# LSTM-based earthquake prediction: enhanced time feature and data representation

Asmae Berhich*, Fatima-Zahra Belouadha and Mohammed Issam Kabbaj

AMIPS Research Team,
E3S Research Center,
Ecole Mohammadia d'Ingénieurs,
Mohammed V University in Rabat,
10090, Morocco
Email: berhich.asmae@gmail.com
Email: belouadha@emi.ac.ma
Email: kabbaj@emi.ac.ma
*Corresponding author

**Abstract:** In the last decades, many studies have emerged with different approaches in the field of earthquake prediction. Statistical and machine learning approaches have been used. However, these contributions remain immature. Some of them have not led to a successful prediction. Others have not been able to predict earthquakes so efficiently. Consequently, research into more relevant methods appropriates to this field will be important, as it would improve accuracy, performance, and dynamicity. This paper suggests applying the well-known deep learning algorithm long short-term memory to predict earthquakes in Moroccan regions. The features used in the prediction takes the most influencing and correlated datasets, it calculates an appropriate time feature that is simpler and more precise. The optimal hyperparameters values of our models are retrieved by the grid search technique. The performance of our model is compared with deep neural networks. The final results demonstrate that our model is more effective.

**Keywords:** earthquake prediction; deep learning; neural networks; LSTM; long short-term memory; seismic dataset; recurrent neural networks; magnitude prediction; hyperparameters optimisation; grid search.

**Reference** to this paper should be made as follows: Berhich, A., Belouadha, F-Z. and Kabbaj, M.I. (2021) 'LSTM-based earthquake prediction: enhanced time feature and data representation', *Int. J. High Performance Systems Architecture*, Vol. 10, No. 1, pp.1–11.

**Biographical notes:** Asmae Berhich is a PhD student at the University Mohammed V in Rabat, Morocco. She received a Master's degree in Engineering Information Systems from Cadi Ayad University in Marrakech. Her current field of research is with the earthquake prediction using machine learning algorithms. She is interested in artificial neural networks and deep learning.

Fatima-Zahra Belouadha received her PhD in Computer Science from the Mohammadia School of Engineers (Ecol Mohammadia d'Ingénieurs – EMI), Mohamed V University in Rabat, Morocco. She is a Professor at the Computer Science Department of EMI since 1996. She was a member of EMI and Mohammed V University in Rabat Councils, and she is currently a Member of the AMIPS Research Team, EMI, University Mohammed V in Rabat, Morocco. Active in the research area, she has published many papers on semantic Web service composition and composability. Her current research interests are focused on business process management, semantic data mining, big data and machine learning.

Mohammed Issam Kabbaj receives his Doctorate on Mathematics and Computer Science in Toulouse, France. He is Professor of Higher Education at Mohammadia School of Engineers (EMI) in Morocco since 2010.

This paper is a revised and expanded version of a paper entitled 'Machine learning based earthquake prediction' presented at *International Conference on Modern Intelligent Systems Concepts*, Mohammed V University in Rabat, Faculty of sciences, Morocco, 12–13 December, 2018.

## 1 Introduction

Natural disasters are sudden events, resulting from the natural process of the earth like floods, hurricanes, tornadoes, volcanic eruptions, earthquakes, tsunamis, and other processes. Earthquakes may be the cause of the most devastating damages ever seen on earth; they can destroy a city or even spread their damage over a whole region within seconds causing loss of lives and properties, not to mention their social, economic and environmental impact (Špičák and Vaněk, 2016). To mitigate the underlying risks, scientists have tested multiple approaches applied to highest seismic regions in the world such as Japan, California, Greece, China, etc., e.g., in 1975 scientists have successfully forecasted a strong earthquake, in Haicheng, China, using geoelectrical measurements. As mentioned in Wang et al. (2006), the prediction in Haicheng was a blend of confusion, but it was the first successful prediction. However, one year later, the scientists failed to predict the Tangshan earthquake, and the consequences were dramatic.

Earthquake prediction is certainly a worthwhile matter. If it gets as efficient as we hope, it can help to preserve thousands of lives and saving great amounts of money necessary for reconstruction. Accordingly, it is clearly recommended to apply the most effective models and tools to solve this problem. Therefore, for this purpose, seismic historical data are used, in order to extract their nature and characteristics. Besides, seismologists have used multiples approaches, for instance geophysical, mathematical, statistical and computational models. However, not all of their predictions have had really an accurate result. Sometimes it was due to the complex calculations, and sometimes it was because of failing to analyse the huge recorded seismic data. Other recent research contributions have adopted machine learning and deep learning techniques to predict an earthquake. However, earthquake prediction that focuses on empirical analysis due to the nature of the problem it addresses, has not yet led to a successful prediction of an earthquake and it hence, remains an immature science (Hayakawa, 2015). Failing in predicting correct results can lead to a dilemma. On one hand, false alarms with no quakes taking place may drive the society to a panic and economic disruption, on the other hand, not giving a warning at the appropriate time, to a major earthquake that does occur, might guide to great losses of lives and economic destructions. Consequently, research into more relevant methods appropriates and adapted to this field will be important as it would improve accuracy, performance and dynamicity in terms of real-time processing.

Artificial neural networks and deep learning are widely applied to hard-learned and complex datasets. Hence, many works tried artificial neural networks and deep learning because of their randomness nature and hyper parametrisation quality to solve the earthquake prediction. Not all of the works were typically compelling especially in terms of predicting all the important characteristics of the future earthquakes.

Our work will fill the full meaning of the spatiotemporal earthquakes prediction by giving the four parameters as outputs: Magnitude, location, time, and accuracy of the prediction. The historical data we use belongs to the Moroccan regions given by the geophysical institute of CNRST.

In this research paper, we build a model based on LSTM a deep learning algorithm, and a useful method for time series analysis and sequence data. The dataset representation used in the prediction is enhanced by calculating a simple feature that presents the date and time of events perfectly, especially in a very simple way. Our model gives performant and effective results because we select the best hyper parametrisation for it, using the grid search technique. And, to evaluate and compare the performance of our model we apply the deep neural networks (DNN).

The remainder of this paper is organised in four sections. Section 2 gives an overview of earthquake prediction and classifies the techniques used in this field. Section 3 presents the dataset that is typically used in literature and it describes the architecture of the LSTM algorithm. Section 4 explains the usefulness of our data and model representation. Section 5 is a comparative synthesis that discusses the relevance and performance of our approach. Finally, Section 6 concludes the objective of our work.

## 2 Earthquake prediction: overview and techniques classification

Earthquake prediction, precisely the spatiotemporal magnitude prediction, is a branch of the seismology science. The aim of this branch is to help the authorities to focus their efforts for reducing the socioeconomic damages and losses of seismic events that would occur in the future. In practice, its purpose is to provide four important elements (Allen, 1976):

- a specific magnitude range
- a specific span of time
- a specific location or area
- a specific probability of occurrence that determines the performance of prediction.

Earthquake prediction constitutes in fact, a sensible research field because it has a social and economic impact. Bad predictions may lead to huge damages and fatal injuries. For this reason, many research contributions have been performed trying to find more accurate results. These contributions have various specific objectives in terms of the time range of the performed prediction and the use context which refers to the magnitude of the predicted earthquake. According to the time range prediction, three types of prediction can be distinguished: long term prediction (10 to 100 years time scale), intermediate-term prediction (1 to 10 years time scale), and short-term prediction (up to one-year time scale). We note that five-years seismicity models were elaborated for a magnitude greater than or equal to 5.0 (Kagan et al., 2007; Ebel et al., 2007). Similarly, a smoothed-seismicity model constructed on small earthquakes for mapping large ones was proposed in Helmstetter et al. (2006). Besides, two one-day forecast methods for earthquakes larger than or equal to 4.0 were presented in Ebel et al. (2007). In Rhoades (2007), they propose a method for long-range forecasting based on preceding minor earthquakes for forecasting large events.

As mentioned above, many research contributions have been proposed in the earthquake prediction field to improve the quality of results. The proposed approaches experiment different techniques trying to achieve this objective and they can be classified into two main categories: probabilistic and statistical models-based approaches and machine learning based approaches.

The probabilistic and statistical models-based approaches perceive the earthquake generation as a stochastic process and hence, are founded on analysing the seismicity distribution and using statistical methods. Many works from this category are proposed by the regional earthquake likelihood model (RELM) working group formed in 2000 to refine the seismic hazard modelling in California and highlight how earthquake occurrence can be physically and statistically characterised. A time-independent model considering that the probability of the earthquake occurrence follows a Poisson distribution was proposed in Petersen et al. (2007). A probabilistic method using foreshock/aftershock statistics for 24-hours forecast was presented in Gerstenberger et al. (2007). Another probabilistic model was established in Shen et al. (2007) as intermediate to long-time forecast. To participate to RELM program, authors in Bird et al. (2007) proposed simple methods for estimating long-term average seismicity of any region, based on a local kinematic model of surface velocities and an existing global calibration of plate-boundary seismicity. In addition, to match with the vision of RELM project, testable earthquake potential maps based on geodesy, geology, historical seismicity and computer simulations of earthquakes were proposed in Ward (2007). A stochastic model called epidemic rate strain (ERS) model, was also proposed in Console et al. (2007). This model allows the computation of the likelihood of a seismic catalogue and reflects at least to some extent the physics of earthquake processes. It merges the classical concept of a purely stochastic model called epidemic type aftershock sequence (ETAS) and the RateState theory for the seismicity rate. Many other contributions (Baykara et al., 2005; Yalım et al., 2007; Erees et al., 2007; Zmazek et al., 2003; Şen, 1998; Şen and Al-Suba'i, 2001) are founded on mathematical models and statistical calculations, such as regression calculations for analysing the characteristics and the risk of the earthquake's occurrences.

On the other hand, the machine learning-based approaches use machine learning methods as recently prominent techniques for automatic pattern recognition from the time series data. They experiment with various learning methods based on learning from training data like Artificial Neural Networks and Support Vector Machines. Many research contributions from this category have been recently proposed. An Artificial Neural Networks based model was developed in Külahci et al. (2009) to study the relationship between radon and earthquakes. Panakkat and Adeli (2009) proposed a new recurrent neural networks (RNN) model to predict earthquake time and location using a vector of eight seismicity indicators as input. To give a better estimation of radon variations, Negarestani et al. (2002) suggested layered neural networks analyse the relationship between radon concentration and environmental parameters for earthquake prediction in Thailand. In Moustra et al. (2011), authors developed three variations of Neural Networks models analysing the seismic electrical signals (SES) recorded in Greece. The first model which is the basic one has only one output which is the magnitude of the predicted impending earthquakes. The second one considers another extra input which is the average magnitude for the 30 previous days and gives the magnitude of predicted upcoming earthquakes. Finally, the third model is the same as the basic one but it predicts in addition to the magnitude, the time lag between the date on which SES were recorded and the date of the impending major earthquake.

In Asim et al. (2018a), authors tried to consider the maximum of information on seismic activity by calculating seismic indicators in different regions. After that, they apply the genetic programming and Adaboost (GP- Adaboost) as an ensemble method to predict earthquakes of magnitude 5.0 and above. The paper (Asim et al., 2018b) predicts earthquakes of three different regions using seismic features of geophysical and seismological concepts. Its authors construct a support vector machine regressor combined by a hybrid neural network merged by three different ANNs, the final model is applied to their data.

Additionally, other contributions (Asencio–Cortés et al., 2017, 2018; Buscema et al., 2015) use different machine learning classifiers like Naive Bayes, Support Vector Machines and Random Forest for earthquake prediction. For instance, Five classifiers were used in Asencio–Cortés et al., (2018) to predict the maximum earthquake magnitude in the upcoming seven days in California.

The ability of deep learning to discover complex patterns in data conduct scientists to make its applications on earthquake prediction, where they benefit from the fact that no feature extraction is required (Mignan and Broccardo, 2019).

For instance, in Li and Liu (2016) an improved variant of particle swarm optimisation (PSO) was applied combined with backpropagation neural networks to predict earthquakes. In Mahmoudi et al. (2016), they develop 128 different MLP networks to find the best architecture of the magnitude earthquake prediction model. The work in Narayanakumar and Raja (2016) proposes a three-layer feed-forward BP NN to predict earthquakes in the region of Himalaya. The input datasets used are the seismic indicators and historical data. An earthquake location and time prediction of moderate to large earthquakes using seismic indicators and RNNs is presented in Panakkat and

Adeli (2009), it considers two cases of studies: location decomposition and time decomposition. In Parameswaran et al. (2020), they have implemented the LSTM network using AdaGrad optimiser to predict the coming earthquakes. Their dataset considers a whole area instead of using subregions. The model is a two-dimensional LSTM that uses a normalised input dataset. The results show that training the LSTM AdaGrad is more accurate using the LSTM only.

## 3 Earthquake prediction: method and datasets

Earthquake prediction aims generally to analyse various seismic data to forecast earthquakes. This context matches well with the deep learning and machine learning objectives, which considers learning problem as a problem of learning from experience concerning some tasks and performance measures. For this reason, more and more recent works tend to adopt machine learning and deep learning techniques as prominent techniques that could lead to emerging models for earthquake prediction from large recorded datasets. This section focuses on this category of work. It is divided into three subsections which respectively present the earthquake prediction process, the dataset representation, and the LSTM model architecture that we use in this work.

### *3.1 Process for earthquake prediction*

In the field of machine learning and deep learning based earthquake prediction, the adopted prediction process usually includes five steps as described below:

- *Data acquisition*: This step consists in collecting important and huge data of historical earthquakes which can lead to achieving the targeted objectives. The dataset we use in this work is recorded and provided by the National Geophysics Institute (CNRST) from 1900 to 2019. It belongs to the Moroccan regions and it contains 32396 seismic events.
- *Pre-processing*: It aims to clean the collected data. In this step, it is important to remove noises and eliminate or transform all non-significant details, remove redundancy, eliminate details that do not affect the prediction, usually delete incomplete data and normalise data having a large size.
- *In this step*, we delete the redundancy and remove the negative values of magnitudes from our datasets since they present not felt events. The negative values in our datasets are outliers that skew the training process of our model and lead to bad results.
- *Feature extraction and generation*: This step consists in extracting the relevant seismic characteristics from the data. The relevance of selected information is estimated based to its ability to affect the prediction. In this step, we replace the time attributes with the one simple and appropriate time parameter, which presents the number of seconds between seismic events.
- *Error metric definition*: In this step, it is important to select the metric most appropriate for determining the best set of parameters. In this work, the used metrics are the mean squared error, the mean absolute error, and accuracy.
- *Processing*: This is the main step that consists in training the adopted model on the dataset and evaluating and comparing the predicted results with the measured values based on the defined error metrics. Before that, we normalise the final dataset using the Min-Max scaler which transforms datasets to an exact same scale, in a range between 0 and 1.

### *3.2 Dataset representation*

In the literature, earthquake prediction is based on analysing seismic features and different anomalies, which have reliable relationship to earthquakes. These anomalies constitute patterns of occurrences of seismic events during a time period or during full/new moon periods. They refer generally to strange or irregular animal behaviours, soil gas and liquid movements and concentrations before the earthquake formation, physical characteristics of rocks, electrical signals, thermal and electromagnetic anomalies, water level, unusual weather and atypical cloud.

With regard to the works (Asencio–Cortés et al., 2018; Külahci et al., 2009; Moustra et al., 2011; Alarifi et al., 2012; Reyes et al., 2013), we have identified a set of main seismic features and anomalies indicators used as datasets to predict the earthquake magnitudes using machine learning methods. These features and indicators are as follows:

- *Magnitude distribution*: It is the distribution of historical magnitudes retrieved from the recorded data. It constitutes the most important seismic feature because it represents the fluctuation of magnitude (Alarifi et al., 2012). In addition, this information gives an overview of the rate of earthquakes in a specific region.
- *Source depth*: It is the depth at which an earthquake occurs. This earthquake feature shows its classification and determines where the most seismic events are concentrated.
- *Earthquake location*: It determines the locations where the seismic events occurred. This feature is practically represented by the geographic coordinate’s longitude and latitude.
- *Date and time*: They are usually represented in the recorded data by the year, month, day and time.
- *Number of earthquakes*: It refers to the total number of earthquakes in any given region and time period. This indicator is calculated using Gutenberg-Richter’s law which represents the relationship between the

magnitudes and the total number of earthquakes. In the literature, Richter (2018) has noted that the number of shocks decreases very rapidly for the higher magnitudes. In the later studies, Gutenberg and Richter (1941) have suggested an exponential distribution for the number of earthquakes vs. the magnitude. Again, Gutenberg and Richter (1944) have transformed this law into a linear law expressing this relation for the magnitude frequency distribution as illustrated in equation (1). $N(M)$ is the number of events with magnitude larger or equal to $M$, $a$ is the seismic activity and $b$ known as $b$-value is a size distribution factor.

$$\log 10\ N(M) = a - bM \tag{1}$$

- *b-value*: The $b$-value is a feature that reflects the physical characteristics of the area under analysis (Lee and Yang, 2006). It constitutes a very important feature since it reflects the geophysical properties of rocks and fluid in a specific region. In Gibowicz (1974) and Wiemer et al., (2002), authors show that the $b$-value increases after large earthquakes in New Zealand and decreases before the next aftershocks. They conclude that the variation of the $b$-value over time refers to aftershocks and that this value tends to decrease when many earthquakes occur in a local area during a short period of time.

- *Aftershocks rate*: It refers to the rate of aftershocks in the elapsed time t since the mainshock. This indicator is calculated using Omori-Utsu's law. We note that Omori's law determines the rate of aftershocks with time which falls-off very rapidly by time after the mainshock (Utsu, 1999). It is calculated according to equation (2). $N(t)$ is the rate of aftershocks in the elapsed time $t$ since the mainshock, $K$ is the amplitude and c is the time offset parameter that is typically much less than one day. Omori-Utsu law is the modified version of Omori's proposed in 1967 by Utsu (1961). As illustrated in equation (3), it uses the p-value which is the fitness parameter that modifies the decay rate and typically falls in the range 0.8–1.2.

$$N(t) = K/(c + t) \tag{2}$$

$$N(t) = K/(c + t)^p \tag{3}$$

- *Average difference in magnitude*: It refers to the average difference in magnitude between a mainshock and its largest aftershock. Bath's law states that it is constant and typically equals to 1.1–1.2 Mw regardless of the mainshock magnitude (Båth, 1965).

- *Radon concentration*: It refers to soil radon content. Radon is a radioactive noble gas. Evaluating its content aims to detect changes in its level, this would be useful as a potential earthquake predictor.

- *Electrical signals*: They are geoelectric voltages referring to low frequency electric signals. Known as SES in VAN method of physics professors Panayiotis Varotsos, Kessar Alexopoulos and Konstantine Nomicos (VAN), these signals have been recommended in 1981 as indicators for predicting earthquakes of magnitude larger than 2.8 within all of Greece up to seven hours beforehand. VAN method (Varotsos et al., 1986, 1988) is an experimental method of earthquake prediction. It is based on observing and assessing SES that occur several hours to days before the earthquake which can be used as warning signs. This method has successfully predicted about 60% of Greek earthquakes of magnitude larger than 5.3 on the Richter scale (Uyeda, 1997).

### *3.3 LSTM architecture*

In earthquake prediction research field, the most used methods are based on neural networks and DNN models.

ANN is a mathematical model for information processing, inspired by the way that biological nervous systems process the information. ANN is in practice, a network of nodes, called neurons, connected by directed links. The methods founded on DNN and ANN are in fact, based on training a multi-layer ANN where each layer learns to transform its input data into a more abstract and composite representation. These methods are founded on a layered-based iterative procedure using non-linear transformations of data for pattern recognition. They explore an ANN with multiple hidden layers between input and output layers. All methods proposed in Alarifi et al. (2012), Ozerdem et al. (2006), Galkina and Grafeeva (2019), Su and Zhu (2009) and Bhatia et al. (2018) used feed-forward ANNs. These ANNs are the most powerful and most popular neural networks for nonlinear regression Asencio–Cortés et al. (2018). A Feed-forward network sends the information between neurons in only one direction forward, from the input neurons, through the hidden layers to the output neurons. The DL method used in Bengio et al. (2014) explores a multi-layer feed-forward ANN that is trained with stochastic gradient descent using back-propagation learning algorithm which is useful for feed-forward networks. It adjusts the weights of each unit in such a way that the error between the desired output and the actual output is reduced (Rumelhart et al., 1986). Besides, DL enables performing supervised, semi-supervised or unsupervised learning. However, DNNs, like ANNs, present some limits in terms of overfitting and computation time. They are likely to overfit because of the added layers of abstraction, which allow modelling rare dependencies in the training data.

To deal with these problems, deep learning comes with strong algorithms, especially for time series analysis. Deep learning methods and approaches do not need feature engineering and extraction, and it gives effective results even with unstructured data. Deep learning is usually known by the high-quality results in complex predictions and pattern recognition.

Earthquakes datasets are time-series data and difficult to forecast. To predict them reliably, it is crucial to try one of the most powerful algorithms dedicated to sequence data called LSTM.

LSTM is a RNN proposed by Hochreiter and Schmidhuber (1997). RNN are a type of artificial neural network enhanced by their memory state. RNNs are applicable in time series data because such data are usually dependent on each other. However, classical RNNs do not present the best solution. They suffer from vanishing and exploding gradient and become untrainable.

LSTM resolves the problems of RNNs, it contains three important components called gates: input gate, forget gate, output gate; and two memory cells: hidden state and internal state.

These steps explain the functioning of LSTMs (see equations (4)–(8):

First, the input data is initially squashed by the Tanh function to make them very small values and present them in a non-linear manner.

Second, the squashed data pass to the input gate. The latter takes the relevant information and filters the no required elements by multiplying them with a sigmoid function.

Third, the internal state or the memory of the current state takes the information stored in the previous state and adds it to the input data. It uses an addition operation instead of multiplication to avoid the vanishing problem.

After that, the recurrence of states is enforced by a forget gate. This one decides which state elements should be memorised or forgotten using a sigmoid function.

Finally, a Tanh function squashes the outputs then the forget gate decides which elements should be stored and set as outputs of the current cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\check{c}_t = \tan h(W_c[h_{t-1}, x_t] + b_c) \tag{5}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{6}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

$$c_t = f_t * c_{t-1} + i_t * \check{c}_t \tag{8}$$

$$h_t = o_t * \tan h(c_t) \tag{9}$$

where $i_t$, $\check{c}_t$, $f_t$, $o_t$, $c_t$, $h_t$ are the input gate, cell input activation, forget gate, output gate, cell state, and the hidden state respectively. $W_i$, $W_c$, $W_f$ and $W_o$ are their weight matrices respectively. $b_i$, $b_c$, $b_f$, and $b_o$ are the biases. $X_t$ is the input, $h_{t-1}$ is the last hidden state, $h_t$ is the internal state. $\sigma$ is the sigmoid function.

## 4 Earthquake prediction approach using LSTM algorithm and an enhanced time feature

This section focuses on explaining our proposed earthquake prediction model. Initially, we start by presenting and proving the importance of using our enhanced data representation, and then we define the architecture of our LSTM-based model.

### 4.1 Data and time feature representation

Data representation is an essential step in machine learning and deep learning, which could ease learning complex tasks by the model. Simplifying and reducing the dimensionality of data is very crucial, especially when learning non-correlated data. In the case of earthquake prediction multiple works tried to compute and generate the geophysical indicators (Section 3), these indicators were calculated from the basic dataset like magnitudes, foreshocks, and aftershocks. This way is computationally expensive and takes the same trends and patterns of the basic dataset by presenting them in another structure.

Our work is based on the DL algorithm LSTM, a strong model that is capable to learn the patterns from the basic features without generating others. In the work (Ozerdem et al., 2006) authors demonstrate that ANN and deep learning do not need a feature generation process because they are skilled in learning and extracting the insights from datasets by themselves. Where they evaluate the performance of DL models with feature generation and without it, and finally they found that DL does not need feature engineering.

For these reasons, our model focuses on training the main and important dataset features which are: the magnitude distribution, source depth, earthquake location (longitude and latitude), and an inferred time feature.

In the field of machine learning-based prediction, the date and time are hard to learn. On one hand, they could be presented in split attributes (year, month, day, hour, minute, and second) which expand the dimensionality of the model. On the other hand, they could be composed where they do not present any correlation with the other features.

To solve these problems, some works use a sequential number to save the chronology of the seismic events and ignore the date and time attributes. In contrast, the date and time parameters are very important in earthquake prediction since they are time-series data. The year and month or even the time could be related to the cause of a seismic event. Also, it is possible to take just the year since it is usually the most correlated with magnitude distribution.

**Table 1** Correlation coefficients of different date and time representations (SEQUENTIAL number, year, month, day, hour, minute, second and the inferred time that we propose) with magnitude distribution

| *Correlation with magnitude distribution* | *Date and time representation* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *Seq* | *Year* | *Month* | *Day* | *Hour* | *Min* | *Sec* | *Inferred time* |
| Coefficients | 59% | 55.46% | 0% | 0% | 0% | 0% | 0% | 56% |

In this regard, we suggest calculating another simpler and expressive feature, which is inferred from date and time and saves the exact information without wasting any parameter of them. This time feature refers to the number of elapsed seconds from the first seismic event. To make things easier, this feature transforms the date and time components to one parameter that counts the number of seconds to each event. Giving such exact and clear information will enhance and gain our model a great training quality, especially when it is correlated with the magnitude distribution (see Table 1).

## *4.2 Method and hyperparameters identification*

LSTM is one of the effective methods of RNNs. RNNs are deep learning algorithms, adapted to sequence data, they are extremely expressive since they are capable to learn complex vector-to-vector mappings.

Our approach is based on the application of LSTM algorithm using an enhanced data representation. This family of algorithms is known by its high parametrisation. The values of parameters defined when using an algorithm affect its behaviour in terms of error tolerance, number of iterations, variants, etc. The learning time and accuracy of the algorithm can sometimes depend greatly on the choice of appropriate parameters. Algorithms with large numbers of parameters require generally, more testing to find the right combination. For this reason, we apply the search grid: a hyperparameter optimisation technique, which allows us to test and compare the performance of several different combinations of parameters, to give the optimal parameterisation of our model. The grid search is typically computationally expensive, and we use it in this work since we cannot manually tune this number of parameters.

The purpose of using the grid search technique in our approach is to find the best values for these hyperparameters:

- The number of memory cells of the LSTM model.
- The Batch-size that presents the number of samples that will be propagated in the network.
- Epochs, the maximum number of iterations that the model needs to correctly learn.
- An activation function that calculates the output of each node to determine the final output, the choice of the activation function is very crucial since it presents an important effect on the model's ability to converge and on the convergence speed. The type of activation function depends on the nature of the problem and datasets. In our case, we use non-linear functions because the seismic activity is non-linear and complex.
- A dropout rate to specify the probability of setting each unit to 0 at each update and that is the main attribute of the dropout function. The dropout function is used to drop out of the network some neurons, where it does not consider them during the training process. This function ignores the co-dependency between neurons and helps the model to avoid overfitting.
- An Optimiser, the method used to update the values of weights and learning rates during the training phase to reduce the losses. There are several existing methods like SGD, Adagrad, Adadelta, and Adam.

In Berhich et al. (2020), we built an LSTM model in two cases of studies with the same data of the Moroccan regions we use in this work. The first case gives the prediction of earthquakes using all datasets and the LSTM model. The second case uses two LSTM models with data decomposition. The decomposition aims to separate large earthquakes from small and medium earthquakes to generate two groups of datasets. The paper compares the two cases and demonstrates that earthquake prediction with data decomposition is more effective than using the whole dataset at one time.

**Figure 1** Flow chart of the proposed LSTM model

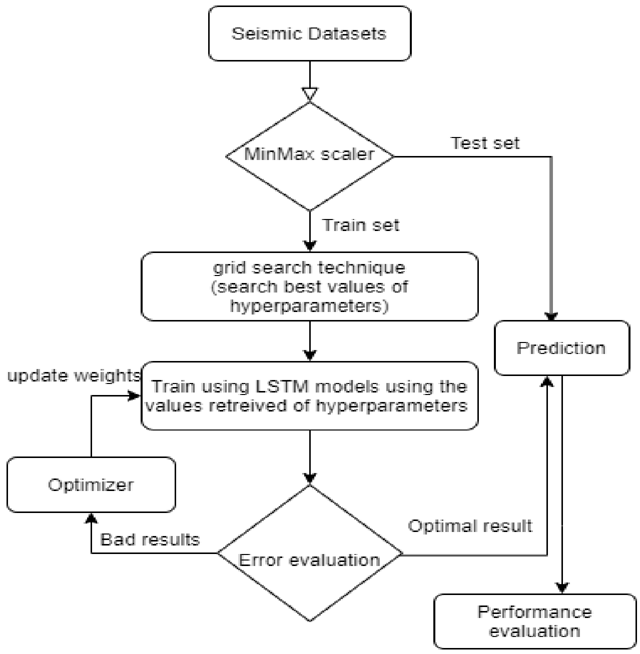



**Table 2** Best combination values of hyperparameters retrieved by the grid search technique when using the LSTM and multi layered NN models

| | *Models* | |
|---|---|---|
| *Hyper parameters* | *LSTM* | *Multi-layer NN* |
| Number of memory cells | 75 | 50 |
| Batch size | 128 | 128 |
| Epochs | 100 | 100 |
| Activation function | reLu | reLu |
| Dropout rate | 0.05 | 0.05 |
| Optimiser | Adamax | Adam |
| Loss function | Mean squared error | Mean squared error |

The flowchart in Figure 1 illustrates our model clearly. At first, we start by normalising the datasets using the Min-Max scaler. After that, the scaled data is split into 80% for training and 20% for the testing data. Then, the grid search technique will define the optimal combination of hyperparameters from a set of different values that we define before (see Table 2). Afterward, the model is trained using the best parameter values retrieved, till that it converges to the minimum error using the best optimiser and loss function for error evaluation. Finally, the trained model is applied to testing data to evaluate its performance. The performance metrics we use for evaluation are mean absolute error, mean squared error, and accuracy.

## 5 Evaluation and discussion

Obtaining the most accurate answer possible is necessary for earthquake prediction. Claiming wrong results can lead to a big loss. In this perspective, we present in this section the results of our LSTM predictive model. The experiment results are illustrated in Table 3.

**Table 3** Spatiotemporal magnitude prediction experiment's results of LSTM and multi-layer NN models, using the performance metrics accuracy, MAE and MSE, and the elapsed time during training process

| *Performance metrics* | *Models* | |
|---|---|---|
| | *LSTM* | *Multi-layer NN* |
| MAE | 0.031 | 0.074 |
| MSE | 0.0032 | 0.012 |
| Accuracy | 99% | 96% |
| Elapsed time by seconds | 654.40 | 98.910 |

To evaluate the performance of our work we use the metrics: mean absolute error, mean squared error, and accuracy.

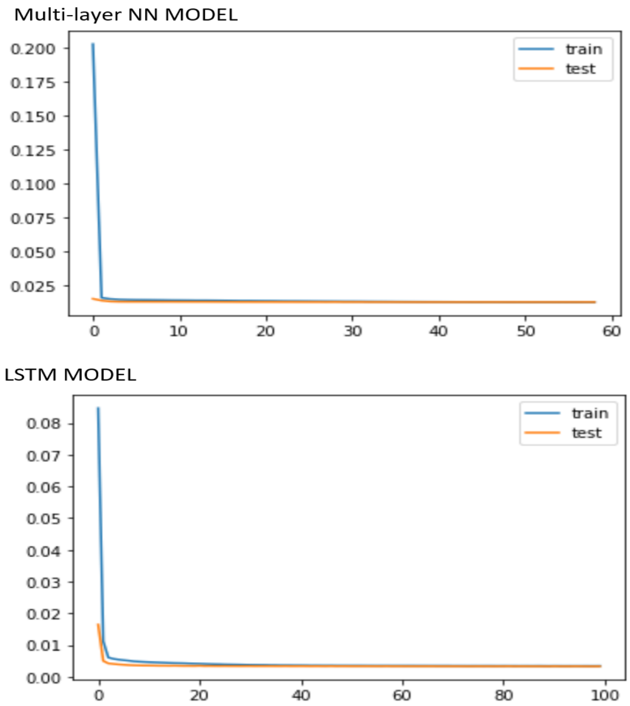
One of the ill-posed problems of earthquake prediction is the difficulty of comparing the performances of the works in literature, because of the variety of studied regions, datasets, and performance metrics used by authors. For this reason, we are applying a multi-layer neural network (NN) on datasets to evaluate the performance of the proposed model. The hyperparameters of the multilayer NN are defined using the same grid search technique. We choose the multi-layer NN for evaluation because it is widely used in literature as it is mentioned in Section 2.

According to the results (Table 3), we find that the LSTM model that we propose is outperforming the multi-layer NN using the three metrics, where the MAE is 0.031 when using LSTM and 0.074 when using multi-layer NN, the MSE is 0.0032 when using LSTM and 0.012 when using multi-layer NN, and finally, the accuracy of LSTM model is more effective than multi-layer NN by 3%, but it is slower because it takes its time to converge to the minimum error.

The models in Figure 2 illustrates the fitting curves of the model's LSTM and multi-layer NN. The fitting is good in both models. No underfitting or overfitting is observed.

As we can see, the approaches proposed in the literature for earthquake prediction, namely those based on machine learning and artificial neural networks, use different approaches based on various elements and indicators as a dataset. The success of some of them has been attained by chance. In 2011, the International Commission on Earthquake Forecasting for Civil Protection (ICEF) considered the search for useful precursors as unsuccessful. It concluded that the prediction capability claimed by VAN could not be validated (Jordan et al., 2011). In contrast, the geophysical indicators, namely, the number of earthquakes, the b-value, the aftershocks rate and the average difference in magnitude, have been recommended as input features for the machine learning models in Asencio–Cortés et al. (2018) and Asim et al. (2018a) since they allow achieving an important accuracy of prediction. These indicators are calculated from the principal features that we use in this work. That is why we do not need to generate them since the LSTM algorithms are capable to give such performant results without any feature extraction or generation. Besides, our enhanced time feature adds a great potential and quality of prediction, because of its simplicity, and the fact that it preserves the tendency of seismic activity in one soft parameter.

**Figure 2** Fitting curves of LSTM and multi-layer NN training process (see online version for colours)




In brief, this work attains four important goals:

- a reduced number of features
- an enhanced and simplified feature that contains the trends of date and time in one information
- a spatiotemporal magnitude prediction
- the best and optimal combination values of hyperparameters are identified by the grid search technique.

Finally, this improved earthquake prediction model is outperforming our previous LSTM model presented in Berhich et al. (2020) and it gives more performant and better results.

## 6 Conclusion

In this paper, we suggest an enhanced approach for spatiotemporal earthquake magnitude prediction using the datasets of the Moroccan regions. Our approach applies an LSTM model on six features: the time, magnitude, source depth, longitude, and latitude. The time feature we propose is a simple and better transformation of date and time features, which could be complex and hard to learn. To identify the model's hyperparameters we use the grid search technique, which gives the optimal combination of hyperparameter values. Our LSTM model is compared with a multilayer neural network model that we apply on the same dataset. The final results improve that our enhanced LSTM model provides effective forecasting and achieves favourable performance compared with others.

## References


Alarifi, A.S.N., Alarifi, N.S.N. and Al-Humidan, S. (2012) 'Earthquakes magnitude predication using artificial neural network in northern Red Sea area', *J. King Saud Univ. – Sci.*, Vol. 24, No. 4, pp.301–313, doi: 10.1016/j.jksus.2011.05.002.

Allen, C.R. (1976) 'Responsibilities in earthquake prediction', *Bull. Seismol. Soc. Am.*, Vol. 66, No. 6, pp.2069–2074, [Online], Available: http://www.bssaonline.org/content/66/6/2069%5Cnpapers3://publication/uuid/9D60F83C-E85F-4B1E-8602-76C09AD4773B

Asencio–Cortés, G., Morales–Esteban, A., Shang, X. and Martínez–Álvarez, F. (2018) 'Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure', *Comput. Geosci.*, Vol. 115, June, pp.198–210, doi: 10.1016/J.CAGEO.2017.10.011.

Asencio–Cortés, G., Scitovski, S., Scitovski, R. and Martínez–Álvarez, F. (2017) 'Temporal analysis of Croatian seismogenic zones to improve earthquake magnitude prediction', *Earth Sci. Informatics*, Vol. 10, No. 3, September, pp.303–320, doi: 10.1007/s12145-017-0295-5.

Asim, K.M., Idris, A., Iqbal, T. and Martínez-Álvarez, F. (2018a) 'Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification', *Soil Dyn. Earthq. Eng.*, Vol. 111, August, pp.1–7, doi: 10.1016/j.soildyn.2018.04.020.

Asim, K.M., Idris, A., Iqbal, T. and Martínez-Álvarez, F. (2018b) 'Earthquake prediction model using support vector regressor and hybrid neural networks', *PLoS One*, Vol. 13, No. 7, July, p.e0199004, doi: 10.1371/journal.pone.0199004.

Båth, M. (1965) 'Lateral inhomogeneities of the upper mantle', *Tectonophysics*, Vol. 2, No. 6, December, pp.483–514, doi: 10.1016/0040-1951(65)90003-X.

Baykara, O., Doğru, M., İnceöz, M. and Aksoy, E. (2005) 'Measurements of radon emanation from soil samples in triple-junction of North and East Anatolian active faults systems in Turkey', *Radiat. Meas.*, Vol. 39, No. 2, February, pp.209–212, doi: 10.1016/J.RADMEAS.2004.04.011.

Bengio, Y., Simard, P. and Frasconi, P. (2014) 'Artificial adaptive systems to predict the magnitude of earthquakes', *IEEE Trans. Neural Netw.*, Vol. 5, No. 2, p.157, doi: 10.4430/bgta0144.

Berhich, A., Belouadha, F. and Kabbaj, M.I. (2020) *LSTM-based Models for Earthquake Prediction*, p.6, doi: https://doi.org/10.1145/3386723.3387865.

Bhatia, A., Pasari, S. and Mehta, A. (2018) 'Earthquake forecasting using artificial neural networks', *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. – ISPRS Arch.*, Vol. 42, No. 5, pp.823–827, doi: 10.5194/isprs-archives-XLII-5-823-2018.

Bird, P., Letters, Z.L-S.R. and undefined (2007) 'Seismic hazard inferred from tectonics: California', *cseptesting.org*, [Online], Available: http://www.cseptesting.org/documents/files/relm/Bird.pdf (Accessed 17 October, 2018).

Buscema, P.M., Massini, G. and Maurelli, G. (2015) 'Artificial adaptive systems to predict the magnitude of earthquakes', *Boll. di Geofis. Teor. ed Appl.*, Vol. 56, No. 2, pp.227–256, doi: 10.4430/bgta0144.

Console, R., Murru, M., Catalli, F. and Falcone, G. (2007) 'Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: comparison with a purely stochastic ETAS model', *Seismol. Res. Lett.*, Vol. 78, No. 1, January, pp.49–56, doi: 10.1785/gssrl.78.1.49.

Ebel, J.E., Chambers, D.W., Kafka, A.L. and Baglivo, J.A. (2007) 'Non-Poissonian earthquake clustering and the hidden markov model as bases for earthquake forecasting in California', *Seismol. Res. Lett.*, Vol. 78, No. 1, January, pp.57–65, doi: 10.1785/gssrl.78.1.57.

Erees, F.S., Aytas, S., Sac, M.M., Yener, G. and Salk, M. (2007) 'Radon concentrations in thermal waters related to seismic events along faults in the Denizli Basin, Western Turkey', *Radiat. Meas.*, Vol. 42, No. 1, January, pp.80–86, doi: 10.1016/J.RADMEAS.2006.06.003.

Galkina, A. and Grafeeva, N. (2019) 'Machine learning methods for earthquake prediction: a survey', *CEUR Workshop Proc.*, Vol. 2372, pp.25–32.

Gerstenberger, M., Jones, L., Letters, S.W-S.R. and undefined (2007) 'Short-term aftershock probabilities: case studies in California', *Citeseer* [Online], Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.956.190&rep=rep1&type=pdf (Accessed 17 October, 2018).

Gibowicz, S.J. (1974) 'Frequency-magnitude, depth, and time relations for earthquakes in an island arc: North Island, New Zealand', *Tectonophysics*, Vol. 23, No. 3, August, pp.283–297, doi: 10.1016/0040-1951(74)90028-6.

Gutenberg, B. and Richter, C.F. (1941) *Seismicity of the Earth*, pp.1–126.

Gutenberg, R. and Richter, C.F. (1944) 'Frequency of earthquakes in California', *Bulletin of the Seismological Society of America*, Vol. 34, pp.185–188.

Hayakawa, M. (2015) *Fundamentals of Earthquake Prediction, Earthquake Prediction with Radio Techniques*, 11th ed., Hardcover, Wiley & Sons Ltd., August, 304 pages, ISBN: 978-1-118-77016-0.

Helmstetter, A., Kagan, Y. and Jackson, D. (2006) 'High-resolution time-independent grid-based forecast for M5 earthquakes in California', *Seismol. Res. Lett.*, Vol. 78, No. 1, pp.78–86 [Online], Available: https://hal.archives-ouvertes.fr/hal-00195399/ (Accessed 17 October, 2018).

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Comput.*, Vol. 9, No. 8 November, pp.1735–1780, doi: 10.1162/neco.1997.9.8.1735.

Jordan, T.H., Chen, Y-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., Papadopoulos, G., Sobolev, G., Yamaoka, K. and Zschau, J. (2011) 'Operational earthquake forecasting. State of knowledge and guidelines for utilization', *Ann. Geophys.*, Vol. 54, No. 4, August, doi: 10.4401/ag-5350.

Kagan, Y.Y., Jackson, D.D. and Rong, Y. (2007) 'A testable five-year forecast of moderate and large earthquakes in Southern California based on smoothed seismicity', *Seismol. Res. Lett.*, Vol. 78, No. 1, January, pp.94–98, doi: 10.1785/gssrl.78.1.94.

Külahci, F., Inceöz, M., Do□ru, M., Aksoy, E. and Baykara, O. (2009) 'Artificial neural network model for earthquake prediction with radon monitoring', *Appl. Radiat. Isot.*, Vol. 67, No. 1, pp.212–219, doi: 10.1016/j.apradiso.2008.08.003.

Külahci, F., Inceöz, M., Do□ru, M., Aksoy, E. and Baykara, O. (2009) 'Artificial neural network model for earthquake prediction with radon monitoring', *Appl. Radiat. Isot.*, Vol. 67, No. 1, January, pp.212–219, doi: 10.1016/j.apradiso.2008.08.003.

Lee, K. and Yang, W-S. (2006) 'Historical seismicity of Korea', *Bull. Seismol. Soc. Am.*, Vol. 96, No. 3, June, pp.846–855, doi: 10.1785/0120050050.

Li, C. and Liu, X. (2016) 'An improved PSO-BP neural network and its application to earthquake prediction', *Proceedings of the 28th Chinese Control and Decision Conference, CCDC 2016*, August, pp.3434–3438, doi: 10.1109/CCDC.2016.7531576.

Mahmoudi, J., Arjomand, M.A., Rezaei, M. and Mohammadi, M.H. (2016) 'Predicting the earthquake magnitude using the multilayer perceptron neural network with two hidden layers', *Civ. Eng. J.*, Vol. 2, No. 1, January, pp.1–12, doi: 10.28991/cej-2016-00000008.

Mignan, A. and Broccardo, M. (2019) 'Neural *Network Applications in Earthquake Prediction (1994-2019): Meta-Analytic Insight on their Limitations*, September, pp.1–25 [Online], Available: http://arxiv.org/abs/1910.01178

Moustra, M., Avraamides, M. and Christodoulou, C. (2011) 'Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals', *Expert Syst. Appl.*, Vol. 38, No. 12, November, pp.15032–15039, doi: 10.1016/j.eswa.2011.05.043.

Narayanakumar, S. and Raja, K. (2016) 'A BP artificial neural network model for earthquake magnitude prediction in Himalayas, India', *Circuits Syst.*, Vol. 07, No. 11, pp.3456–3468, doi: 10.4236/cs.2016.711294.

Negarestani, A., Setayeshi, S., Ghannadi-Maragheh, M. and Akashe, B. (2002) 'Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction', *J. Environ. Radioact.*, Vol. 62, No. 3, January, pp.225–233, doi: 10.1016/S0265-931X(01)00165-5.

Ozerdem, M.S., Ustundag, B. and Demirer, R.M. (2006) 'Self-organized maps based neural networks for detection of possible earthquake precursory electric field patterns', *Adv. Eng. Softw.*, Vol. 37, No. 4, pp.207–217, doi: 10.1016/j.advengsoft.2005.07.004.

Panakkat, A. and Adeli, H. (2009) 'Recurrent neural network for approximate earthquake time and location prediction using multiple seismicity indicators', *Comput. Civ. Infrastruct. Eng.*, Vol. 24, No. 4, May, pp.280–292, doi: 10.1111/j.1467-8667.2009.00595.x.

Parameswaran, S, Nandha Gopal, A., Revathi, D. and Senthamarai, M. 2020) *Earthquake Prediction System by LSTM* [Online], Available: www.ijiarec.com (Accessed 24 April, 2020).

Petersen, M.D., Cao, T., Campbell, K.W. and Frankel, A.D. (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.*, Vol. 78, No. 1, January, pp.99–109, doi: 10.1785/gssrl.78.1.99.

Reyes, J., Morales-Esteban, A. and Martínez-Álvarez, F. (2013) 'Neural networks to predict earthquakes in Chile', *Appl. Soft Comput. J.*, Vol. 13, No. 2, pp.1314–1328, doi: 10.1016/j.asoc.2012.10.014.

Rhoades, D.A. (2007) 'Application of the EEPAS model to forecasting earthquakes of moderate magnitude in southern California', *Seismol. Res. Lett.*, Vol. 78, No. 1, January, pp.110–115, doi: 10.1785/gssrl.78.1.110.

Richter, C.F. (2018) 'An instrumental earthquake magnitude scale', *Bull. Seismol. Soc. Am.*, Vol. 25, No. 1, January, pp.1–32 [Online], Available: https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/25/1/1/115102 (Accessed 17 October, 2018).

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) 'Learning internal representations by error propagation (No. ICS-8506)', *Calif. Univ San Diego La Jolla Inst Cogn. Sci.*, Vol. 1, pp.318–362, doi: 10.1016/B978-1-4832-1446-7.50035-2.

Şen, Z. (1998) 'Point cumulative semivariogram for identification of heterogeneities in regional seismicity of Turkey', *Math. Geol.*, Vol. 30, No. 7, pp.767–787, doi: 10.1023/A:1021704507596.

Şen, Z. and Al-Suba'i, K. (2001) 'Seismic hazard assessment in the Tihamat Asir Region, Southwestern Saudi Arabia', *Math. Geol.*, Vol. 33, No. 8, pp.967–991, doi: 10.1023/A:1012579323959.

Shen, Z-K., Jackson, D.D. and Kagan, Y.Y. (2007) 'Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 earthquakes in Southern California', *Seismol. Res. Lett.*, Vol. 78, No. 1, January, pp.116–120, doi: 10.1785/gssrl.78.1.116.

Špičák, A. and Vaněk, J. (2016) 'Earthquake swarms reveal submarine magma unrest induced by distant mega-earthquakes: Andaman Sea region', *J. Asian Earth Sci.*, Vol. 116, No. 116, February, pp.155–163, doi: 10.1016/j.jseaes.2015.11.017.

Su, Y.P. and Zhu, Q.J. (2009) 'Application of ANN to prediction of earthquake influence', *2009 2nd International Conference on Information and Computing Science, ICIC 2009*, Vol. 2, pp.234–237, doi: 10.1109/ICIC.2009.169.

Utsu, T. (1961) 'A statistical study on the occurrence of aftershocks', *Geophys. Mag.*, Vol. 30, pp.521–605 [Online], Available: https://ci.nii.ac.jp/naid/10006083239/ (Accessed 17 October, 2018).

Utsu, T. (1999) 'Representation and analysis of the earthquake size distribution: a historical review and some new approaches', *Seismicity Patterns, their Statistical Significance and Physical Meaning*, Birkhäuser Basel, Basel, pp.509–535.

Uyeda, S. (1997) 'The VAN method of short-term earthquake prediction', *RIKEN International Frontier Program on Earthquake Research. INCEDE Newsletter*, Vol. 5, No. 4, pp.2–4, http://sems-tokaiuniv.jp/old/eprc/res/incede/incede2a.html

Varotsos, P., Alexopoulos, K., Nomicos, K. and Lazaridou, M. (1988) 'Official earthquake prediction procedure in Greece', *Tectonophysics*, Vol. 152, Nos. 3–4 September, pp.193–196, doi: 10.1016/0040-1951(88)90045-5.

Varotsos, P., Alexopoulos, K., Nomicos, K. and Lazaridou, M. (1986) 'Earthquake prediction and electric signals', *Nature*, Vol. 322, No. 6075, July, pp.120–120, doi: 10.1038/322120a0.

Wang, K., Chen, Q.F., Sun, S. and Wang, A. (2006) 'Predicting the 1975 Haicheng earthquake', *Bull. Seismol. Soc. Am*., Vol. 96, No. 3, pp.757–795, doi: 10.1785/0120050191.

Ward, S.N. (2007) 'Methods for evaluating earthquake potential and likelihood in and around California', *Seismol. Res. Lett*., Vol. 78, No. 1, January, pp.121–133, doi: 10.1785/gssrl.78.1.121.

Wiemer, S., Gerstenberger, M. and Hauksson, E. (2002) 'Properties of the aftershock sequence of the 1999 Mw 7.1 hector mine earthquake: implications for aftershock hazard', *Bull. Seismol. Soc. Am*., Vol. 92, No. 4, May, pp.1227–1240, doi: 10.1785/0120000914.

Yalım, H.A., Sandıkcıoğlu, A., Ünal, R. and Orhun, Ö. (2007) 'Measurements of radon concentrations in well waters near the Akşehir fault zone in Afyonkarahisar, Turkey', *Radiat. Meas*., Vol. 42, No. 3, March, pp.505–508, doi: 10.1016/J.RADMEAS.2006.12.013.

Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J. and Kobal, I. (2003) 'Application of decision trees to the analysis of soil radon data for earthquake prediction', *Appl. Radiat. Isot*., Vol. 58, No. 6, June, pp.697–706, doi: 10.1016/S0969-8043(03)00094-0.