnitudes equal or greater than 6.0. This work concludes that the accuracy of the prediction improves when increasing the number of days for prediction to 15 days, and as a consequence, the relevance of the size of the training set.

A multilayer perceptron neural network has been proposed to predict earthquakes from total electron content (TEC) time series in [2]. This neural network detects the variations of ionospheric variables considered as precursors of earthquakes. Namely, variations in the TEC time series are an indicator for the occurrence of an earthquake in the next three days.

Nowadays, hybrid methods that highlight most of the strengths of each technique are the most popular work among the researchers. A wavelet transform with a neural network were combined to obtain an estimation of the time elapsed between two consecutive earthquakes in [9]. A combination of neural networks and fuzzy logic is proposed by Zamani [46] to predict earthquakes in Iran. This work incorporates as a first step a data normalization, and given the importance that the seismic indicators play in the accuracy of the prediction, a feature selection based on principal component analysis has also been included. In

fact, the seismic indicators used as input data in neural networks or other methods related to artificial intelligence to predict earthquakes has been widely studied [6, 24, 32]. For instance, the correlation between temporal variations of the Gutenberg–Richter law's b value and the occurrence of earthquakes has been shown in [27, 30, 38, 44].

The main disadvantage existing in neural networks is that local minima can be reached during the training of the network. For this reason, many works focus on the use of bioinspired metaheuristics such as genetic algorithms, particle swarm or artificial bee colony algorithms to solve the underlying optimization problem [3, 22, 39, 45].

In a non-supervised learning framework, different approaches based on association rules [16, 25] or clustering techniques [11, 28] have been successfully applied to predict earthquakes in areas with a high seismic activity from Spain, Portugal and Chile. Also, Iran has been analysed by means of the Fuzzy C-means algorithm [26]. This time, a supervised learning framework, namely regression or classification techniques, has been preferred by the majority of researchers to forecast earthquakes in the last years. For instance, different kind of neural networks have been applied to classify the input data in different classes depending on the magnitude of the earthquake, and also to predict the occurrence time and location of the earthquake [1, 23, 33].

Despite the intense research activity conducted in the last decade, to the best of the authors' knowledge, approaches based on artificial neural networks have not been applied to predict earthquake magnitudes in Japan.

3 Geophysical fundamentals

This section exposes the geophysical fundamentals underlying the method proposed in this text. In particular, it is mainly focused on the description of the Gutenberg–Richter law (GR) since its b value is the basis of many seismicity indicators used as input for the ANN and other classifiers evaluated in the rest of the paper.

Gutenberg and Richter [13] and Ishimoto and Iida [17] observed that the number of earthquakes, N, of magnitude larger than or equal to M follows a power law distribution according to the following equation:

$$N(M) = \alpha M^{-B} \tag{1}$$

where α and B are adjustment parameters.

This equation was transformed into a linear law by Gutenberg and Richter in 1954 [15]:

$$\log_{10} N(M) = a - bM \tag{2}$$

This law relates the cumulative (or absolute) number of events with magnitude larger than or equal to M with the seismic activity, a, and the size-distribution parameter, b. The b value has been observed to reflect the tectonic of the area under analysis [21]. Moreover, it has been related with the physical characteristics of the area. A high b value shows that the relative number of small and large events is dissimilar; therefore, the region has a low resistance. Contrariwise, a low b value implies that the number of earthquakes of large magnitude is often, implying a higher resistance of the ground.

The least squares method was firstly used to estimate the parameters of Eq. (2). It is shown in Eq. 3. In this equation, N_{i-j} is the number of events with magnitude equal or larger than M_{i-j} in the n events prior to e_i.

$$b = \frac{n\sum_{j=1}^{n}(M_{i-j}\log N_{i-j}) - \sum_{j=1}^{n} M_{i-j}\sum_{j=1}^{n}\log N_{i-j}}{\left(\sum_{j=1}^{n} M_{i-j}\right)^2 - n\sum_{j=1}^{n} M_{i-j}^2} \tag{3}$$

Nonetheless, Shi and Bolt [41] pointed out that the presence of even a few large earthquakes has a significant influence over the results. The maximum likelihood method is alternatively proposed. This method produces robust estimates even when the infrequent large earthquakes appear.

It is assumed that the magnitudes of the earthquakes that occur in a zone, in a period of time, are independent and identically distributed variables. These follow the GR law [37]. This hypothesis is equivalent to suppose that the probability density of the magnitude M is exponential:

$$f(M, \beta) = \beta^{-\beta(M-M_0)} \tag{4}$$

where

$$\beta = \frac{b}{\log(e)} \tag{5}$$

and M_0 is the cut-off magnitude.

Utsu [43] used the maximum likelihood method to estimate β:

$$\beta = \frac{1}{(\overline{M} - M_0)} \tag{6}$$

where $\overline{M}$ is the mean magnitude of the earthquakes. The resolution of this equation is shown in next equation:

$$b = \frac{\log e}{(1/n)\sum_{j=0}^{n-1} M_{i-j} - M_0} \tag{7}$$

The parameters involved are: number of events considered prior to e_i, n; magnitude of the event e_{i-j}, M_{i-j}; and the cut-off magnitude of the seismic zone.

As aforementioned, it is important to highlight that the b value must be constant in a seismogenic zone and different to the adjacent ones. Since the area under study (see

Sect. 5.1) is small and compact, this requirement is met by the used data.

4 Methodology

This section describes the proposed methodology to perform earthquake predictions as well as the algorithms used to assess its effectiveness. An artificial neural network-based method, named EarthQuake Predictor based on Artificial Neural Network (EQP-ANN), is presented, turning the earthquake magnitude prediction into a classification problem suitable to establish a machine learning-based comparison.

The remainder of this section is the following. In first place, in Sect. 4.1 the overall scheme to predict earthquakes is described. Next, Sect. 4.2 describes all the seismic features generated to be used for predictions. The artificial neural network inside EQP-ANN is described in Sect. 4.3. Finally, the settings of the other machine learning algorithms put in comparison are shown in Sect. 4.4.

4.1 Overall prediction procedure

The earthquake prediction is carried out into a supervised learning scheme involving a set of well-known classifiers and using *hold-out* as validation method. These approaches have already been used in the literature [24, 27, 32, 38] and consist in predicting the class for a test set, given a training set previously used to generate a knowledge model.

A schematic representation of the proposed methodology is shown in Fig. 1. First, the target zone is selected: the city of Tokyo (Japan) and surroundings are selected for illustrative purposes (red circle). More details about the geographic area and information retrieval are provided in Sect. 5.1.

The information is retrieved from the seismic catalogue by the US Geological Survey (USGS), freely available online, which feeds the system with data of earthquakes occurred in the past. From these data, both training and test sets are built. As Fig. 1 shows, those sets are defined from events e_r to e_s for training and from e_{s+1} to e_t for test. Five test sets are extracted from the catalogue in order to have several equidistant test points to perform comparisons. The five datasets and their extraction details are explained in Sect. 5.1.

For each event in both training and test sets, two groups of seismic indicators are generated (highlighted in blue and green in Fig. 1) and a binary class is allocated (in orange), so that the classifiers can perform the predictions. The first group, coloured in blue, corresponds to features which are derived from the seismic indicator b. The second group, in green, is independent of b. The seismic indicator b corresponds to the Gutenberg–Richter law's b value [14] (see Sect. 3) and its calculation is explained in Sect. 4.2. The definitions of all used seismic indicators have been taken from two previous works [32, 38]. It is worth noting three indicators (T, μ and c) have been modified in this work, including a parametrization based on a magnitude threshold. All seismic indicators characteristics are explained in detail in Sect. 4.2.

The class C for every event is defined as a logical value, $C = \{0, 1\}$. The class C is equals to 1 if the maximum magnitude for events occurred in the prediction horizon (a determined number of days after e_i) is larger than or equal to a predefined threshold named M_s, and it is $C = 0$ otherwise. Formally, the class C_i associated with the event e_i is defined in Eq. (8), where Ω is the number of events encountered in the catalogue within the prediction horizon. The value of Ω varies depending on the event e_i and the specific catalogue. In this work, the prediction horizon has been set to seven days and the magnitude threshold for the class, M_s, has been set to 5.0.

$$C_i = \begin{cases} 1, & \text{if } \max\{\mathrm{M_j}: \ \mathrm{i} < \mathrm{j} \leq \mathrm{i} + \Omega\} \geq \mathrm{M_s} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

To assess the seismic indicators for a given event e_i, the previous n events are calculated. In this work, n has been set to 50 events, as suggested in [29] and successfully used in [24, 27, 38].

Training and test sets are obtained after the generation of the features and class for events e_i, where $i \in [r, t]$. All the attributes have been normalized between 0 and 1. Once both sets are generated, our method and a battery of well-known supervised classifiers have been used in order to predict the class of the current event, which involves predicting if an event with magnitude larger than 5.0 will occur or not during the seven next days.

Finally, differences between predicted and actual classes are assessed. There is a wide range of quality measures and its use may vary from one researcher to another, depending on what it is wanted to be evaluated. Section 5.2 defines the set of measures proposed in this work to evaluate the performance.

4.2 Seismic indicators

The complete set of seismic indicators which are considered for predictions in this work are indicated in Table 1. These attributes were firstly introduced in [38] (x_1, x_2, x_3, x_4, x_5, x_6 and x_7) and [32] (b, a, η, ΔM, T, μ, c, $dE^{1/2}$ and M_{mean}). Two groups of features are defined to separate those dependent on the

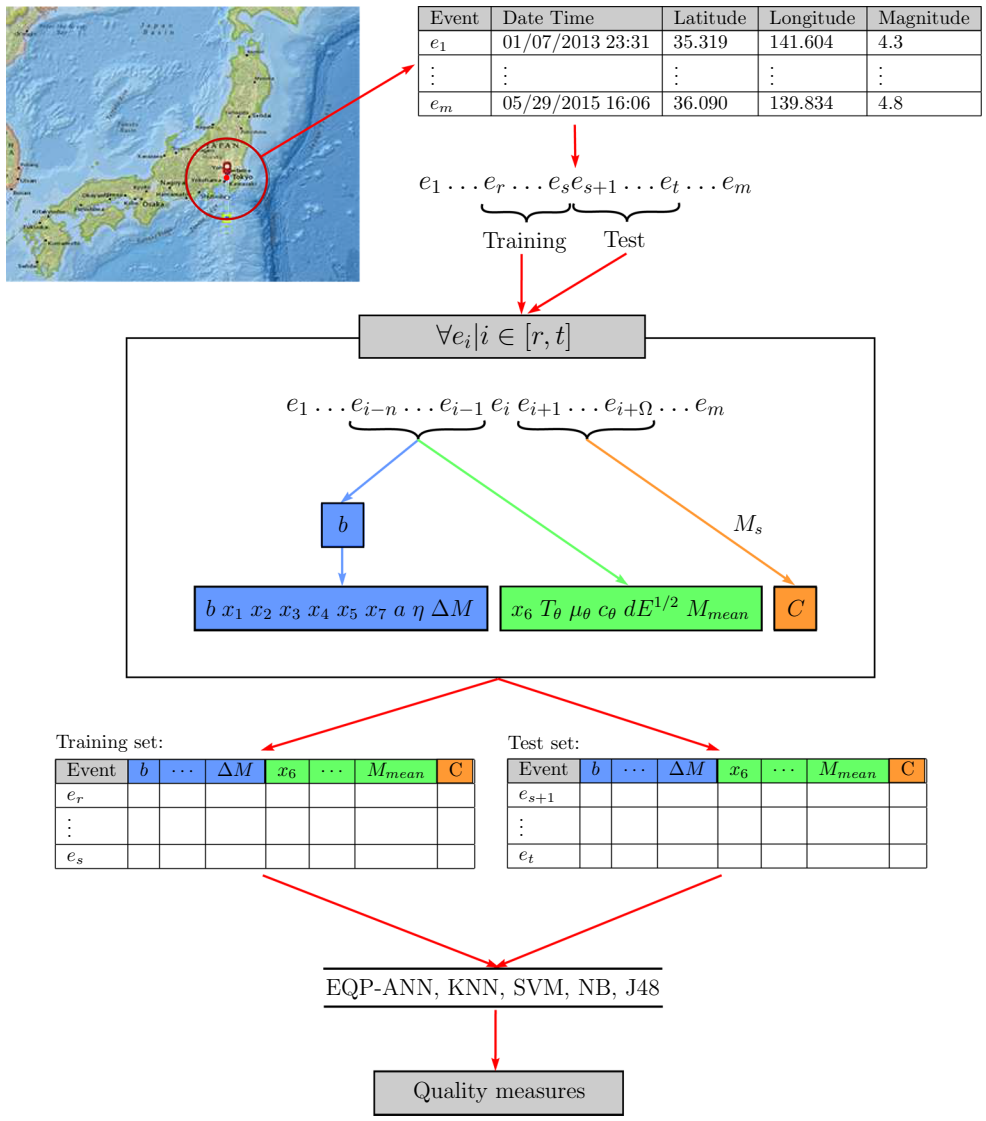

Fig. 1 Scheme of the overall prediction procedure

Gutenberg–Richter law's b value and those independent of such b value.

The group of attributes which are dependent of b is composed of $b, x_1, x_2, x_3, x_4, x_5, x_7, a, \eta$ and ΔM (coloured in blue in Fig. 1 and grouped on top of the Table 1), where b is the Gutenberg–Richter law's b value itself.

The authors in [32] used the least squares method for calculating the b value as shown in Eq. (3). Due to the lack of robustness of this method when large infrequent earthquakes happen (see Sect. 3), [38] used the maximum likelihood method which is described in Eq. (7).

In all previous proposals [24, 27, 32, 38], the attributes x_1, x_2, x_3, x_4, x_5 and x_7 were based on a b calculation based on the Eq. (7), while the attributes b, a, η and ΔM were based on the Eq. (3). In this work, a unique criterion has been used to assess the b value and therefore all b-dependent attributes: the one defined in Eq. (7).

The group of attributes which are independent of b is composed by $x_6, T_\theta, \mu_\theta, c_\theta, dE^{1/2}$ and M_{mean} (coloured in green in Fig. 1 and grouped at the bottom of the Table 1).

The attributes T_θ, μ_θ and c_θ have been modified with respect to their original definitions. A parameter θ, which is a magnitude threshold, is introduced for these attributes.

In order to define the attributes T_θ, μ_θ and c_θ for the event e_i, it is necessary to introduce two new sets H_θ and CH_θ. The set H_θ is composed of the ordered times for the events prior to e_i with magnitude equal or larger than a threshold θ, as shown in Eq. (9), where t_j identifies the date and time of event e_j.

Table 1 Seismic indicators considered for predictions

Name	Description
b	Gutenberg–Richter (GR) law's b value
x_1	Increment of b between the events i and $i-4$
x_2	Increment of b between the events $i-4$ and $i-8$
x_3	Increment of b between the events $i-8$ and $i-12$
x_4	Increment of b between the events $i-12$ and $i-16$
x_5	Increment of b between the events $i-16$ and $i-20$
x_7	Probability of events with magnitude ≥ 6.0 using a probability density function
a	Gutenberg–Richter law's a value
η	Sum of the mean square deviation from the regression line based on GR's law
ΔM	Difference between largest observed magnitude and largest expected based on GR's law
x_6	Maximum magnitude from the events recorded during the last week (OU's law)
T_θ	Elapsed time between the first and nth event with magnitude $\geq\theta$
μ_θ	Mean time among the n last θ-characteristic events
c_θ	Coefficient of variation of the mean time between the n last θ-characteristic events
$dE^{1/2}$	Rate of square root of seismic energy
M_{mean}	Mean magnitude of the last events

$$H_\theta = \{t_j : M_j \geq \theta \wedge j < i\} \tag{9}$$

Let p be the index vector of the set H_θ. Then, T_θ can be defined as the time elapsed between the first and the n-th event, prior to e_i with magnitude equal or larger than θ, as defined in Eq. (10):

$$T_\theta = t_{p(1)} - t_{p(n)} \tag{10}$$

where $p(1)$ and $p(n)$ are the element in the first and the n-th position of the vector, respectively.

Let CH_θ be an ordered set of time for the characteristic events occurred before e_i with magnitude θ and maximum deviation ϕ, as shown in Eq. (11). The value for ϕ has been set to 0.1.

$$CH_\theta = \{t_j : M_j \in [\theta - \phi, \theta + \phi] \wedge j < i\} \tag{11}$$

Let v be the index vector of the set CH_θ. Then, μ_θ is the average elapsed time for n consecutive characteristic events prior to e_i, as it can been observed in Eq. (12):

$$\mu_\theta = \frac{\sum_{j=1}^{n} t_{v(j)} - t_{v(j-1)}}{n} \tag{12}$$

where $v(j)$ is the element in the j-th position of the vector.

Finally, c_θ is the variation coefficient of the elapsed times for the n consecutive characteristic events prior to e_i, as shown in Eq. (13).

$$c_\theta = \frac{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(t_{v(j)} - t_{v(j-1)} - \mu_\theta)^2}}{\mu_\theta} \tag{13}$$

The limits of the parameter θ are defined according to the minimum and maximum magnitudes of the datasets (see Sect. 5.1). For such datasets, θ takes values from 3.6 to 6.2 with a step value of 0.1 (27 values). Therefore, the attributes T_θ, μ_θ and c_θ are vectors in $\mathbb{R}^{27}$.

4.3 Artificial neural network learning machine

The algorithm proposed to perform predictions from the previously described input data is based on an artificial neural network, implemented using a well-known multilayer perceptron that trains using the backpropagation technique. The network has three layers: an input layer; a hidden layer to which all the input nodes are connected; and an output layer.

According to the number of input attributes, there are 94 neurons at the input layer. There is one hidden layer with $(\text{numAttributes} + \text{numClasses})/2 = (94 + 2)/2 = 48$ neurons. Finally, the output layer has two neurons, according to the two possible classes.

One of the most important aspects in a neural network-based algorithm is the learning rate. The learning rate is a multiplicative parameter (between 0 and 1) that determines the step size when searching for a solution and hence how quickly the search converges. If it is too large and the error function has several minima, the search may overshoot and miss a minimum entirely, or it may oscillate. If it is too small, progress towards the minimum may be slow. The learning rate has been set to 0.3.

A limit on the training iterations is introduced for efficiency purposes. This limit sets the number of training epochs, i.e. the number of passes that will be taken through the data. The network stops when the specified number of epochs is reached. This limit has been set to 500 epochs.

The weights in the multilayer perceptron have been learned using the standard gradient descent, but a slight modification has been introduced in order to improve the efficiency. Specifically, a momentum term is included when updating weights. Thus, this term adds to the weight change at the next iteration a small proportion of the weight change from the previous iteration. This smoothes the search process by making changes in the direction less abrupt. This momentum term has been set to 0.2.

Table 2 summarizes the parameters described in this section for the ANN proposed.

Table 2 ANN's architecture and set-up

Parameter	Value
Topology	Multilayer perceptron
Training algorithm	Backpropagation
Number of layers	3
Number of inputs	94
Number of hidden neurons	48
Number of outputs	2
Learning rate	0.3
Number of epochs	500
Momentum term	0.2

4.4 Other machine learning algorithms

The neural network-based algorithm is compared to four classical machine learning approaches: nearest neighbours, support vector machines, Bayesian networks and decision trees.

The nearest neighbours method (KNN) [10] employed is an implementation of the classic k-NN algorithm with $k = 1$, Euclidean distance and a linear search for neighbours without any preponderation of attributes or distances.

The support vector machine method (SVM) used is a fast implementation of support vector machines algorithm using a sequential minimal optimization [20]. The complexity parameter has been set to 1.0 and it uses a polynomial kernel of exponent 1.

The Bayesian networks method (NB) is a simple Naive Bayes implementation for classification [18]. Any special parametrization is not required.

The decision trees method (C4.5) used is an implementation of the classic C4.5 algorithm [36] for decision tree-based classification. The confidence factor for pruning has been set to 0.25 and it uses a minimum of 2 instances per leaf.

5 Results

This section gathers the results for this work. First, Sect. 5.1 describes the datasets and their retrieval procedure. Then, the quality parameters used to assess the performance of the proposed methodology are introduced in Sect. 5.2. Later in Sect. 5.3, the results themselves are reported and commented. Finally, a statistical analysis is carried out in Sect. 5.4 to prove that results between EQP-ANN and other methods are statistically significant.

5.1 Datasets description

The city of Tokyo (Japan) and its surroundings has been selected as the geographic region to test the proposed method. Figure 2 shows a map of Japan with the region of study surrounded by a circumference. The coordinates of the centre are 35°41′02″N, 139°46′28″E and the radius has been set to 200 km.

The catalogue of earthquake events has been retrieved from the NSGS website on 29 May 2015. The last 400 events up to that date have been extracted in order to take five datasets to perform predictions with EQP-ANN and comparisons over other approaches. Table 3 shows the main characteristics of each dataset extracted. The processed data, with all new attributes, can be found at http://eps.upo.es/martinez/data.rar.

The first column of Table 3 provides an identifier for each dataset. The second to fifth columns indicate the dates of beginning and end of both training and test sets, respectively (in mm/dd/yy format). Note that dates of tests are consecutive from DS1 to DS5. The columns #TR and #TE show the number of events in both training and test sets, respectively. The ratios between training and test are 75.0, 74.7, 74.7, 74.8 and 74.8 % for datasets DS1 to DS5, respectively.

The columns TR+/− and TE+/− indicate the number of positive events and negative ones in both training and test sets, respectively. Positive (and negative) criteria is defined in Eq. (8) using $M_s = 5.0$. Finally, columns Mean and SD show the average and standard deviation of magnitudes of events putting both training and test sets together. The minimum magnitude or cut-off magnitude for the seismic zone (M_0), used to retrieve the catalogue, is 3.0. With such filter, the minimum magnitude present in the 400 events is 3.6. The maximum magnitude encountered in those events is 6.2.

Note that the magnitude threshold for predictions ($M_s = 5.0$) is higher than Mean + SD for all datasets, which implies predictions of atypical (and therefore most relevant) high-magnitude earthquakes. Moreover, Fig. 3 shows the distribution of magnitudes for each dataset of

Fig. 2 The region of study: Tokyo (Japan) (35°41′02″N, 139°46′28″E) with a radius of 200 km

Table 3 Characteristics of the five extracted datasets to perform comparisons.

DS	TR begin	TR end	TE begin	TE end	#TR	#TE	TR+/−	TE+/−	Mean	SD
DS1	01/07/13	12/08/13	12/09/13	02/18/14	144	48	20/124	17/31	4.59	0.35
DS2	04/17/13	02/18/14	02/19/14	06/19/14	142	48	20/122	7/41	4.58	0.35
DS3	07/31/13	06/19/14	06/20/14	09/28/14	148	50	18/130	7/43	4.57	0.32
DS4	11/28/13	09/28/14	09/29/14	01/03/15	152	51	15/137	4/47	4.54	0.31
DS5	02/05/14	01/03/15	01/04/15	05/29/15	158	53	10/148	2/51	4.50	0.31

study. As it can be noticed, the magnitudes follow the GR exponential law, as it is mentioned in Sect. 3.

5.2 Quality measures

Several measures are used in this section to assess the performance of the proposed method and compare it to other methods. Hence, the four different possible situations for classifiers are defined by means of the confusion matrix, shown in Table 4. For this particular case, TP is the number of times that an upcoming earthquake has been properly predicted. TN is the number of times that neither a classifier triggered an alarm nor an earthquake has occurred. FP the number of times that a classifier has erroneously predicted. And FN the number of times that a classifier has not trigger an alarm but an earthquake did occur.

The combination of these parameters leads to the calculation of some well-known quality parameters such as predictive positive value (PPV), negative predictive value (NPV), sensitivity (S_n) and specificity (S_p). Furthermore, in order to provide overall quality measures, the average of these four measures is introduced, $\mathrm{Avg}(\mathrm{PPV}, \mathrm{NPV}, S_n, S_p)$, as well as the Matthews Correlation Coefficient (MCC), since it often provides a much more balanced evaluation of the prediction than averaged percentages. It is basically a correlation coefficient between the observed and predicted binary classifications. Values range from -1 to $+1$, where $+1$ means a perfect prediction, 0 no better than random prediction and -1 total disagreement between prediction and observation. All formulas are listed below:

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

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

$$S_n = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{16}$$

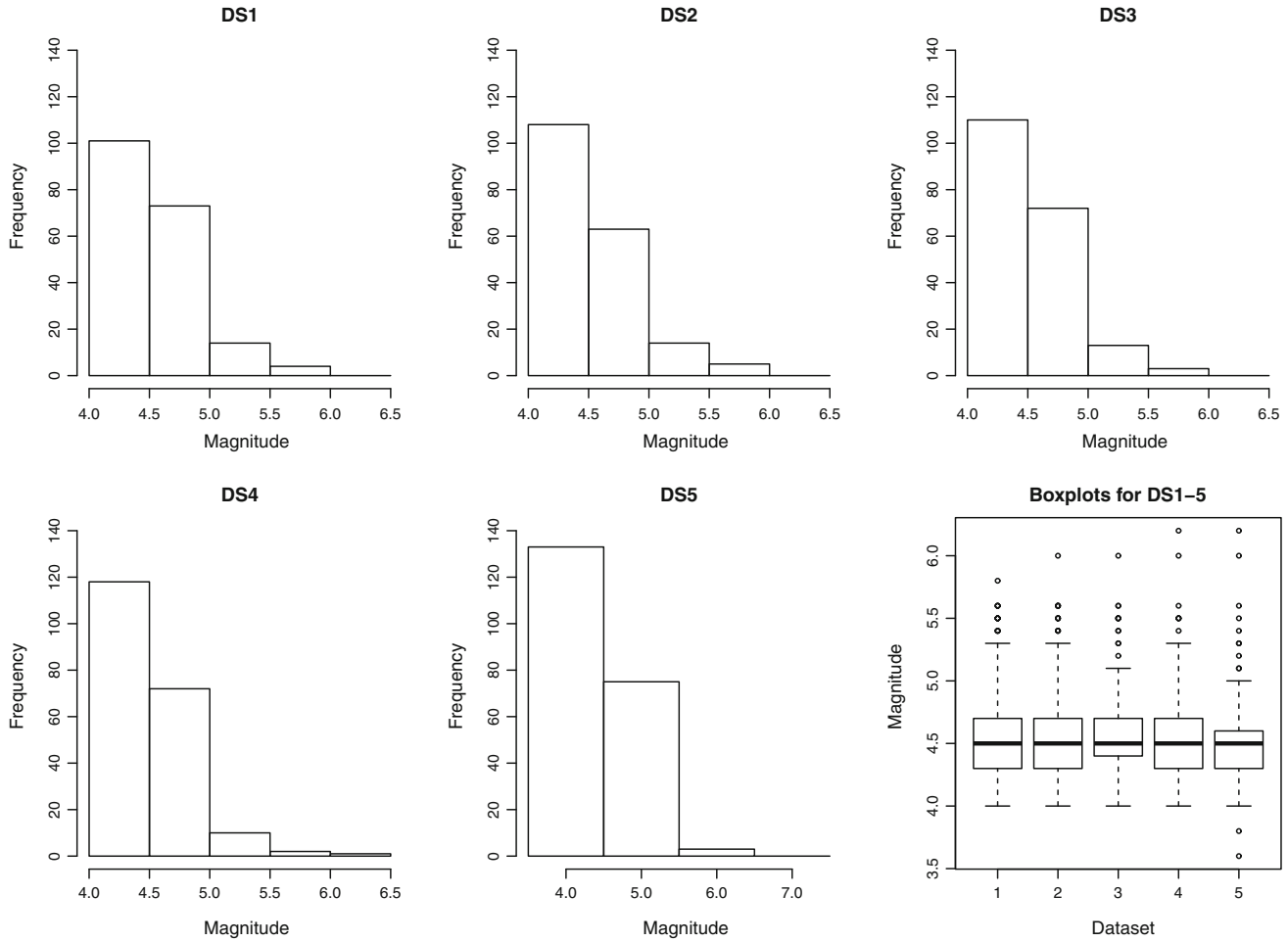

Fig. 3 Distribution of magnitudes in the datasets of study

Table 4 Confusion matrix for classifiers

	Satisfies null hypothesis	Does not satisfy null hypothesis
Test negative	True negative (TN)	False negative (FN)
Test positive	False positive (FP)	True positive (TP)

$$S_{\mathrm{p}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{17}$$

$$\mathrm{Avg}(\mathrm{PPV}, \mathrm{NPV}, S_{\mathrm{n}}, S_{\mathrm{p}}) = 0.25(\mathrm{PPV} + \mathrm{NPV} + S_{\mathrm{n}} + S_{\mathrm{p}}) \tag{18}$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{19}$$

5.3 Experimentation results

This section provides the experimental results for the five consecutive datasets analysed in this work. The results for every dataset are discussed separately. For this purpose, Tables 5, 6, 7, 8 and 9 refer to the results for DS1 to DS5, respectively.

With reference to DS1 (see Table 5), it is worth noting that there is no FP in the predictions by EQP-ANN, in contrast to the remaining methods. This fact leads to a 100 % PPV and 100 % specificity. In other words, EQP-ANN covers all occurrences for the negative class. By contrast, the sensitivity appears in third position, that is it covers less occurrences for the positive class. However, when EQP-ANN predicts a positive case and therefore an earthquake is expected to occur, it does occur. Finally, the MCC index is the highest among the five classifiers, followed by KNN. There is a very significant difference in terms of accuracy with SVM, NB or J48, whose results are not competitive for this dataset.

Table 5 Comparison of effectiveness between EQP-ANN and other approaches when predicting the dataset DS1

	EQP-ANN	KNN	SVM	NB	J48
TP	3	10	2	0	7
TN	31	22	25	20	10
FP	0	9	6	11	21
FN	14	7	15	17	10
PPV	1.00	0.53	0.25	0.00	0.25
NPV	0.69	0.76	0.63	0.54	0.50
Sn	0.18	0.59	0.12	0.00	0.41
Sp	1.00	0.71	0.81	0.65	0.32
Avg	0.72	0.65	0.45	0.30	0.37
MCC	0.35	0.29	−0.10	−0.40	−0.26

Table 6 Comparison of effectiveness between EQP-ANN and other approaches when predicting the dataset DS2

	EQP-ANN	KNN	SVM	NB	J48
TP	5	1	3	1	0
TN	35	34	28	20	28
FP	6	7	13	21	13
FN	2	6	4	6	7
PPV	0.45	0.13	0.19	0.05	0.00
NPV	0.95	0.85	0.88	0.77	0.80
Sn	0.71	0.14	0.43	0.14	0.00
Sp	0.85	0.83	0.68	0.49	0.68
Avg	0.74	0.49	0.54	0.36	0.37
MCC	0.48	−0.03	0.08	−0.26	−0.25

Table 7 Comparison of effectiveness between EQP-ANN and other approaches when predicting the dataset DS3

	EQP-ANN	KNN	SVM	NB	J48
TP	1	6	0	0	1
TN	43	26	43	35	33
FP	0	17	0	8	10
FN	6	1	7	7	6
PPV	1.00	0.26	0.00	0.00	0.09
NPV	0.88	0.96	0.86	0.83	0.85
Sn	0.14	0.86	0.00	0.00	0.14
Sp	1.00	0.60	1.00	0.81	0.77
Avg	0.76	0.67	0.47	0.41	0.46
MCC	0.35	0.32	−1.00	−0.18	−0.08

Table 6 summarizes the results for DS2. Again, EQP-ANN is the method with less FP, which is a very desirable property in predictive seismic systems [24]. Additionally, it has also generated the smaller number of FN. In short, EQP-ANN has outperformed all other methods for all the parameters evaluated. The difference in terms of the MCC index between the EQP-ANN and the other methods is especially remarkable, producing a difference greater than 0.4 units with the second best classifier (SVM).

Particular results for DS3 are reported in Table 7. Although EQP-ANN offers better results in terms of Avg and MCC, KNN discovered many more TP (6 vs. 1) but, to get this done, it threw many more FP (17 vs. 0). These values generated, for KNN, a 86 % sensitivity but with only 26 % PPV. In this dataset, SVM has not thrown any FP but it has not discovered any TP either, in other words, it classified all samples in the negative class and, therefore, its results cannot be taken into consideration.

One of the most remarkable features of DS4 is its high degree of imbalance, as it is shown in Table 3. This has caused very low values of PPV for all the methods except for EQP-ANN, which properly predicted 1 out of 4 positive cases, as well as KNN. Nonetheless, KNN generated 10 FP contrarily to EQP-ANN, which generated no FP. Again, the

Table 8 Comparison of effectiveness between EQP-ANN and other approaches when predicting the dataset DS4

	EQP-ANN	KNN	SVM	NB	J48
TP	1	1	0	0	0
TN	47	37	47	46	23
FP	0	10	0	1	24
FN	3	3	4	4	4
PPV	1.00	0.09	0.00	0.00	0.00
NPV	0.94	0.93	0.92	0.92	0.85
Sn	0.25	0.25	0.00	0.00	0.00
Sp	1.00	0.79	1.00	0.98	0.49
Avg	0.80	0.51	0.48	0.47	0.34
MCC	0.48	0.02	−1.00	−0.04	−0.28

Table 9 Comparison of effectiveness between EQP-ANN and other approaches when predicting the dataset DS5

	EQP-ANN	KNN	SVM	NB	J48
TP	2	0	2	0	0
TN	46	51	30	50	51
FP	5	0	21	1	0
FN	0	2	0	2	2
PPV	0.29	0.00	0.09	0.00	0.00
NPV	1.00	0.96	1.00	0.96	0.96
Sn	1.00	0.00	1.00	0.00	0.00
Sp	0.90	1.00	0.59	0.98	1.00
Avg	0.80	0.49	0.67	0.49	0.49
MCC	0.51	−1.00	0.23	−0.03	−1.00

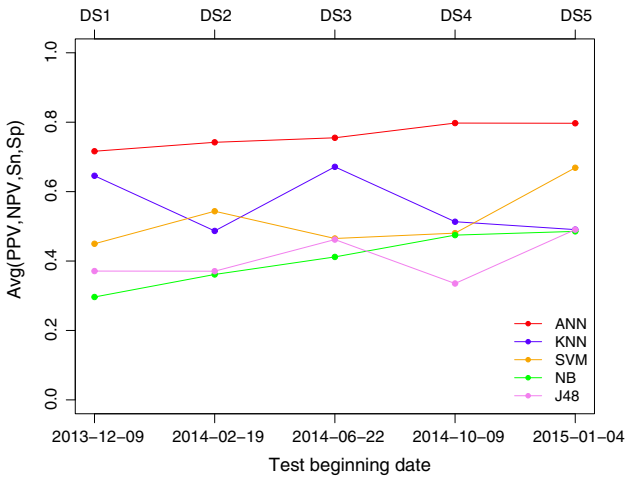

Fig. 4 Performance comparison between EQP-ANN and KNN, SVM, NB, J48 for each dataset

value for the MCC index has been largely better for EQP-ANN, thus outperforming in 0.46 units the second best value.

Finally, Table 9 shows the results for DS5. This dataset only contains two positive cases versus 51 negatives, being the most imbalanced among all analysed datasets. This fact turns the prediction into a very difficult task. It can be noticed that only EQP-ANN and SVM discovered earthquakes and, in fact, both properly predicted the two TP cases. Nevertheless, SVM generated the largest value for FP, namely a value of 21.

From a joint analysis of the five tables, a conclusion can be easily drawn. Since MCC and Avg values for EQP-ANN are the best ones for all of them, it can be concluded that EQP-ANN is the most suitable classifier for these datasets. To prove this statement, statistical tests are carried out in the next section. Generally speaking, results are satisfactory in average terms. However, there is much work to be done since too many FP in almost all predictions. The big challenge is to increase the sensitivity (TP) without increasing FP.

Figure 4 illustrates the average value (Avg) of PPV, NPV, Sn and Sp for the five datasets. It can be observed that EQP-ANN reached better results in terms of this parameter for all the five datasets. Moreover, results related to EQP-ANN exhibited a slightly upward trend in their values, even if a high degree of imbalance is encountered in the datasets over time. The second best algorithm is KNN but with high variability in the results, turning this method into a one not as stable as desired and, maybe, not the one to be evaluated for future works. SVM obtained stable results over time, but with values around 50 %, except for the last dataset in which it reached almost 70 %. Finally, the methods with worse results, with an average accuracy lower than 50 %, have been J48 and NB.

5.4 Statistical analysis

This section is devoted to determine if EQP-ANN has significant statistical difference regarding the other evaluated methods. For this purpose, the Kruskal Wallis and Friedman tests have been performed, as well as a post hoc test using Mann–Whitney tests with Bonferroni correction in order to determine a pairwise analysis for EQP-ANN with respect to the other methods [12]. Such analyses have been carried out by using the free online tool STATService [34].

The Kruskal Wallis test revealed significant differences among methods in comparison over the five datasets of study ($\chi^2 = 17.899, p$ value $= 0.0012$, therefore $p < 0.01$). Moreover, the Friedman test also revealed the same result ($\chi^2 = 17.333, p$ value $= 0.0016, p < 0.01$).

The post hoc test using Mann–Whitney tests with Bonferroni correction showed significant differences between EQP-ANN and KNN (p value $= 0.0079, p < 0.05, r = 0.8307$), between EQP-ANN and SVM (p value $= 0.0079, p < 0.01, r = 0.8282$), between EQP-ANN and NB (p value $= 0.0079, p < 0.01, r = 0.8282$) and between EQP-ANN and J48 (p value $= 0.0079, p < 0.01, r = 0.8307$).

Therefore, this study concludes that differences between EQP-ANN and the other methods are statistically significant.

6 Conclusions

The prediction of earthquakes of medium–large magnitude for the city of Tokyo and surroundings has been firstly addressed by means of artificial neural networks in this work. The results achieved can be considered as satisfactory since it has reached values greater than 70 % for all the five consecutive datasets considered. To evaluate its effectiveness, it has been compared to other well-known and powerful methods. It outperformed all of them in terms of the quality parameters typically used to assess this kind of systems. When medium–large magnitudes are intended to be predicted, a high degree of imbalance can be seen in datasets. In this sense, future work should be directed towards the use of classification techniques specially designed to predict imbalanced datasets.

Acknowledgments The authors would like to thank Spanish Ministry of Economy and Competitiveness, Junta de Andalucía and Pablo de Olavide University for the support under projects TIN2014-55894-C2-2-R, P12-TIC-1728 and APPB813097, respectively.

References

1. Adeli H, Panakkat A (2009) A probabilistic neural network for earthquake magnitude prediction. Neural Networks 22:1018–1024
2. Akhoondzadeh M (2013) A MLP neural network as an investigator of TEC time series to detect seismo-ionospheric anomalies. Adv Space Res 51(11):2048–2057
3. Alavi AH, Gandomi AH (2011) Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. Comput Struct 89:2176–2194
4. Alexandridis A, Chondrodima E, Efthimiou E, Papadakis G, Vallianatos F, Triantis D (2014) Large earthquake occurrence estimation based on radial basis function neural networks. IEEE Trans Geosci Remote Sens 52(9):5443–5453
5. Allen CR (1982) Responsibilities in earthquake prediction. Bull Seismol Soc Am 66:2069–2074
6. Asencio-Cortés G, Martínez-Álvarez F, Morales-Esteban A, Reyes J, Troncoso A (2015) Improving earthquake prediction with principal component analysis: application to Chile. Lect Notes Artif Intell 9121:393–404
7. Azam F, Sharif M, Yasmin M, Mohsin S (2014) Artificial intelligence based techniques for earthquake prediction: a review. Sci Int 26(4):1495–1502
8. Barnes GL (2003) Origins of the Japanese islands: the new big picture. Jpn Rev 15:3–50
9. Chen YL, Cheng Y, Huang Q (2012) Research on the magnitude time series prediction based on wavelet neural network. Appl Mech Mater 128:233–236
10. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
11. Florido E, Martínez-Álvarez F, Morales-Esteban A, Reyes J, Aznarte JL (2015) Detecting precursory patterns to enhance earthquake prediction in Chile. Comput Geosci 76:112–120
12. García S, Fernández A, Luengo J, Herrera F (2009) A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Comput 13(10):959–977
13. Gutenberg B, Richter CF (1942) Earthquake magnitude, intensity, energy and acceleration. Bull Seismol Soc Am 32(3):163–191
14. Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. Bull Seismol Soc Am 34:185–188
15. Gutenberg B, Richter CF (1954) Seismicity of the Earth. Princeton University, Princeton
16. 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
17. Ishimoto M, Iida K (1939) Observations sur les seismes enregistrés par le microsismographe construit dernièrement. Bull Earthq Res Inst 17:443–478
18. John GH, Langley P (1995) Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the eleventh conference on uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc, pp 338–345
19. 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
20. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK (2001) Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput 13(3):637–649
21. Lee K, Yang W-S (2006) Historical seismicity of Korea. Bull Seismol Soc Am 71(3):846–855
22. Li J, Kang F, Ma Z (2013) An artificial bee colony algorithm for locating the critical slip surface in slope stability analysis. Eng Opt 45(2):207–223
23. Madahizadeh R, Allamehzadeh M (2009) Prediction of aftershocks distribution using artificial neural networks and its application on the May 12, 2008 Sichuan earthquake. J Seismol Earthq Eng 11(3):111–120
24. Martínez-Álvarez F, Reyes J, Morales-Esteban A, Rubio-Escudero C (2013) Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula. Knowl Based Syst 50:198–210
25. Martínez-Álvarez F, Troncoso A, Morales-Esteban A, Riquelme JC (2011) Computational intelligence techniques for predicting earthquakes. Lect Notes Artif Intell 6679(2):287–294
26. 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
27. 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
28. Morales-Esteban A, Martínez-Álvarez F, Troncoso A, de Justo JL, Rubio-Escudero C (2010) Pattern recognition to forecast seismic time series. Expert Syst Appl 37(12):8333–8342
29. Nuannin P (2006) The potential of b-value variations as earthquake precursors for small and large events. Technical Report 183, Uppsala University, Sweden
30. Nuannin P, Kulhanek O, Persson L (2005) Spatial and temporal b value anomalies preceding the devastating off coast of nw sumatra earthquake of December 26, 2004. Geophys Res Lett 32:L11307
31. Otari GV, Kulkarni RV (2012) A review of application of data mining in earthquake prediction. Int J Comput Sci Inf Technol 3(2):3570–3574
32. Panakkat A, Adeli H (2007) Neural network models for earthquake magnitude prediction using multiple seismicity indicators. Int J Neural Syst 17(1):13–33
33. Panakkat A, Adeli H (2009) Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators. Comput Aided Civil Infrastruct Eng 24:280–292
34. Parejo JA, García J, Ruiz-Cortés A, Riquelme JC (2012) STAT service: Herramienta de análisis estadístico como soporte para la investigación con Metaheurísticas. In: Actas del VIII Congreso Expañol sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados
35. 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
36. Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco
37. Ranalli G (1969) A statistical study of aftershock sequences. Annali di Geofisica 22:359–397
38. Reyes J, Morales-Esteban A, Martínez-Álvarez F (2013) Neural networks to predict earthquakes in Chile. Appl Soft Comput 13(2):1314–1328
39. Shah FM, Gazhali R, Nawi NM (2011) Using artificial bee colony algorithm for MLP training on earthquakes time series data prediction. J Comput 3(6):135–142
40. Shen ZZ, 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
41. Shi Y, Bolt BA (1982) The standard error of the magnitude-frequency b value. Bull Seismol Soc Am 72(5):1677–1687
42. Tiampo KF, Shcherbakov R (2012) Seismicity-based earthquake forecasting techniques: ten years of progress. Tectonophysics 522–523:89–121

43. Utsu T (1965) A method for determining the value of b in a formula $logn = a - bm$ showing the magnitude-frequency relation for earthquakes. Geophys Bull Hokkaido Univ 13:99–103
44. Wiemer S, Schorlemmer D (2007) ALM: an asperity-based likelihood model for California. Seismol Res Lett 78(1):134–143
45. Yadav A, Deep K, Kumar S (2012) Improving local and regional earthquake locations using and advance inversion technique. World J Modell Simul 8:135–141
46. 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