❒   1304

Earthquake trend prediction using long short-term memory RNN

Vardaan K.[1], Tanvi Bhandarkar[2], Nikhil Satish[3], S. Sridhar[4], R. Sivakumar[5], Snehasish Ghosh[6]

[1,2,3,4]Department of Computer Science and Engineering, SRM Institute of Science and Technology, India
[5,6]Earthquake Research Cell, Department of Civil Engineering, SRM Institute of Science and Technology, India

Article Info	ABSTRACT
Article history: Received Sep 25, 2017 Revised Oct 15, 2018 Accepted Nov 30, 2018 *Keywords:* Artificial neural network Earthquake Feed forward neural network Long short-term memory Recurrent neural network	The prediction of a natural calamity such as earthquakes has been an area of interest for a long time but accurate results in earthquake forecasting have evaded scientists, even leading some to deem it intrinsically impossible to forecast them accurately. In this paper an attempt to forecast earthquakes and trends using a data of a series of past earthquakes. A type of recurrent neural network called Long Short-Term Memory (LSTM) is used to model the sequence of earthquakes. The trained model is then used to predict the future trend of earthquakes. An ordinary Feed Forward Neural Network (FFNN) solution for the same problem was done for comparison. The LSTM neural network was found to outperform the FFNN. The R^2 score of the LSTM is better than the FFNN's by 59%.

Corresponding Author:

Vardaan Kishore,
Department of Computer Science Engineering,
SRM Institute of Science and Technology,
4046, estancia tower-4, Gudvancherry, Chennai, Pin-603203, India.
Email: vardaan_kishore@srmuniv.edu.in

1.   INTRODUCTION

Earthquakes are a natural hazard which can cause a lot of damage and loss to human life. It does not follow any set of patterns in occurence and thus predicting the trend has always been an important area of research. A fault in geology is a fracture in rock across which there has been significant rock mass movement. Faults are created by the action of plate tectonic forces. The energy release caused by rapid movement of plates across certain faults can be called active faults and is the most common cause of earthquakes. This energy travels to the surface of the Earth as waves. There are three kinds of seismic waves. S and P waves penetrate the interior of the earth and are hence called body waves [1]. The third kind of waves and the most destructive are surface waves, which are similar to water and travel across the surface of the earth.

Due to its destructive potential, humankind has long been searching for an earthquake trend prediction method. Predicting an earthquake implies stating the exact time, magnitude and location of a coming earthquake. Prediction models come under either short-term prediction (<1-year time scale), long term prediction (10 to 100 years time scale) or intermediate term prediction (1 to 10 years time scale) [2], [3]. Great effort has been made by the scientific community but due to the intrinsic random nature of the phenomenon itself, no valid and reliable method has yet been found. Nevertheless, earthquakes generation is not a cyclical process due to the variation of rupture area and earthquake-mediated interactions along other faults. This means that the time between events can be extremely irregular. Consequently, the prediction of the time, or a relatively close time interval, of an oncoming large earthquake is still a difficult task. Although considerable research is devoted to the science of short-term earthquake forecasting, standardisation of

operational procedures is in a nascent stage of development. The problem is challenging because large earthquakes cannot be reliably predicted for specific regions over time scales that span less than decades.

There are two general approaches to predict earthquakes, precursors based and trend based. Precursors are anomalous phenomena that might signal an impending earthquake [4] such as radon gas emissions, unusual animal behaviour, electromagnetic anomalies etc. Trend based methods involve identifying patterns of seismicity that precede an earthquake. In this paper, a trends-based approach is adopted and the LSTM neural network is used to capture the trend involving statistical techniques.

The relationship between the maximum of earthquake affecting coefficient and site and basement condition was studied, also proposed a model based on earthquake magnitude prediction using artificial neural network in the northern read sea area [5]. A multilayer using compression data for precursor detection in electromagnetic wave observation was proposed [6]. A time series approach composed of seismic events occurred in Greece was applied [7]. A study between radon and earthquake using an artificial neural networks model was done [8]. A relationship between radon concentration and environmental parameters for earthquake prediction was modelled using an ANN in the region of Thailand [9]. A neural network for classification after analysing the electric field data and the seismicity collected from different stations was studied and results were pretty accurate [10]. Investigated the seismic damage identification by using a PCA-compressed response function and artificial neural networks [11]. Prediction of earthquake damages and reliability analysis using fuzzy sets [12]. The variation of Total Electron Content (TEC) as an anomaly as an indication of earthquake a few days or hours before it, this was used by them to build a model [13]. Recursive sample-entropy technique for earthquake forecasting, where the earth data based on VAN method was used for the modelling [14]. Models based on measurement of elastic and electromagnetic waves to predict earthquakes and tsunami was done [15]. Earthquake hazard assessment was done using EaHaAsTo tool for visualization [16]. Determined the threshold energy leading to seismic activity [17].

2.   RESEARCH METHOD

An artificial neural network is a mathematical model that mimics the biological neurons in brain. A neural network is a set of input, output and hidden layers. These layers have nodes which are interconnected through links. These links have some associated numeric weight which determines how much the input contributes to and affects the results. The weights and activation functions can be modified by a process called learning which is governed by a learning rule [18]. In this paper we have compared the structures of Feed Forward Neural Network (FFNN) and Recurrent Neural Network (RNN) on time-series based data.

2.1. Feed forward neural network

Feedforward networks are acyclic network usually arranged in layers, where each neuron receives inputs only from the immediately preceding layer. The architecture of an FFNN with 2 hidden layers is shown in Figure 1.
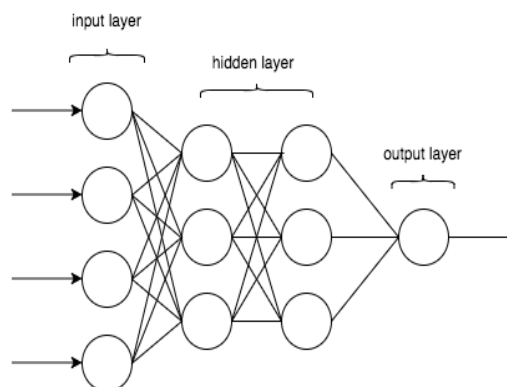

Figure 1. Feed Forward Neural Network with input layer, output layer and 2 hidden layers

FFNNs are still successfully applied to many problems but still cannot capture long term dependencies. Many models have implicitly captured time by concatenating each input with some number of its immediate predecessors and successors, presenting the machine learning model with a sliding window of

context about each point of interest [19]. This approach is followed by concatenating a fixed number (fourteen) of past earthquake data and giving them as input to the FFNM with the next earthquake as the target. The model uses two hidden layers with 20 nodes and 60 nodes respectively. All the nodes having the sigmoid function as their activation. The learning rule used it the 'rmsprop'. This model is trained for 1000 epochs on the dataset. These attributes were selected after employing a grid search method which selected the best architecture based on the error rate.

2.2. Long short-term memory

Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) architecture (an artificial neural network) proposed by Sepp Hochreiter and Jürgen Schmidhuber in 1997 [20]. RNNs can capture the dynamics of sequences via cycles in the network. But some RNNs suffer from the vanishing and exploding gradients problem in which gradients are either squashed to zero or increase without bound during back propagation through a large number of time steps. LSTM is introduced primarily to overcome the problem of vanishing gradients. It has chain like structure, having three or four neural network layer or "gates" which are implemented using logistic function.

The information given in [19] depicts about the forward pass and backward pass in LSTMs. In terms of the forward pass, the LSTM can learn when to let activation into the internal state. As long as the input gate takes value zero, no activation can get in. Similarly, the output gate learns when to let the value out. When both gates are closed, the activation is trapped in memory cell, neither growing nor shrinking. In terms of backwards pass, the constant error carousel enables the gradient to propagate back across many time steps, neither exploding nor vanishing. Figure 2 shows a chunk of neural network, A, which looks at some input x_t and outputs a value h_t. A loop allows information to be passed from one step of the network to the next.
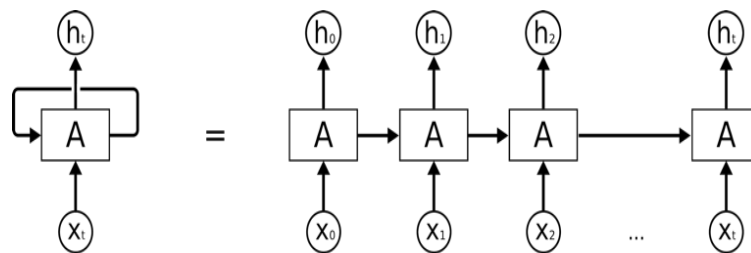

Figure 2. RNN-LSTM architecture. The input of x_t is fed into the memory cell A which gives output h_t

The LSTM RNN model proposed in this study includes two hidden layers with 40 hidden units each that are LSTM cells. The backpropagation through time is limited to 15 steps. A dropout layer is included between the 2 hidden layers for regularisation [21]. It will randomly exclude 30% of the activations of the previous layer from propagating to prevent overfitting. The Root Mean Square (RMS) loss is reduced using the Adagrad algorithm which increases the learning rate for more sparse parameters and decreases the learning rate for less sparse ones. This strategy often improves convergence performance over standard stochastic gradient descent in settings where data is sparse [22]. The initial learning rate is taken to be 7 and is exponentially decreased when the RMS loss does not improve for more than 10 epochs. The training was stopped after the loss started to fluctuate despite very low learning rate. The number of epochs came to be 1600. These parameters were selected after trying out other architectures.

3. RESULTS AND ANALYSIS
3.1. Data exploration

Number of occurrences of earthquakes recorded over various regions is shown in Figure 3. The Afghanistan - Tajikistan region recorded close to 5000 earthquakes being the highest. The Lakshadweep region recorded the lowest number of earthquakes. The distribution of count of earthquakes based on the magnitude recorded on each earthquake is shown in Figure 4. Earthquakes of magnitude 3.4 on the Richter scale were highest in amount being close to 45 occurrences whereas that of higher range close to 7.9 are low in occurrences. Figure 5 shows the average magnitude of earthquakes in each region. The western xizang-india border has the highest average magnitude whereas Thailand and North-Eastern India has the lowest. Figure 6 shows the average depth of each earthquake recorded in kilometres. An earthquake of 140 km is the

deepest recorded earthquake, occurred in the Tajikistan region. The earthquakes recorded in the the Lakshadweep and Laos region were the lowest in depth being very close to the surface.
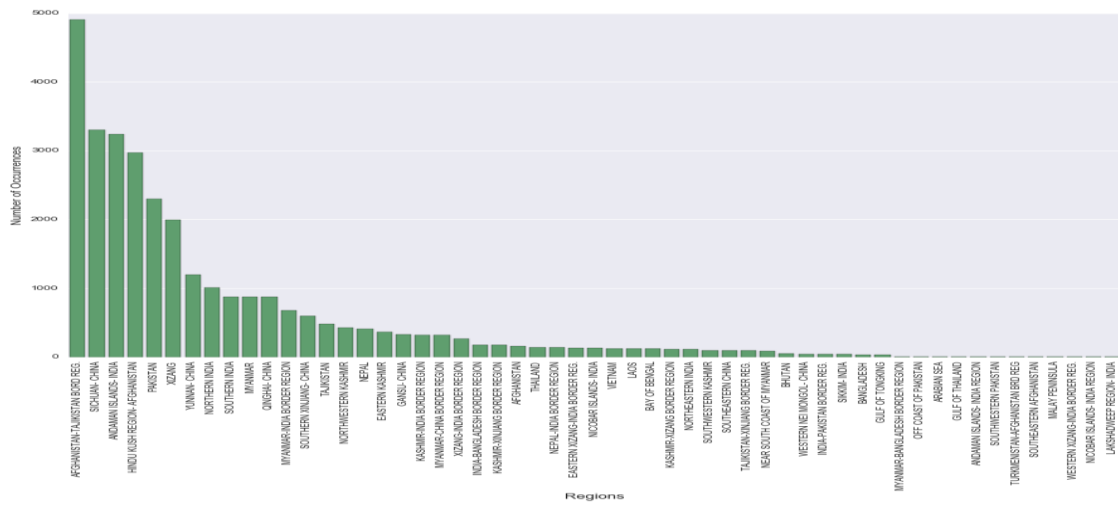

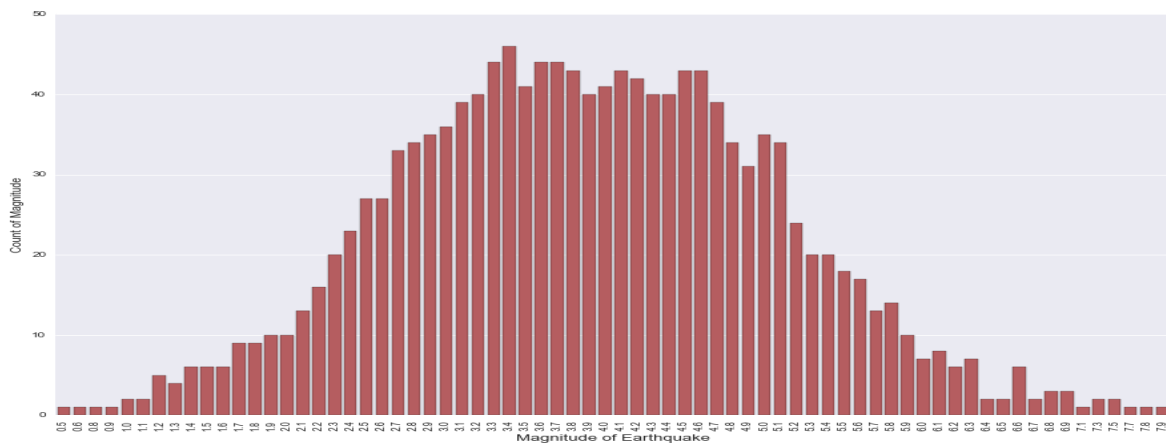Figure 3. Number of occurrences of earthquake based on the regions

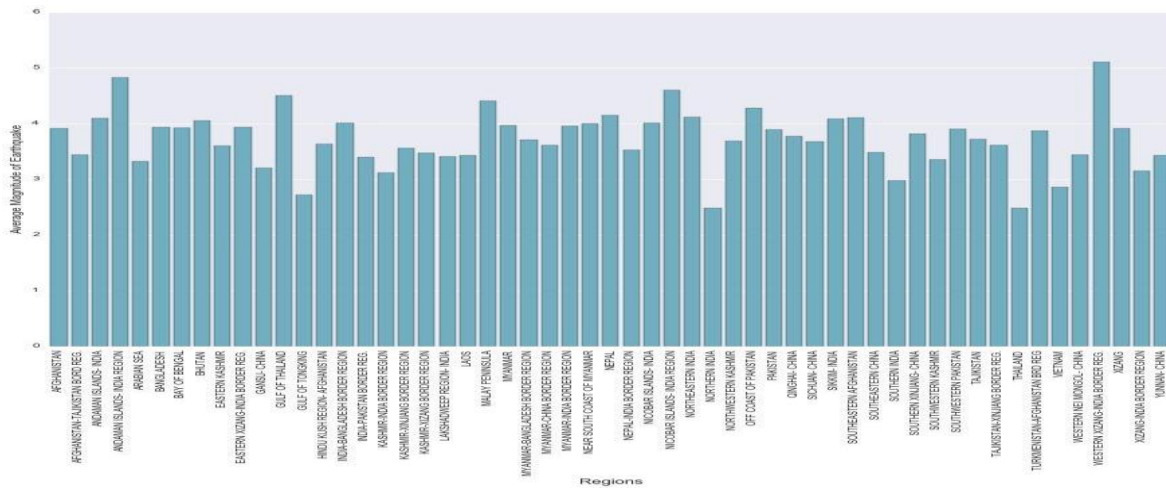Figure 4. Distribution of earthquake magnitudes from low to high

Figure 5. Average magnitude of earthquake in each region

Figure 6. Average depth of occurring earthquake

3.2. Experimental Results
3.2.1. Feed forward neural network

Figure 7 shows the Mean Square Error after every epoch for FFNN. As the model converges, the error becomes static and achieves highest value showing its inability to model sequence data. Figure 8 shows the R^2 score for each variable used for prediction and are negative which indicates that the FFNM is not able to capture the chaotic nature of the attributes. Figure 9 shows the original data set plotted on map. Figure 10 shows the predicted 6000 earthquakes by the FFNN plotted on the map, as seen the region is clustered showing a weak prediction trend. Figure 11 shows the region wise distribution of earthquakes of the earthquakes predicted by the FFNN, as observed the data is skewed and it shows the Xizang region will have more than 7000 earthquakes in the prediction trend.

Figure 7. MSE after every epoch for FFNN

Figure 8. R^2 Score for each variable at the end of training of FFNN

Figure 9. Given data plotted on a map

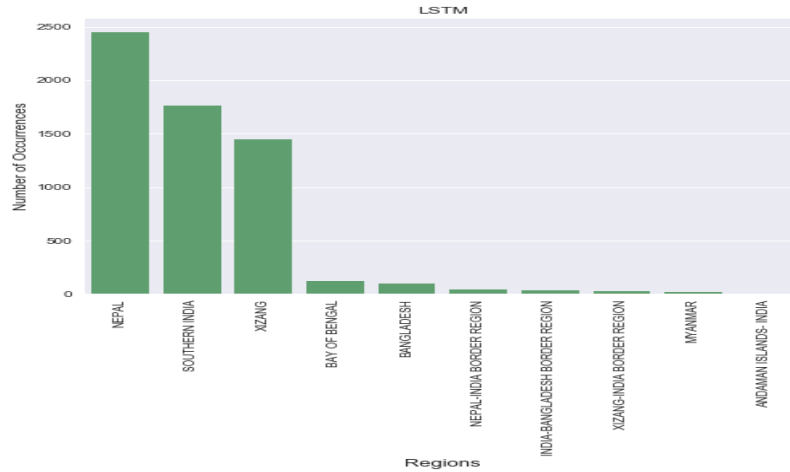Figure 10. Next 8988 FFNN predicted earthquakes

Figure 11. Region wise distribution of predicted earthquake by FFNN

3.2.2. Long short-term memory

Figure 12 shows the MSE after every epoch for LSTM. As we can see the MSE converges at a point lower than that of the FFNN model, this show that the model is converging and the error rate is low as well. Figure 13 shows the R^2 score of every variable on the test set. The scores are all positive. There is a drastic improvement in the score for Timestamp and this pulls up the overall score to -0.252 which is 59% more than the score for FFNN but is still negative. This indicates that though explicitly capturing sequence information by using LSTMs leads to better results than traditional neural networks. The R^2 score is used to evaluate our models. R^2 (coefficient of determination) regression score function. Best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse). A constant model that always predicts the expected value of y, disregarding the input features, would get a R^2 score of 0.0. R^2 scores are considered to be better than the MSE because it is scaled between 0-1, where as MSE is not scaled to any particular values. R^2 can be interpreted more easily. Table 1 gives the comparison of the R^2 scores between the Feed Forward Neural Network and Long Short-Term Memory RNN.
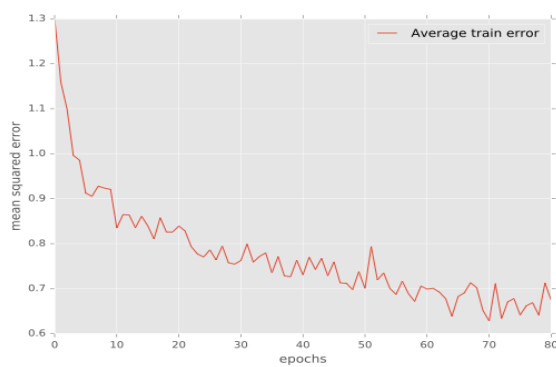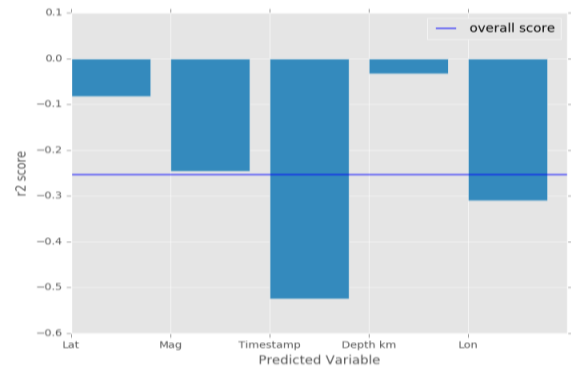

Figure 12. MSE after every epoch for LSTM

Figure 13. R^2 score for each variable at the end of training of LSTM

Table 1. Comparison of R^2 Score

Attribute	FFNN	LSTM
Overall	-0.6188	-0.252
TimeStamp	-4.677	-0.524
Depth	-0.201	-0.032
Magnitude	-0.004	-0.246
Longitude	-0.040	-0.31
Latitude	-0.029	-0.083

By plotting the earthquakes predicted from the data set on a map across epochs we can get a feel of how the data is spread. Figure 14 shows the Future 6000 predicted earthquakes on the map; the spread-out points show that the LSTM is better than the FFNN in prediction. Figure 15 shows the region wise distribution of the earthquakes predicted by the LSTM network. Nepal region is predicted to have the highest number of earthquakes in the future.
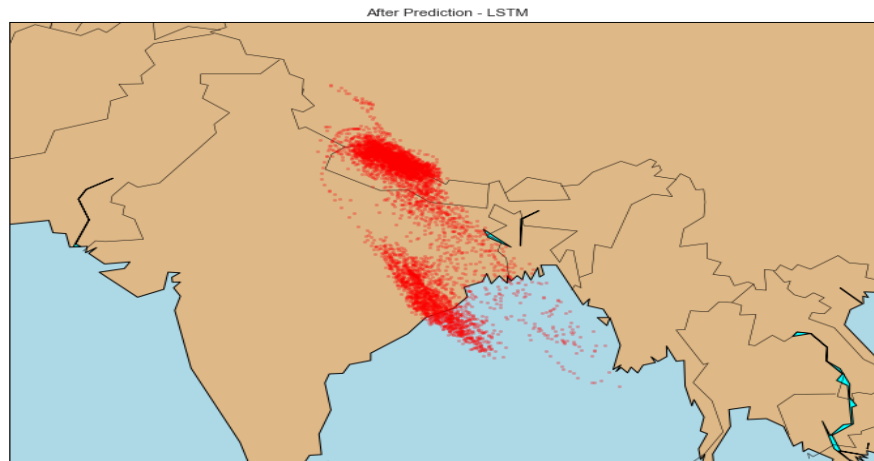

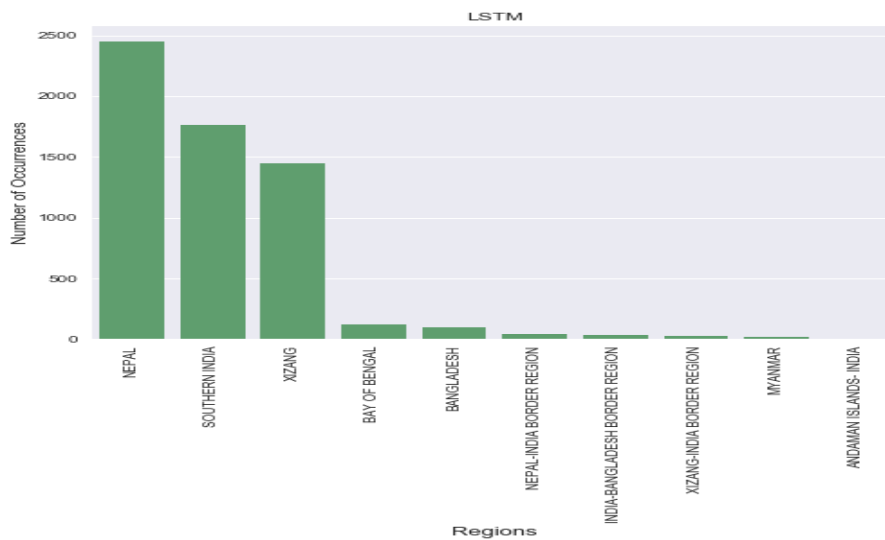Figure 14. Map of Next 6001 predicted earthquake of LSTM

Figure 15. Region wise distribution of earthquake from LSTM prediction

Table 2 shows comparison of future predicted earthquakes of both models against the original dataset, it can be observed that the FFNN shows a large number of earthquakes in Xizang and totally zero earthquakes for Southern India, this indicates that the model is weak. LSTM shows a changing earthquake trend in the future with much more earthquakes along Nepal and South India.

Table 2. Comparison of Future Predicted Earthquakes of Both Models against the Original Dataset

Area	Original Data	FFNN	LSTM
Xizang	4800	7400	1400
Nepal	300	500	2400
Southern India	900	0	1750

4.   CONCLUSION

LSTM model predicts earthquake magnitude, depth, time, latitude and longitude and got significant results that indicate that the present model does indeed capture some trends in the given earthquake data. The results clearly show the superiority of LSTMs over regular FFNMs in the task of modelling the sequence of earthquakes. The R^2 score yielded by LSTM on the test data is 59% higher than the FFNN's. Future work can be done by collecting data of earthquakes in regions like Sikkim in India which are earthquake prone. The present model can be improved by combining different deep learning models for real time use.

REFERENCES

[1]   P. Shearer, "Introduction to Seismology," Cambridge University Press, 1999.
[2]   Masashi Hayakawa, "Earthquakes and EQ Prediction," John Wiley Sons, Singapore Pte. Ltd, pp. 1–17, *2015*.
[3]   Seiya Uyeda, Toshiyasu Nagao, and Masashi Kamo-gawa, "Short-Term Earthquake Prediction: Current Status of Seismo-Electromagnetics," *Tectonophysics,* vol. 470(3), pp. 205 – 213, *2009.*
[4]   R. J. Geller, "Earthquake Prediction: A Critical Review," *Geophysical Journal International,* vol. 131, pp. 425–450, December 1997.
[5]   Zhu Qing-Jie Su, You-Po, "Application of Ann to Prediction of Earthquake Influence," *Second International Conference on Information and Computing Science,* vol. 2, pp. 234 – 237, 2009.
[6]   A. Itai, H. Yasukawa, I. Takumi, and M. Hata, "Multi- Layer Neural Network For Precursor Signal Detection In Electromagnetic Wave Observation Applied to Great Earthquake Prediction," In *NSIP 2005 Abstracts. IEEE-Eurasip Nonlinear Signal and Image Processing, 2005,* pp. 31–, May 2005.
[7]   V. P. Plagianakos and E. Tzanaki, "Chaotic Analysis of Seismic Time Series and Short-Term Forecasting Using Neural Networks," In *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on, volume 3,* vol. 3, pp. 1598–1602, 2001.
[8]   Külahci F., Inceoz M., Doğˇru, E. Aksoy, O. Baykara, "Ann For Earthquake Prediction With Radon Monitoring," *Applied Radiation and Isotopes,* vol. 67, pp. 212 – 219, 2009.
[9]   A. Negarestani, S. Setayeshi, M. Ghannadi-Maragheh, and B. Akashe, "Layered Neural Networks-Based Analysis of Radon Concentration and Environmental Parameters in Earthquake Prediction," *J Environ Radioact,* vol. 62 (3), pp. 225 – 233, 2002.
[10]   Mehmet Sirac Ozerdem, Berk Ustundag, and R. Murat Demirer, "Self-Organized Maps Based Neural Networks For Detection Of Possible Earthquake Precursory Electric Field Patterns," *Advances in Engineering Software*, vol. 37(4), pp. 207 – 217, 2006.
[11]   Y.Q. Ni, X.T. Zhou, and J.M. Ko, "Experimental Investigation of Seismic Damage Identification Using Pca-Compressed Frequency Response Functions And Neural Networks," *Journal of Sound and Vibration*, vol. 290 (1), pp. 242 – 263, 2006.
[12]   Zhong Zhihuan and Yu Junjing, "Prediction of Earthquake Damages and Reliability Analysis Using Fuzzy Sets," In *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pp. 173–176, Dec 1990.
[13]   M. H. Jusoh, N. Ya'acob, H. Saad, A. A. Sulaiman, N. H. Baba, R. A. Awang, and Z. I. Khan, "Earthquake Prediction Technique Based on Gps Dual Frequency System in Equatorial Region," In *2008 IEEE International RF and Microwave Conference*, pp. 372 – 376, Dec 2008.
[14]   Shohei Shimizu, Koichi Sugisaki, and Hiromitsu Ohmori, "Recursive Sample-Entropy Method and Its Application For Complexity Observation Of Earth Current," In *2008 International Conference on Control, Automation and Systems,* pp. 1250 – 1253, Oct 2008.
[15]   G. P. Turmov, V. I. Korochentsev, E. V. Gorodetskaya, A. M. Mironenko, D. V. Kislitsin, and O. A. Starodubtsev, "Forecast Of Underwater Earthquakes With A Great Degree Of Probability," In *Proceedings of the 2000 International Symposium on Underwater Technology* (Cat. No.00EX418), pp. 110 – 115, 2000.
[16]   R. Sivakumar and Snehasish Ghosh, "Earthquake Hazard Assessment through Geospatial Model and Development of Eahaasto Tool for Visualization: An Integrated Geological and Geoinformatics Approach," *Environmental Earth Sciences*, vol. 76 (12), pp. 442, Jun 2017.
[17]   R. Sivakumar and Snehasish Ghosh, "Determination of Threshold Energy for the Development of Seismic Energy Anomaly Model through Integrated Geotectonic and Geoinformatics Approach," *Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, vol. 86 (2), pp.711 – 740, 2017.
[18]   A. Zell, *Simulation neuronaler Netze*. Oldenbourg, 1997.
[19]   Zachary Chase Lipton, "A Critical Review of Recurrent Neural Networks for Sequence Learning," *CoRR*, abs/1506.00019, 2015.
[20]   Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory," *Neural Comput.,* vol. 9 (8), pp. 1735–1780, November 1997.
[21]   Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals, "Recurrent Neural Network Regularization," *CoRR,* abs/1409.2329, 2014.
[22]   John Duchi, Elad Hazan, and Yoram Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *J. Mach. Learn. Res.,* vol. 12, pp. 2121 – 2159, July 2011.

BIOGRAPHIES OF AUTHORS

Tanvi Bhandarkar is pursuing her bachelor degree in Computer Science and Engineering from SRM Institute of Science and technology. Her research interests include applications of intelligent systems, deep learning and pattern recognition.

Vardaan Khanted is an Engineering Student at SRM University, pursuing Computer Science Engineering and is working in the area of Artificial Intelligence and Machine Learning mainly focusing on cross-domains applications of it.

Nikhil is an Engineering Student at SRM University, pursuing Computer Science Engineering and is working in the area of Artificial Intelligence and Machine Learning mainly focusing on cross-domains applications of it.

Shehasish Ghosh is currently the Senior Research Fellow at Earthquake Research Cell at SRM University. His research areas are Remote sensing & GIS, Geodynamic behavior and earthquake energy study, Active fault mapping & classification, Development of GIS customization tool.

Dr. R. Sivakumar is an Associate Professor in the Department of Civil Engineering and has got rich research experience in Remote sensing & GIS. His other research interests are Geological remote sensing, Geoinformatics in urban hydrology, Earthquake studies, Disaster studies & mitigation, Hyperspectral remote sensing, Digital Image processing.

S. S. Sridhar received his Bachelor of Engineering, B.E. (Computer Engineering) from Madurai Kamaraj University, India in 1990 and Master of Science, M.S (Software Systems) from Birla Institute of Technology and Science, Pilani, India in 1995 and Ph. D in Constructive Neural Networks from SRM University. His research interest included Pattern Recognition, Earthquake studies, Artificial Intelligence, Image Processing and Brain Computer Interfacing.