ISSN: 0975-766X
CODEN: IJPTFI
Research Article

Available Online through
www.ijptonline.com

NEURAL NETWORK TECHNIQUE IN DATA MINING FOR PREDICTION OF EARTH QUAKE

K.Karthikeyan, Sayantani Basu
[1]Associate professor, Statistics & O.R division, SAS, VIT University, Vellore-14.
[2]SCSE, VIT University, Vellore-14.
Emails: kkarthikeyan67@yahoo.co.in; Sayantani.basu2014@vit.ac.in

Received on 28-10-2015 *Accepted on 25-11-2015*

Abstract

In this article we introduce an efficient algorithm to predict magnitude and damage of earthquake by using the techniques of data mining (multi-linear regression) and neural networks (backpropagation network).Here we consider the data sets of six earthquake-prone regions of the world. The parameters considered in both techniques are month, year, latitude and longitude and moment magnitude as dependent variable. The neural network is a three-layer perceptron that uses back propagation method and the moment magnitudes have been converted to its nearest-integer binary equivalent to facilitate damage prediction of future earthquakes based on the Mercalli intensity scale by using C++ program.

Keywords: Earthquake prediction model, neural network, multi-linear regression, data mining, disaster prediction

I. Introduction

Earthquakes have become an everyday occurrence in today’s world, sometimes causing immense damage and loss of life and property. Since earthquakes are a natural disaster, they are thought to be erratic and sudden. Several attempts by researchers and seismologists have been made in order to perfect the prediction mechanism of natural disasters like earthquakes. However, no user-friendly earthquake prediction models are available which are accessible to people and are easily operable. Earthquakes usually occur along faults, where the Earth’s crust is broken. They are usually characterized by massive trembling, the more massive ones leading to shifting of building foundations, collapse of high rises and a large number of casualties. The prediction of the moment magnitude and damage to be caused by such earthquakes beforehand can minimize casualties as well as economic losses. In this article, we introduce new technique for prediction of the future occurrence of earthquake which includes Multiple Linear Regression (MLR) and Artificial Neural Networks (ANNs). An accurate prediction model can be built by considering the optimum and appropriate factors. The factors considered for this article are the month, year, latitude and longitude. Six earthquake-prone regions of the world have been studied for this purpose: i) Honshu, Japan ii) Kuril Islands, Russia iii) Nicobar Islands, India, iv) Solomon Islands, v) Sumatra, Indonesia and vi) Tonga Islands. The above analysis methods were applied to the earthquake datasets of each of the mentioned regions. The output or dependent variable in MLR is the moment magnitude. A Multi-Layer Perceptron (MLP) that learns by online back propagation method was the neural network implemented for damage prediction using the Mercalli intensity scale. Finally, the regression equations and neural network codes were implemented in a working prototype in the form of a user-friendly C++ program.

II. Related Works

Time series magnitude data and Seismic Electric Signals (SES) have been used by scientists to predict[1] the magnitude of the earthquake on the following day using artificial neural networks with an efficiency of 80.55% for all seismic events. The *b*-value, the Bath's law, and the Omori–Utsu's law as parameters are used to predict seismic activity in Chile [2]. Also, the time interval between two large successive earthquakes is proportional to the coseismal displacement of the previous earthquake [3]. Therefore, the frequency of earthquake occurrence, or the number of earthquakes occurring in a particular area in a year is an important parameter to be considered for future earthquake prediction. Regression trees have been used to find patterns that simulate seismic temporal data to help in the prediction of earthquakes [4]. A regression model is used to predict the occurrence and nature of velocity pulses in future ground motions [5]. Hence, the location of each earthquake is also an important parameter for the prediction of future earthquakes. Real-time[6] prediction models based on Back propagation Networks and Multiple Regression have been made for predicting debris-flow disasters in Taiwan. A considerable amount of research work [7] has also been done with regard to earthquake prediction using the M8 algorithm, which predicts the occurrence of future earthquakes with magnitude 8.0 and above.

In view of the previous work done in this area it is observed that attempts have been made to introduce future earthquake prediction models. The following highlights are the salient features of the present prediction model:

(i) It is robust and user-friendly prediction model.

(ii) It is a versatile model and can be applied to any disaster and any place.

(iii) It considers the parameters month, year, latitude and longitude of earthquake occurrence of a region. No experiment-based parameters need to be entered by the user.

III. Methodology

In this article we use Data mining and Artificial Neural Networks (online back propagation network) to obtain an optimal prediction model. It is built based on pattern recognition in earthquake data occurring for the last 10 years. The factors that are selected for prediction of future earthquakes are: (1) month, (2) year, (3) latitude and (4) longitude. The month and year are considered in order to study the frequency of earthquake occurrence in a particular area. Similarly, the latitude and longitude give an accurate idea about the exact location of earthquake occurrence. The six datasets of 10 years of the following places were studied: i) Honshu, Japan, ii) Kuril Islands, Russia, iii) Nicobar Islands, India, iv) Solomon Islands, v) Sumatra, Indonesia and vi) Tonga Islands. The datasets were obtained from the *'Global Centroid Moment Tensor CMT Catalog'*.

The factors month, year, latitude and longitude were considered as the independent variables for multiple linear regression and the moment magnitude was considered as the dependent variable. The six equations for the above mentioned regions were calculated using the *'Weka Data Mining Suite'*. The moment magnitude predicted for future earthquakes comprises the *'Magnitude Prediction'* component of this analysis. Next, six artificial neural networks (ANNs) were built that learn using online back propagation method.

The networks each consist of 4 inputs, 1 hidden layer and 4 outputs, with a learning rate of $\alpha = 0.7$ and a momentum of $\beta = 0.7$. The epoch was approximately 10,000 for each neural network. The inputs were taken as the month, year, latitude and longitude, while the output consisted of the nearest-whole number 4-bit binary equivalent of the moment magnitude in each dataset. These binary equivalents were generated using a C++ algorithm. The inputs were then fed to each ANN.

The output was in the form of a 4-bit binary code. When converted to base 10, it gave the damage prediction, that is, the level of damage caused due to the earthquake. The networks were then built and the corresponding codes were generated using the *'Multiple Back-Propagation Software'*. The multi-linear regression equations as well as the back propagation codes generated for the six regions were then implemented into a simple, efficient and user-friendly C++ program. This program accepts as user-input the future place of prediction, month, year, latitude and longitude and outputs the magnitude prediction (using data mining) and damage prediction (using neural networks). The damage prediction analysis has been obtained from the *Mercalli intensity scale*, which describes the effects caused due to an earthquake of a certain magnitude.

IV. Earthquake Prediction Model

This research involved Data Mining (Multiple Linear Regression) and Neural Networks (back propagation network) techniques for building the earthquake prediction model. The related theories are discussed as follows:

I. Multiple Linear Regression Method

Let Y be the dependent variable which will predict the future moment magnitude. Y is a linear function of the independent variable x_i (i=1,2,…,k). The multiple linear regression model can be written in the following matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_n \end{bmatrix} \quad (1)$$

Writing the above matrix form in terms of linear equations,

$$Y_1 = B_0 + B_1 x_{11} + B_2 x_{12} + \cdots + B_n x_{1n}$$

$$Y_2 = B_0 + B_1 x_{21} + B_2 x_{22} + \cdots + B_n x_{2n}$$

$$Y_n = B_0 + B_1 x_{n1} + B_2 x_{n2} + \cdots + B_n x_{kn}$$

The above equation can be written in the following form:

$$\sum_{i=1}^{n} Y = nB_0 + B_1 \sum_{i=1}^{n} x_{i1} + B_2 \sum_{i=1}^{n} x_{i2} + \cdots + B_n \sum_{i=1}^{n} x_{in}$$

$$Y = nB_0 + B_1 x_1 + B_2 x_2 + \cdots + B_n x_n \quad (2)$$

II. Back propagation Network Method

Artificial neural networks (ANNs) are a category of Machine Learning algorithms inspired by biological neural networks. They are used to recognize and estimate patterns in a given dataset depending on a large number of inputs and classify unknown data accordingly. *Figure 2* shows the ANN for our earthquake prediction model. The shapes are comparable to 'neurons' while the connecting lines are comparable to 'synapses'. In the present model, the four squares comprise the input layer, the circle in the middle indicates the hidden layer and the four circles comprise the output layer.

The synapses take values from the input layer, multiply it by a specific weight and transfer it to the next layer by using a technique called 'forward propagation'. Finally, the neurons in the output layer add together outputs from all the synapses and apply an activation function (the sigmoid activation function, the graph of which is shown in Figure 1):

$f(x) = \frac{1}{1+e^{-kx}}$ (3)

Online back propagation algorithm is a method of supervised learning for training a class of neural networks called multilayer perceptron (MLPs). Back propagation algorithm aims at minimizing the sum of squared errors using batch gradient descent in MLPs. This model can then be used for making valuable predictions on new data.

Algorithm for back propagation for (Three-layer) Multilayer Perceptron

1. Initialize the network weights with small values.
2. Feed forward the input data through the neural network and generate the output activations.
3. Back propagate the output activations through the neural network using the training pattern target and calculate the difference between the input and output values for all the output and hidden neurons.
4. Multiply the output delta value obtained in Step 3 with the input activation to compute the gradient of each weight. The sign of the gradient indicates the direction of increase of errors and is minimized by updating the weights in the opposite direction.
5. Subtract the learning rate of the gradient from the weight
6. Update the weights.
7. Repeat steps 2-6 until the error is minimized as much as possible and the performance of the network is reasonable.

This algorithm is called the '*back* propagation algorithm' since the computed errors are propagated backward, starting from the output layer to the hidden layer, then from the hidden layer to the input layer. The related flowchart is shown in Figure 3.In order to minimize the errors in the neural network, a gradient descent algorithm is applied. This method helps to find the appropriate weights so that the network error is minimized.

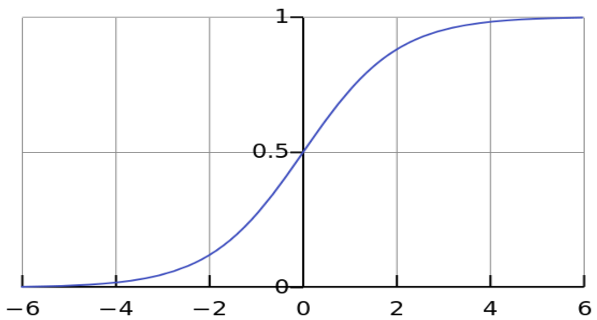

Figure 1: Sigmoid Activation Function.

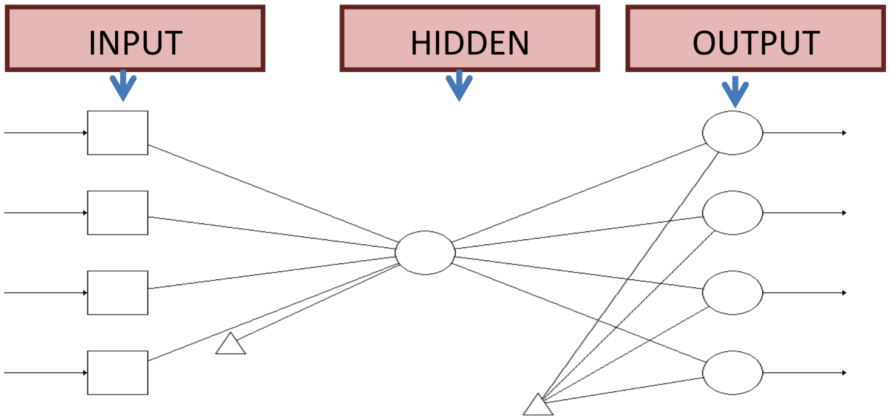

Figure 2: The neural network implemented using back propagation.

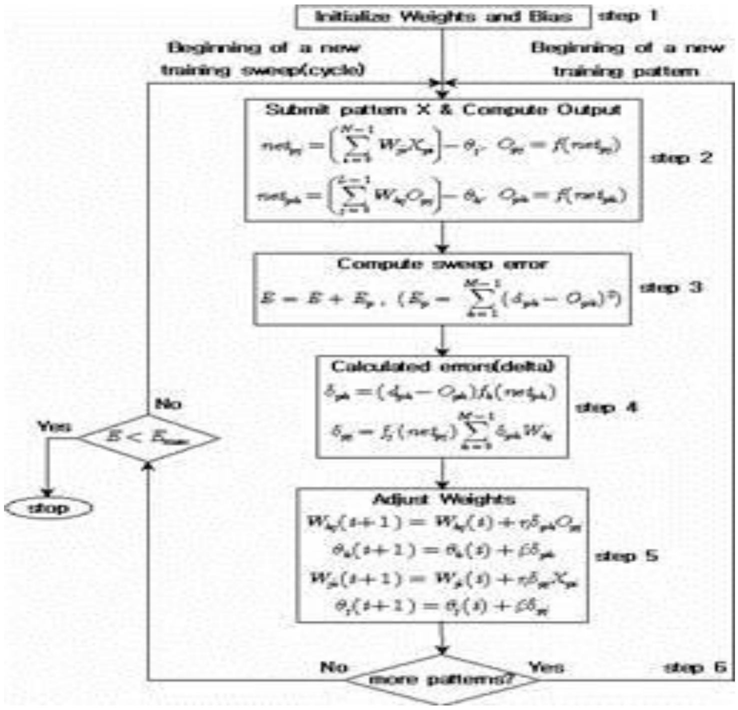

Figure 3: The basic back propagation flowchart.

Magnitude	Typical Maximum Modified Mercalli Intensity
Under 2.0	I
2.0 – 2.9	II – III
3.0 – 3.9	III – IV
4.0 – 4.9	IV – V
5.0 – 5.9	V – VI
6.0 – 6.9	VI – VII
7.0 – 7.9	VII – VIII
8.0 or higher	VIII or higher

V. Moderate	Felt by nearly everyone; many awakened. Some dishes, windows broken. Unstable objects overturned. Pendulum clocks may stop.
VI. Strong	Felt by all, many frightened. Some heavy furniture moved; a few instances of fallen plaster. Damage slight.
VII. Very Strong	Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable damage in poorly built or badly designed structures; some chimneys broken.
VIII. Severe	Damage slight in specially designed structures; considerable damage in ordinary substantial buildings with partial collapse. Damage great in poorly built structures. Fall of chimneys, factory stacks, columns, monuments, walls. Heavy furniture overturned.
IX. Violent	Damage considerable in specially designed structures; well-designed frame structures thrown out of plumb. Damage great in substantial buildings, with partial collapse. Buildings shifted off foundations.

Figure 4: The Mercalli Intensity scale.

V. Pseudo Code

LetN_i,N_h,N_o be the neurons in the input, hidden and output layers respectively. Let W_{ih} be the weights between the input layer and hidden layer and W_{ho} denote the weights of the synapses connecting the hidden layer and the output layer.

for (number of iterations (size of epoch))

for (number of data)

//Feed Forward Step

$$N_h = f(\sum(W_{ih} * N_i))$$

//where $f(x) = \frac{1.0}{1.0+\exp(-x)}$(sigmoid activation function)

$$N_o = f(\sum(W_{ho} * N_h))$$

//Back propagation (output layer to hidden layer)

$$N_{o(error)} = f'(N_o)(N'_o - N_o)$$

$$\Delta W_{ho} = \beta * N_{o(error)} * N_h$$

//where β is the momentum and ΔW_{ho} is the rate of change of W_{ho}

$$W_{ho(new)} = W_{ho} + \Delta W_{ho} + (\alpha * \Delta_t)$$

//where α is the learning rate and Δ_t is the previous change in weight

//Back propagation (hidden layer to input layer)

$$N_{h(error)} = N_{o(error)} * W_{ho}$$

$$\Delta W_{ih} = \beta * N_{h(error)} * N_i$$

$$W_{ih(new)} = W_{ih} + \Delta W_{ih} + (\alpha * \Delta_t)$$

//Weight Update

$$W_{ho} = W_{ho(new)}$$

$$W_{ih} = W_{ih(new)}$$

VI. Observations and Results

The datasets are obtained using the Global *Centroid Moment Tensor Catalog*. The multiple regression analysis was done using *Weka Data Mining Suite*. The six regression equations obtained from the six different regions are given below:

i) Honshu, Japan: magnitude =(-0.0012 * month) + (-0.0028 * year) + (0.0091 * latitude) + (-0.0157 * longitude) +13.227

ii) Kuril Islands, Russia: magnitude= (-0.0091 * month) + (-0.0204 * year) + (0.0713 * latitude) + (-0.0249 * longitude) +47.32

iii) Nicobar Islands, India: magnitude =(0.0127 * month) + (0.0153 * year) + (-0.1171 * latitude) + (-0.1965 * longitude) -5.8037

(iv) *Solomon Islands:* magnitude =(-0.0175 * month) + (-0.0059 * year) + (-0.0048 * latitude) + (0.0263 * longitude) +13.6187

(v) *Sumatra, Indonesia:* magnitude =(0.0032 * month) + (-0.0043 * year) + (-0.016 * latitude) + (-0.0224 * longitude) + 16.4973

(vi) *Tonga Islands:* magnitude = (0.0001 * month) + (0.0046 * year) + (0.0143 * latitude) + (-0.0253 * longitude) -7.6855

The analysis showed that different places have different regression equations. It was also observed that Sumatra, Indonesia has the highest number of earthquakes (OR highest frequency of earthquakes) within a span of 10 years. The graphs showing moment magnitude for future years (till 2025) is shown in *Figure 5*. The values of latitude and longitude were considered as an average of their respective ranges. The graphs were obtained using *Microsoft Excel 2010*. According to this prediction model, it is observed that, in the future, the magnitude of earthquakes will increase in Nicobar Islands, India and Tonga Islands while there will be a slight decrease in the four other regions. Also, a steady increase or decrease is observed in the magnitude in each case. The back propagation network analysis was done using *Mutiple Back-Propagation Software*. The neural network implemented is shown in Figure 2. The codes generated after a learning epoch of approximately 10000 for each dataset were then used for the damage prediction component of this study. The final C++ code implementation shown in *Figure 4* gives an output of the moment magnitude as well as damage prediction using the Mercalli scale. The user will first have to enter the choice for the place and then enter the month and the year of the prediction. Then the latitude and longitude have to be entered. To make it simple and user-friendly, the ranges for the latitude and longitude have been mentioned.

The program finally outputs the 'Predicted Moment Magnitude using Linear Regression' and the 'Damage Prediction using Back Propagation Neural Networks' based on the data entered by the user.

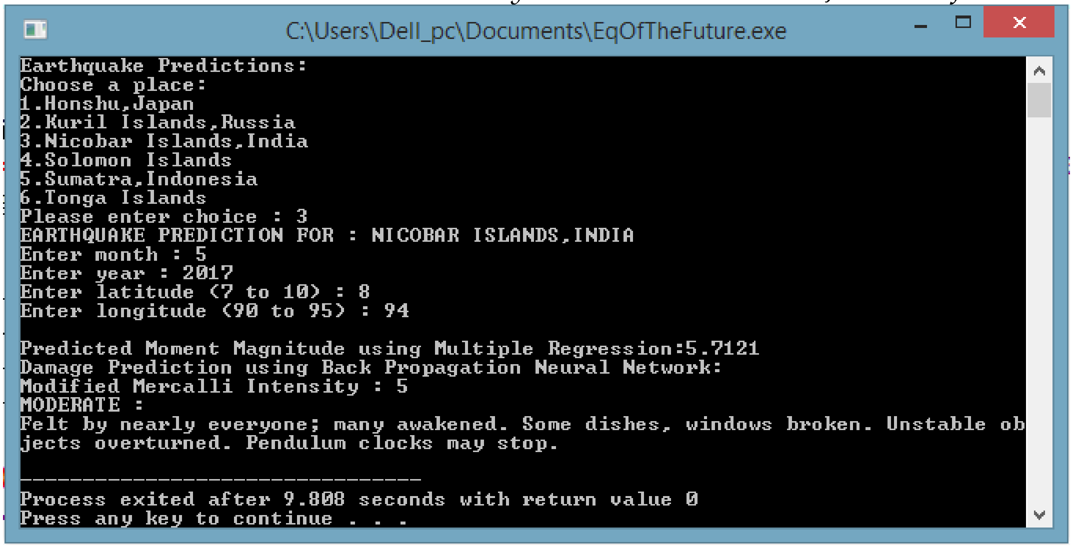

Figure 4: The final C++ prototype output.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.70295	5.70015	5.69735	5.69455	5.69175	5.68895	5.68615	5.68335	5.68055	5.67775	5.67495	5.67215	5.66935	5.66655	5.66375	5.66095	5.65815	5.65535	5.65255	5.64975	5.64695

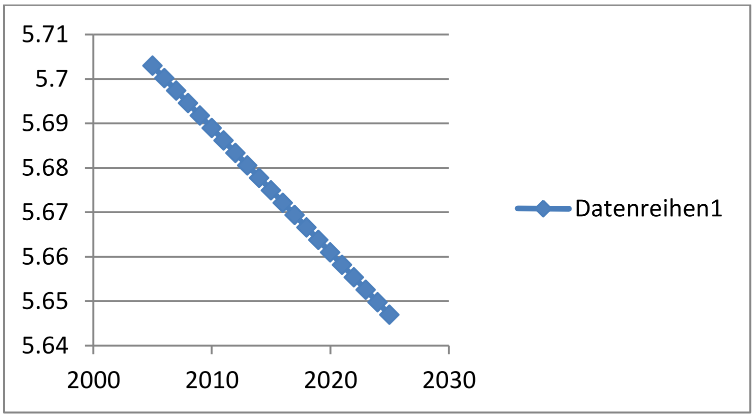

Figure 5(a): Future moment magnitude prediction for Honshu, Japan.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.7701	5.7497	5.7293	5.7089	5.6885	5.6681	5.6477	5.6273	5.6069	5.5865	5.5661	5.5457	5.5253	5.5049	5.4845	5.4641	5.4437	5.4233	5.4029	5.3825	5.3621

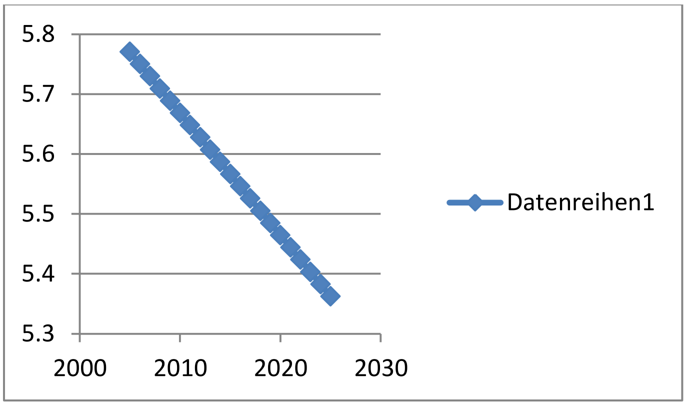

Figure 5(b): Future moment magnitude prediction for Kuril Islands, Russiak.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.78375	5.79905	5.81435	5.82965	5.84495	5.86025	5.87555	5.89085	5.90615	5.92145	5.93675	5.95205	5.96735	5.98265	5.99795	6.01325	6.02855	6.04385	6.05915	6.07445	6.08975

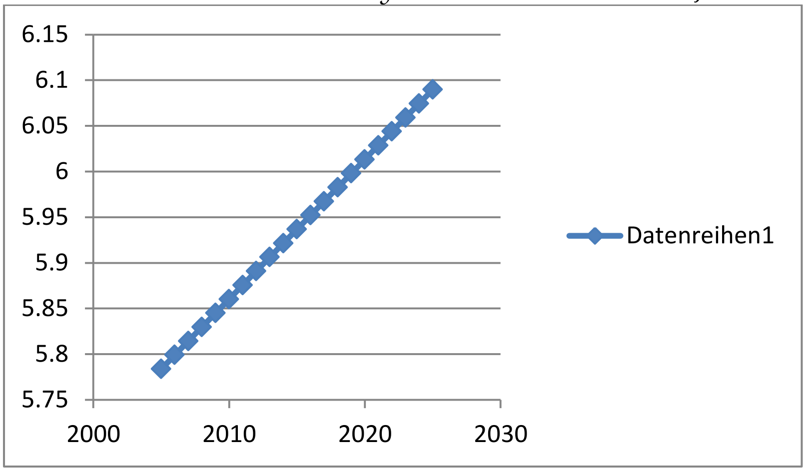

Figure 5(c): Future moment magnitude prediction for Nicobar Islands, India.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.8585	5.8526	5.8467	5.8408	5.8349	5.829	5.8231	5.8172	5.8113	5.8054	5.7995	5.7936	5.7877	5.7818	5.7759	5.77	5.7641	5.7582	5.7523	5.7464	5.7405

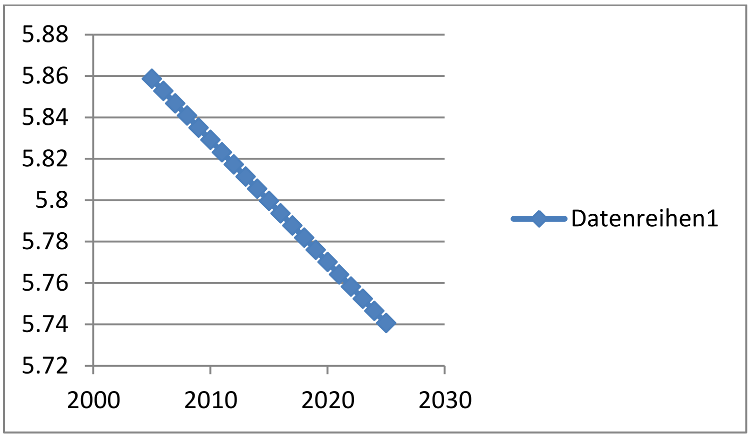

Figure 5(d): Future moment magnitude prediction for Solomon Islands.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.7126	5.7083	5.704	5.6997	5.6954	5.6911	5.6868	5.6825	5.6782	5.6739	5.6696	5.6653	5.661	5.6567	5.6524	5.6481	5.6438	5.6395	5.6352	5.6309	5.6266

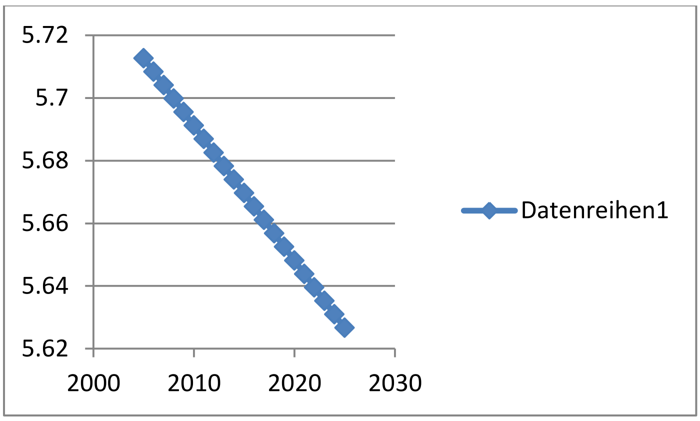

Figure 5(e): Future moment magnitude prediction for Sumatra, Indonesia.

2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
5.60925	5.61385	5.61845	5.62305	5.62765	5.63225	5.63685	5.64145	5.64605	5.65065	5.65525	5.65985	5.66445	5.66905	5.67365	5.67825	5.68285	5.68745	5.69205	5.69665	5.70125

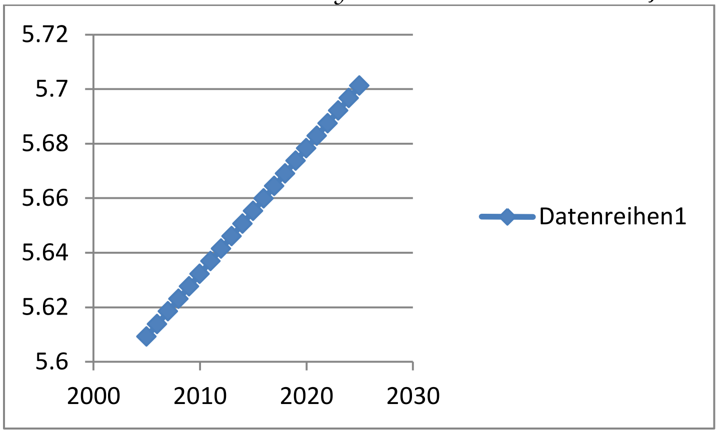

Figure 5(f): Future moment magnitude prediction for Tonga Islands.

VII. Conclusion

In this article, we have shown that the model has an accuracy of 93% on tested datasets which was obtained by training the data for an epoch of ten thousand iterations and this model can accurately and successfully be extended to any region as well as any disaster, given the correct parameters. Back propagation proved to be a more accurate predictor compared to multiple linear regression. Earthquakes have a unique pattern of occurrence in every region. This model can help in early evacuation according to future earthquake predictions, as well as prevent casualties, loss of lives and economic losses that occur due to earthquakes happening worldwide every year.

Acknowledgement

We would like to acknowledge the 'Global Centroid Moment Catalog' for their dataset result and thank the Jet Propulsion Laboratory for the back propagation flowchart (*Figure 3*).

References

1. Moustra, Avraamides, Christodoulou (2011): Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals**,** *Expert Systems with Applications, 38(12), 15032–15039.*

2. Reyes, Morales-Esteban, Martínez-Álvarez (2013): Neural networks to predict earthquakes in Chile, *Applied Soft Computing, 13(2), 1314–1328.*

3. Shimazaki, Nakata (1980): Time-predictable Recurrence Model for Large Earthquakes, *Geophysical Research Letters, 7(4), 279-282.*

4. Martínez-Álvarez**,** Troncoso, Morales-Esteban, Riquelme (2011): Computational Intelligence Techniques for Predicting Earthquakes, *Hybrid Artificial Intelligence Systems, 6679: 287-294.*

5. Shahi, Baker (2011): Regression models for predicting the probability of near-fault earthquake ground motion pulses, and their period, *Baker Research Group, Stanford University.*

6. Kung, Chen, Ku (2012): Designing intelligent disaster prediction models and systems for debris-flow disasters in Taiwan, *Expert Systems with Applications, 39(5), 5838–5856.*

7. Molchan (2010): Earthquake Prediction Analysis: The M8 algorithm, *Geophysical Journal International, 183(3), 1525-1537.*

8. Lopes, N. and Ribeiro, B. (2010): A Strategy for Dealing with Missing Values by Using Selective Activation Neurons in a Multi-Topology Framework. *IEEE World Congress on Computational Intelligence WCCI 2010.*

9. Lopes, N. and Ribeiro, B. (2009): GPU Implementation of the Multiple Back-Propagation Algorithm, *in proceedings of Intelligent Data Engineering and Automated Learning, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, volume 5788, pages 449-456.*

10. Lopes, N. and Ribeiro, B. (2003): An Efficient Gradient-Based Learning Algorithm Applied to Neural Networks with Selective Actuation Neurons, *in Neural, Parallel and Scientific Computations, volume 11, pages 253-272, Dynamic Publishers.*

11. Lopes, N. and Ribeiro, B. (2001): Hybrid learning in a multi neural network architechture, *in INNS-IEEE International Joint Conference on Neural Networks, ICJNN, volume 4, pages 2788-2793, Washington D.C., USA.*

12. Dziewonski, A. M., T.-A. Chou and J. H. Woodhouse, Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.*, 86, 2825-2852, 1981.doi:10.1029/JB086iB04p02825.

13. Ekström, G., M. Nettles, and A. M. Dziewonski, The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, 200-201, 1-9, 2012.doi:10.1016/j.pepi.2012.04.002.

Corresponding Author:

K. Karthikeyan*,

Emails: kkarthikeyan67@yahoo.co.in