World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:7, 2024

A Machine Learning Approach for Earthquake Prediction in Various Zones Based on Solar Activity

Viacheslav Shkuratskyy, Aminu Bello Usman, Michael O'Dea, Mujeeb Ur Rehman, Saifur Rahman Sabuj

***Abstract*—**This paper examines relationships between solar activity and earthquakes, it applied machine learning techniques: K-nearest neighbour, support vector regression, random forest regression, and long short-term memory network. Data from the SILSO World Data Center, the NOAA National Center, the GOES satellite, NASA OMNIWeb, and the United States Geological Survey were used for the experiment. The 23rd and 24th solar cycles, daily sunspot number, solar wind velocity, proton density, and proton temperature were all included in the dataset. The study also examined sunspots, solar wind, and solar flares, which all reflect solar activity, and earthquake frequency distribution by magnitude and depth. The findings showed that the long short-term memory network model predicts earthquakes more correctly than the other models applied in the study, and solar activity is more likely to effect earthquakes of lower magnitude and shallow depth than earthquakes of magnitude 5.5 or larger with intermediate depth and deep depth

***Keywords*—**K-Nearest Neighbour, Support Vector Regression, Random Forest Regression, Long Short-Term Memory Network, earthquakes, solar activity, sunspot number, solar wind, solar flares.

I. Introduction

An earthquake is characterized by a variety of fundamental factors, such as its depth, hypocentre, and magnitude. The distance between the Earth's surface and 700 kilometres below the surface is the depth of an earthquake. The hypocentre, which designates the beginning of an earthquake, is located in the shallow (0-70 km), intermediate (70-300 km), and deep (300-700 km) zones of this subterranean area. The size of an earthquake is determined by its magnitude. For instance, an earthquake of magnitude 5.3 is regarded as moderate, but an earthquake of magnitude 6.3 is regarded as powerful. Earthquakes are caused by a range of natural and artificial reasons, and they often happen along plate tectonic borders. As shown in Fig. 1, there are two types of earthquake impacts: internal and exterior Earth effects. The first kind of earthquake is caused by tectonic activity or by things that happen inside the earth, like rain, volcanoes, or landslides. The second type of earthquake cause is non-tectonic or external earth effects, such as sun and moon gravitation and solar activity.

While earthquakes occur on the Earth's surface, solar activity events occur on the Sun's surface, with a distance of approximately 1.5×10^{11} m between them [1]. Apparently, there does not appear to be any connection between the sun and earthquakes. Despite this, it still is not known how or how much solar activity affects earthquakes, some studies have shown a link between these two phenomena [2]. Wolf [3] was one of the first to show that these two seemingly unrelated events are linked (earthquake and solar activity). Wolf's assertions are supported by other researches, e.g. [4] and [5], which has shown that earthquakes are affected by solar activity and the 11-year solar cycle, and that the placement of active zones on the Sun affects the frequency of major earthquakes of higher magnitudes ($M \geq 6.5$). There are contrary findings to this as well, for example, [6] and [7] claimed that there is no statistically significant evidence that solar activity events contribute to the occurrence of earthquakes. However, these authors stated that they did not have data to prove that solar activity events do not cause earthquakes. In summary, the exact relationship between solar activity and the occurrence of earthquakes remains uncertain, as there are conflicting findings in the research; some studies support the idea that solar activity influences earthquake frequency and magnitude, while others find no significant evidence to support this claim.

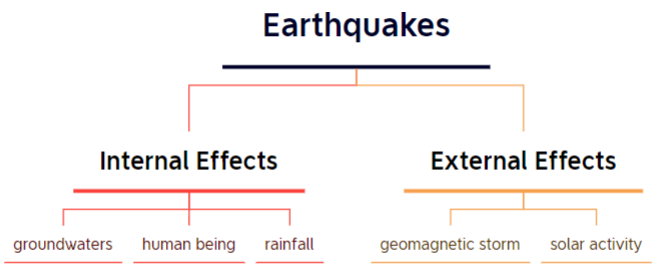

Fig. 1 A categorization of the several factors that might cause earthquakes

Every year, new and improved technology significantly enhances the amount of data on solar activity and earthquakes. Alongside this, smarter algorithms and greater computing power are available to process these massive amounts of data. Some researchers in the field of computer science have been making strides in this direction by using artificial intelligence (AI) methods. As learning is a prerequisite for AI [8], machine

Viacheslav Shkuratskyy is with Department of Computer Science, York St John, Lord Mayor's Walk, York, YO31 7EX, UK, (corresponding author, e-mail: viacheslav.shkurat@yorksj.ac.uk).

Aminu Bello Usman is with School of Computer Science, University of Sunderland, Edinburgh Building, City Campus, Chester Road, Sunderland, SR1 3SD, UK (e-mail: aminu.usman@sunderland.ac.uk).

Michael O'Dea is with School of Computer Science, University of York, York YO10 5DD, UK (e-mail: michael.odea@york.ac.uk).

Mujeeb Ur Rehman is with School of Computer Science and Informatics, De Montfort University, Gateway House, Leicester LE1 9BH, UK (e-mail: mujeeb.rehman@dmu.ac.uk).

Saifur Rahman Sabuj is with Department of Electrical and Electronic Engineering, Kha 224 Bir Uttam Rafiqul Islam Avenue, Merul Badda, Dhaka, Bangladesh (e-mail: srsabuj@bracu.ac.bd).

learning (ML) is a significant component of the field [9].

Although the expanding volume of solar and climatic data, increased computer processing capacity and use of advanced data analysis tools helped identify catastrophe risks, threats, and timing and location. There have still been few research studies that have used ML for effective solar activity-based natural catastrophe prediction. Solar and natural catastrophe data are often unstructured and large, making them challenging to analyse and handle.

This study seeks to determine how well ML can forecast earthquakes based on solar activity. The goal of this research is to investigate how effective ML is at predicting earthquakes based on solar activity. To achieve the aim, it uses studies four algorithms with distinct problem-solving methodologies four algorithms:

1) K-Nearest Neighbour is one of the uncomplicated algorithms with fast training speed [10].
2) Support Vector Regression is a kernel-based algorithm [11].
3) Random Forest Regression is an ensemble learning algorithm with a tree structure [10], [12].
4) LSTM Network is a neural networks algorithm [13].

In comparison to previous studies, the study uses four algorithms with different problem-solving strategies to predict earthquakes based on solar activity events. Also, it classifies earthquakes by their depth and magnitude, utilising ML methods to investigate potential connections between solar activity and earthquakes. The study aims to utilise this as a springboard for more in-depth studies of earthquakes in the future. Moreover, it leverages seismology data to bolster its findings with ML techniques. While this study is just the tip of the iceberg in terms of earthquake prediction using ML techniques, the ramifications for the future are profound.

II. Related Work

ML research investigating earthquakes or solar activity whilst still uncommon is increasing. For instance, [14] predicted the quantity of sunspots for the 25th solar cycle using data on sunspot numbers between 1818 and 2020. According to the analysis, the 25th solar cycle will take place between 2021 and 2025. In [15], authors conducted the research on this subject. Classifying solar wind plasma often involves only two categories: "fast wind" and "slow wind" both of which are determined by the speed of the wind. In order to examine earthquake prediction utilizing seismic parameters and earthquake data, Asim et al. [16] utilised a comprehensive array of ML techniques, including pattern recognition neural networks, recurrent neural networks, random forests, and a linear programming boost ensemble classifier. When compared to the other algorithms, each method provided a distinct set of findings.

One of the first studies that considered both seismological factors and solar activity was by Wolf in 1853 first linked earthquakes to solar activity [3]. Odintsov et al. [17], [5] claimed that the 11-year solar cycle causes earthquakes. They also found that high-speed solar winds also cause earthquakes with Richter magnitudes greater than 5.5. The study [18] showed that the Earth's crust's current density relies on its electrical conductivity and may affect earthquakes. After solar exposure-like effects of electric current on the Earth's crust, they saw a spike in earthquakes under 3 magnitudes. They also correlated solar flares to earthquakes.

Data-driven research has found a connection between solar activity and worldwide earthquakes. For instance, Marchitelli et al. [19] employed statistical techniques to identify a connection between solar wind and earthquakes with Richter magnitudes equal to or higher than 5.6. Support vector regression was used to establish a connection between earthquakes and solar activity in [20]. However, in contrast to the Marchitelli study, they showed that earthquakes with a Richter magnitude of less than six are affected by solar activity. The study [21], using SVM, showed the correlation between solar and earthquakes. However, they used earthquakes with magnitude more than 6. In contrast, [6] asserted that there is no statistically significant evidence to support the concept that solar-terrestrial interaction increases the frequency of earthquakes using the X2 and student's t-tests. They did not, however, refute the idea that solar activity has no impact on earthquakes. Akhoondzadeh and De Santis [7] analysed solar activity and strong earthquakes with magnitude more than 7. They used stimulated datasets and argued that they did not find a relationship between solar activity and earthquakes.

This study delves into the hypothesis concerning the efficacy of ML in earthquake prediction based on solar activity. In pursuit of this hypothesis, several pivotal sub-questions necessitate exploration. For instance, an inquiry into the pertinent characteristics and classifications of earthquakes and solar activity crucial for accurate prediction arises. Furthermore, it is imperative to discern whether solar activity events exert uniform influence across various types of earthquakes, and to ascertain the significance of additional factors such as time delay. Additionally, the selection of optimal ML algorithms to address the core research question becomes pivotal, with a particular focus on identifying the algorithm that yields the highest predictive accuracy among those under consideration.

III. Data Collection and Preparation

This study uses open-source earthquake and solar-activity datasets. Since solar activity is measured throughout cycles, data were collected between the 23rd and 24th solar cycles (1996-2020), even though these cycles have already finished. Daily earthquake datasets were downloaded from the United States Geological Survey (USGS) website [22].

Seismic activity is divided into three categories based on the crust of the Earth's effect on the electric current. It was established two distinct sets of categories for earthquakes. Tremors with a Richter magnitude of less than 5.5 fall under the first group. Earthquakes of magnitude 5.5 or above on the Richter scale fall under the second group.

Sunspot number, solar wind, and solar flares represent solar activity. The SILSO World Data Center provides sunspot number data [23]. The solar wind delivers solar activity to the earth. According to Wood et al. [24], solar wind measurements include velocity, density, and temperature. These data come from NASA OMNIWeb [25]. Solar flares are categorized by strength: A, B, C, M, and X. A-class solar flares are the smallest and X-class the biggest [26]. The NOAA National Center for Environmental Information websites [27] and GOES-R series [28] provide solar flares data for the whole research period.

As a result, information about solar activity and earthquakes in three different depth zones was used: the shallow, the intermediate, and the deep. There are three datasets: one for solar activity and earthquakes from the shallow zone, one for solar activity and earthquakes from the intermediate zone, and one for solar activity and earthquakes from the deep zone are being used.

As the Sun and the Earth are so far apart, it takes time for information of solar activity to reach the Earth. According to Wood et al. [24], the speed of the solar wind is between 300 and 800 km/s. This means that solar activity takes between two and seven days to reach Earth. Also, Sytinskii [4] said that most earthquakes happen two to three days after the sun has moved through the central solar meridian. Because of this, the earthquake data are two to seven days late. In addition, Novikov et al. [18] say that more earthquakes could be caused by activity on the sun. As a result, this study leaves outliers in the data, even though they might change the final result. Odintsov et al. [17] say that when it comes to solar activity, a fast solar wind can have a big effect on earthquakes. This study uses different normalisation scalers to run the normalisation process and reduce the effect of outliers. The best normalization results come from the Quantile Transformer scalar [29]. Also, Nishii et al. [20], say that not all things that happen on the sun may influence earthquakes. The study uses the Principal Component Analysis algorithm to reduce the number of variables in the data relating to how the sun works. The Principal Component Analysis algorithm results say that the first six principal components explain more than 96% of the differences in data about solar activity.

The study first determines if the correlation between solar activity and earthquakes is linear or nonlinear, and only then does it utilize ML methods. In order to evaluate this, an ML approach based on linear regression and calculated using the R-squared error is used. By using this approach we found that earthquakes and solar activity have a nonlinear relationship. Therefore, ML methods like SVM with a linear kernel that make predictions using a linear function cannot be used for this study.

IV. Performance Evaluation

Multiple metrics are used in accordance to the findings of Chai and Draxler [30] who argue that it is a good practice to use a range of metrics. The findings from various datasets are examined and the results are assessed using the normalized by mean values of the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). To compare the models, the normalised values of RMSE (NRMSE) and MAE (NMAE), which are calculated using (1) and (2) [31] are used.

$$NRMSE_{EQ} = \frac{RMSE_{EQ}}{\overline{EQ}} \tag{1}$$

where $RMSE_{EQ}$ represents RMSE of a model, EQ represents mean of earthquakes.

$$NMAE_{EQ} = \frac{MAE_{EQ}}{\overline{EQ}} \tag{2}$$

where MAE_{EQ} represents MAE of a model.

Since the study discovered that there is a nonlinear link between earthquakes and solar activity, it utilizes this information to help choose the best method for testing looking for a correlation between the two. Additionally, it aims to use algorithms that approach challenges in creative ways. To implement the selected algorithms, the python libraries Scikit-learn–Supervised learning [32] and Keras: Python deep learning API [33] are used. One of the simplest and fastest training techniques is the K-Nearest Neighbour (KNN) algorithm, which is based on Euclidean distance. The "K"-value is the most crucial factor when using KNN. Using the elbow technique, the value of 17 for the ideal "K"-value was obtained.

The other technique uses Support Vector Regression (SVR), a kernel-based algorithm. A kernel serves as the primary SVR parameter. SVR may resolve both linear and nonlinear problems depending on the selected kernel. The RBF kernel was chosen because it is nonlinear, extensively utilized, well-researched, and has just one kernel parameter [11].

The subsequent algorithm, tree-structured Random Forest Regression (RFR), is an example of ensemble learning. The two most important factors for RFR are the number of regression trees and the number of features. Two features are chosen since the accuracy of the model increases when the number of features and correlation across trees are reduced. According to Oshiro et al. [34], the number of trees should range between 64 and 128. However, the greater the number of trees, the better the outcome, as RFR does not have an overfitting issue, but opting for this approach takes longer. This is why the default number in the Scikit-learn–Supervised learning library [32] is chosen.

Long Short-Term Memory (LSTM), a neural network technique, is also implemented. As a function, this study utilizes the sigmoid function to develop a neural network model. To lower the generalisation power and cost of the network [35], the number of hidden LSTM layers is set to two. For calculating the number of concealed nodes, the utilized equation (3) is from the Keras: Python deep learning API [33].

$$N_h = \frac{N_s}{(\alpha * (N_i + N_o))} \tag{3}$$

where N_s represents number of samples in training data set, N_i represents number of input neurons, N_o represents number of output neurons, and α represents scaling factor (from 2 to 10).

Since [36] reveals that LSTM with epochs of 25 and 30 yield the best results, 70 epochs are chosen, a number that is more than 25-30 but smaller than100. This is because Sunny et al. [37] found that a model with 100 epochs and 2 hidden layers yields the best results.

A. Solar Activity and Shallow Zone Earthquakes

1) Earthquakes under 5.5 Richter Scale

The study observed that the placements of the algorithms, both optimum and optimally optimal, remain constant over the whole experimental domain. KNN and RFR have the highest errors values, while LSTM and SVR have the lowest error values. Fig. 2 displays observed and projected earthquake values with a three-day lag. The position of the prediction lines in the other portions of the experiment is identical to that of the three-day delay portion. The study also observed that seismic activity usually peaks. This emphasises the importance of outliers in the seismic data. Thus, NRMSE values are probably recommended in this situation. It was discovered that although the NMAE result was best with a six-day delay (0.4575), the NRMSE result is best with a three-day delay (0.5431). The values of the mistakes, frequently cluster together. More proof of this may be found in the positions of the lines indicating the projected values of the algorithms and the lines representing the actual values of earthquakes.

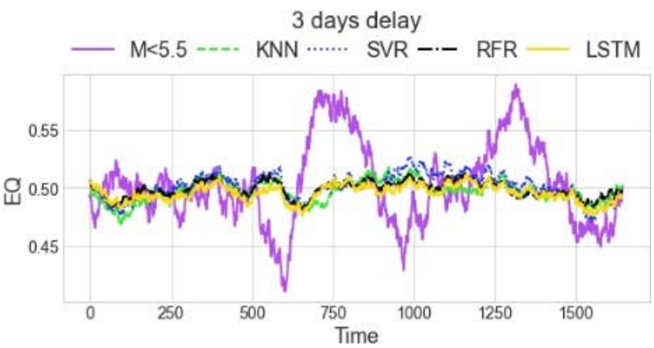

Fig. 2 Comparing actual and predicted values for Shallow zone earthquakes $M < 5.5$

2) Earthquakes Equal and above 5.5 Richter Scale

In this part of the experiment, RFR and LSTM have the lowest error values in terms of NRMSE. Whereas, in terms of NMAE, SVR has the lowest error values, followed by LSTM. The optimal results for NRMSE = 0.8764 with a three-day delay and NMAE = 0.8485 with a five-day delay are both achieved with a delay of three days. Furthermore, similar to the early parts of the experiment, the earthquakes have peaks (Fig. 3). Extreme numbers in the earthquake data have more bearing on the overall picture. For this and the preceding reason, RMSE values are probably the best choice. Additionally, both measures in this data set exceed the outcomes seen in the data set for shallow zone earthquakes with a magnitude of 5.5 or less. Additionally, the normalized RMSE values are quite near to 1, and sometimes even exceed 1. Fig. 3 shows that the LSTM and RFR prediction lines are quite near to each other and to the averages of the actual values when comparing predicted and observed earthquake magnitudes. When compared to the upper bound of real values, the SVR prediction line is rather near the actual values. The KNN prediction line is likewise close to the middle of the real values distribution, however it does not perfectly replicate the actual values line. Fig. 3 shows that while the LSTM prediction lines have been shifting, the lines produced by conventional ML methods have stayed stationary.

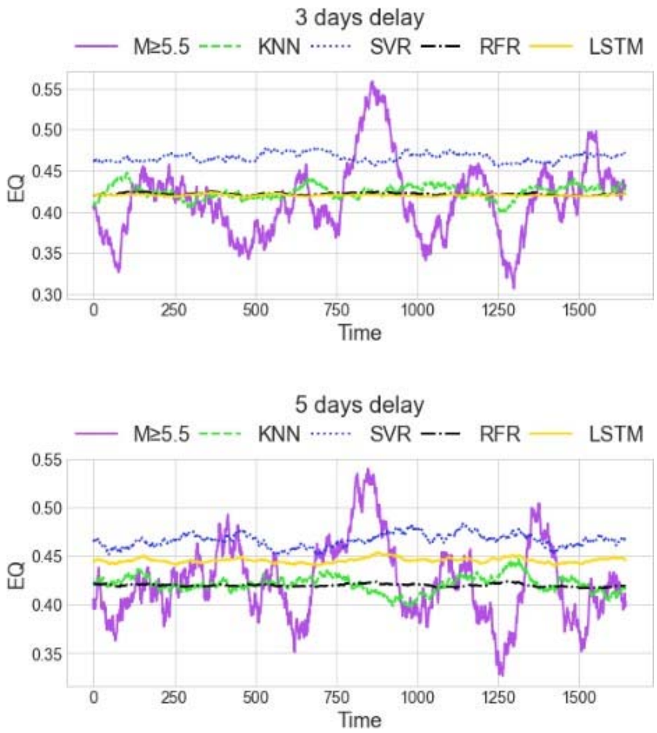

Fig. 3 Shallow zone earthquakes $M \geq 5.5$: Compare actual and predicted values

B. Solar Activity and Intermediate Zone Earthquakes

1) Earthquakes under 5.5 Richter Scale

LSTM and SVR achieve the lowest error values in both metrics in this experiment section. In RMSE, LSTM has the lowest error values, while in MAE, SVR has the lowest error values. In both metrics, RFR and KNN have the highest error values. In terms of NRMSE = 0.5386, the five-day delay part has the best result, while the six-day delay part has the best result in terms of NMAE = 0.4384. Intermediate depth data errors exhibit a higher magnitude compared to errors observed in shallow depth data. Furthermore, akin to earlier phases of the experiment, seismic events display distinct peaks, as illustrated in Fig. 4. Notably, the earthquake data manifest significant deviations from the mean, indicating that RMSE values are likely the preferred metric for evaluating performance in both scenarios. In Fig. 4, the discrepancy between actual and predicted earthquake values is visually evident. Notably, the prediction lines of the algorithms appear closely clustered, with the LSTM prediction line slightly distinct from the rest. Interestingly, in certain instances, the LSTM prediction line closely aligns with the actual values, demonstrating its potential for outperforming other algorithms in accurately forecasting seismic events.

2) Earthquakes Equal and above 5.5 Richter Scale

In this experiment section, the relative standard error (RMSE) is straightforward, and all methods are in the same relative delay locations. In both metrics the four-day delay part has the best result, NRMSE = 2.2464 and NMAE = 1.4377. However, the normalised error values are quite high, i.e. are more than “1”, this shows that the prediction is not perfect. The subsequent segments of the experiment involving deeper earthquakes exhibit higher error values compared to their counterparts focusing on shallow seismic events. Additionally, the earthquake dataset showcases notable deviations from the mean, featuring significant values that underscore the preference for RMSE values in both this scenario and the preceding one. Fig. 5 provides a visual representation of the disparity between actual and predicted earthquake values, revealing results that fall short of ideal. While the prediction lines of LSTM, RFR, and KNN algorithms closely approximate the averages of actual values, the SVR prediction line consistently resides towards the lower end of the actual values spectrum. Furthermore, none of the algorithmic prediction lines faithfully follow the trajectory of actual values, highlighting areas for improvement in the predictive accuracy of the models.

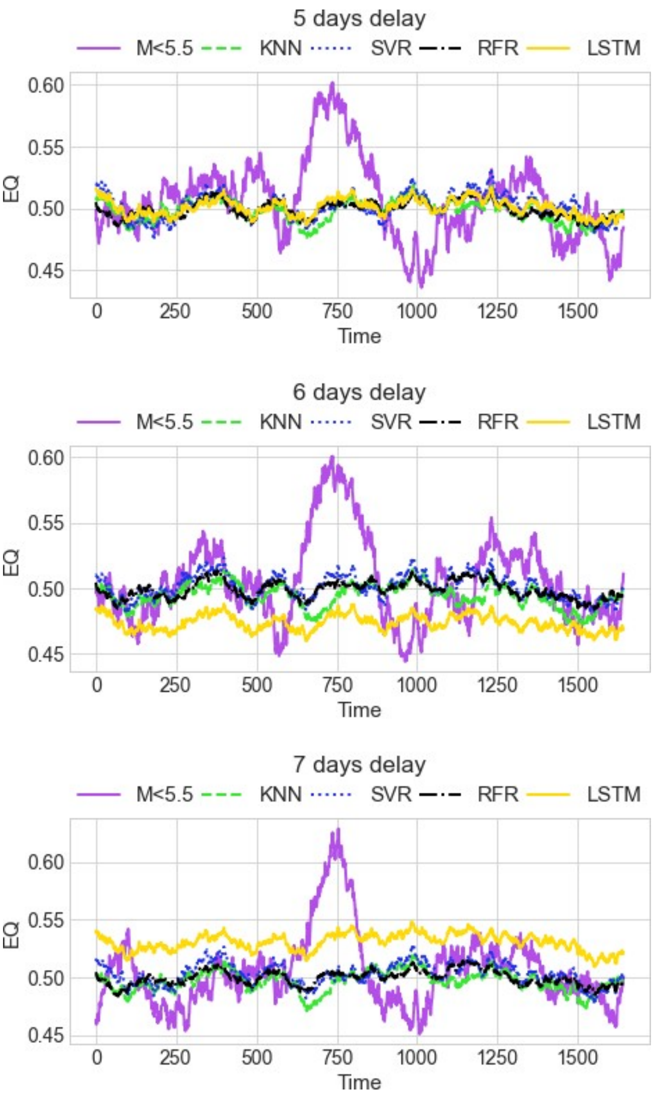

Fig. 4 Intermediate zone earthquakes $M < 5.5$: Compare actual and predicted values

C. Solar Activity and Deep Zone Earthquakes

1) Earthquakes under 5.5 Richter Scale

The experimental results in this section show that RFR and LSTM achieved the highest accuracy values across both metrics. The values of the highest accuracy in both metrics are close to each other. The two-day delay part has the best result for NRMSE = 0.6253, while the seven-day delay part has the best result for NMAE = 0.5177. The data from the earthquakes (Fig. 6) also include notable values that deviate more from the mean. As a result, RMSE values are preferred. The normalised RMSE by standard deviation values are quite near to “1” or even higher. Fig. 6 also shows that the difference between the actual and predicted values of earthquakes change the location of the LSTM prediction line while the traditional ML algorithms maintain the location of their prediction lines as pretty much the same.

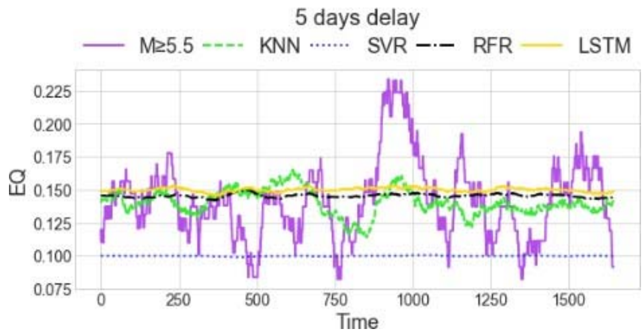

Fig. 5 Intermediate zone earthquakes $M \geq 5.5$: Compare actual and predicted values

2) Earthquakes Equal and above 5.5 Richter Scale

In this experiment section the range variations between two metrics are rather large. In terms of RMSE, the first two positions are occupied by LSTM, whereas in terms of MAE, the first three positions are occupied by LSTM, KNN, and RFR. For NRMSE = 3.3972, the six-day delay part produces the best results, whereas for NMAE = 1.6002, the three-day delay part produces the best results. However, the normalised error values are too high, more than “1” and so the accuracy of the prediction is poor. There are also notable values that are located far from the mean in the earthquake data (Fig. 7). So RMSE values are preferred. Fig. 7 shows the example of the difference between actual and predicted values of earthquakes. It illustrates that the prediction outputs of various algorithms diverge from the actual trend. While LSTM and RFR forecasts cluster near the mean of observed values, the KNN predictions also tend towards the mean albeit with notable peaks. In contrast, the SVR predictions stand out, surpassing those of other algorithms. The locations of the prediction lines, for each algorithm are similar to each part of the experiment.

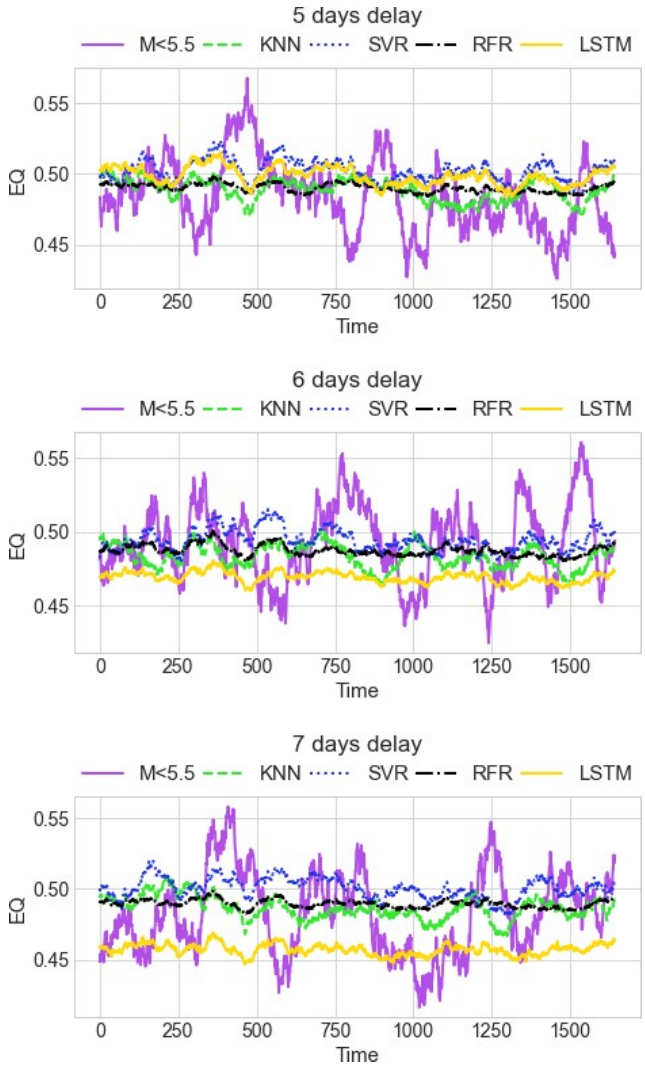

Fig. 6 Deep zone earthquakes $M < 5.5$: Compare actual and predicted values

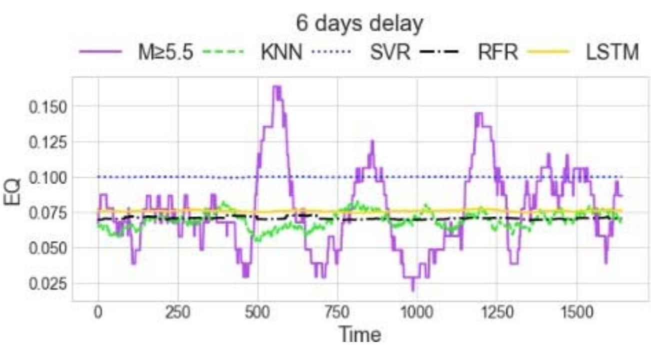

Fig. 7 Deep zone earthquakes $M \geq 5.5$: Compare actual and predicted values

V. Evaluation

This study examines the relationship between solar activity and earthquakes and the effectiveness of using ML in earthquake prediction based on solar activity. One of the most significant earthquake characteristics that can be used in earthquake prediction is earthquake magnitude and depth, which is also confirmed by the literature [5], [18]. The experiment divides global earthquake data by their depth and magnitude. For solar activity the selection is based on previous seismological and space studies [5], [18], such as sunspot number, solar wind (solar wind speed, proton density, proton temperature), and solar flares (A, B, C, M, X classes). For evaluation of the results, normalised RMSE and MAE values are used. The experiment shows that earthquakes have upper and lower peaks. The values that deviate significantly from the mean in the earthquake data are significant. Because RMSE is the square root of the average of squared errors and gives large errors a lot of weight, their values are more useful here. The experiment found that the correlation between solar activity and earthquakes of Richter magnitude $M < 5.5$ has smaller normalised error values than the correlation between solar activity and earthquakes of Richter magnitude $M \geq 5.5$. Based on these results, it seems that solar activity may have a greater effect on earthquakes of magnitude $M < 5.5$ than earthquakes of magnitude $M \geq 5.5$. This conclusion is supported by the findings of [20], which found that solar activity events have the greatest impact on earthquakes with Richter magnitudes of less than 4. However, Odintsov et al. [5] found a stronger correlation between solar activity and earthquakes of magnitude 5.5 or greater, thus these results should be interpreted with caution.

Furthermore, the error values tend to rise in tandem with the depth of earthquakes. So, it can be assumed that solar activity mainly affects shallow zone earthquakes. The RMSE and MAE values, normalised by mean, are larger than “1” in cases of greater depth and greater magnitude. This suggests that solar activity may not have any bearing on very deep and very powerful earthquakes. The results are supported by the research of Novikov et al. [18]. They conducted an experiment showing that earthquakes can be affected by electric current generated by solar activity. In the deeper layers of the Earth’s crust, the density of the current increases as the electrical conductivity of the crust rises.

According to the results summary, LSTM has a higher accuracy than other models in earthquake prediction LSTM attempts to analyse all of the data before making a prediction about the next number. KNN is typically applied to nearby related points of a data point. SVR attempts to predict the value for each row as a separate training sample based on the data it has collected. RFR employs an ensemble technique and does not suffer from overfitting. The error values, on the other hand, were fairly close to one another. That is why, additional experiments with varying parameters, such as K- value in KNN, kernel in SVR, number of trees in RFR, and nodes and epochs in LSTM, are required to determine which algorithm will provide the best accuracy.

VI. Conclusion and Future Work

The LSTM model showed the highest prediction accuracy when compared to the other algorithms. However, there are not many differences between the accuracy values of each algorithm. The findings show prediction accuracy is far from ideal even though the accuracy values of each algorithm are not significantly different. The study discovers that the error values increase with the depth and magnitude of earthquakes, which supports argument of [18] that earthquakes are affected by the electric current produced by solar activity events.

One of the ways to increase the accuracy is to change the attributes of ML algorithms. In the case of KNN, the only changeable attribute is K-value. However, based on the finding of K-values, the error values do not change significantly after K = 17. The primary factor in SVR is the kernel function. In SVR, four kernels are frequently used. In addition to changing kernels, it is also necessary to set the parameters for the new kernels. To find the most precise solution, future studies could try modifying these kernels and comparing the error values. For the RFR algorithm changing the number of trees is one of the first steps in the process to improve accuracy. The accuracy of the prediction will also change if RFR's parameters (like the maximum tree depth) are changed. To increase the LSTM model's prediction accuracy, the number of neurons in the hidden layer can be changed (making the model wider), hidden layers can be added (making the model deeper), or a combination of both methods can be used. Precision will also change with different node counts and epochs. Given that this was the best performing algorithm in this study, it would suggest that this is likely to be the most obvious of future direction to explore.

The accuracy of the results can also be improved by expanding data sets and including new variables, such as extending the time period to more than two solar cycles (24 years), including more solar activity events, and using images of solar activity events like solar flares or solar wind. These measures will also improve a neural network algorithms' accuracy; however, this is not the best scenario for conventional ML algorithms. It is also crucial that any dataset used is reliable and a strong foundation for the study. There are other sources that provide data on solar activity and earthquakes, these could be incorporated into future studies in order to improve accuracy.

The study suggests that the LSTM model has a greater potential for predicting earthquakes based on solar activity events even using the most basic settings, it has more parameters that can be changed to increase the final prediction accuracy. The LSTM model, on the other hand, costs more than the conventional ML algorithms used in this study because it consumes more time, energy, and expensive resources. Although this is becoming less of an issue, it is still worth considering persevering with research using conventional ML algorithms that incorporate new algorithms and different parameter settings. The purpose of this study was to investigate the relationship between solar activity events and earthquakes. It has found that there is evidence to conclude that there is a link after analysing their connections over two solar cycles using ML methods.

REFERENCES

[1] N. Meyer-Vernet, *Basics of the solar wind*. Cambridge University Press, 2007.

[2] J. Gribbin, "Relation of sunspot and earthquake activity," *Science*, vol. 173, no. 3996, p. 558–558, Aug 1971. (Online). Available: https://www.science.org/doi/10.1126/science.173.3996.558.b

[3] R. Wolf, "On the periodic return of the minimum of sun-sport; the agreement between those periods and the variations of magnetic declination," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 5, no. 29, pp. 67–67, 1853.

[4] A. Sytinskii, "Relation between seismic activity of the earth and solar activity." *Uspekhi Fizicheskih Nauk*, vol. 111, no. 10, p. 367, 1973. (Online). Available: http://ufn.ru/ru/articles/1973/10/i/

[5] S. D. Odintsov, G. S. Ivanov-Kholodnyi, and K. Georgieva, "Solar activity and global seismicity of the earth," *Bulletin of the Russian Academy of Sciences: Physics*, vol. 71, no. 4, p. 593–595, Apr 2007. (Online). Available: http://link.springer.com/10.3103/S1062873807040466

[6] J. J. Love and J. N. Thomas, "Insignificant solar-terrestrial triggering of earthquakes: Insignificant triggering," *Geophysical Research Letters*, vol. 40, no. 6, p. 1165–1170, Mar 2013. (Online). Available: http://doi.wiley.com/10.1002/grl.50211

[7] M. Akhoondzadeh and A. De Santis, "Is the apparent correlation between solar-geomagnetic activity and occurrence of powerful earth- quakes a casual artifact?" *Atmosphere*, vol. 13, no. 7, p. 1131, 2022.

[8] E. L. Thorndike, "Animal intelligence: experimental studies. new brunswick," 2000.

[9] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," *Reports on Progress in Physics*, vol. 81, no. 7, p. 074001, 2018.

[10] E. Alpaydin, *Introduction to machine learning*, third edition ed., ser. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2014.

[11] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, p. 199–222, Aug 2004. (Online). Available: http://link.springer.com/10.1023/B:STCO.0000035301.49549.88

[12] L. Breiman, *Machine Learning*, vol. 45, no. 1, p. 5–32, 2001. (Online). Available: http://link.springer.com/10.1023/A:1010933404324

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, p. 1735–1780, Nov 1997. (Online). Available: https://direct.mit.edu/neco/article/9/8/1735-1780/6109

[14] T. Khan, F. Arafat, M. U. Mojumdar, A. Rajbongshi, S. M. T. Siddiquee, and N. R. Chakraborty, "A machine learning approach for predicting the sunspot of solar cycle," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2020, pp. 1–4.

[15] U. Feldman, "On the sources of fast and slow solar wind," *Journal of Geophysical Research*, vol. 110, no. A7, p. A07109, 2005. (Online). Available: http://doi.wiley.com/10.1029/2004JA010918

[16] K. M. Asim, F. Mart´ınez-A´ lvarez, A. Basit, and T. Iqbal, "Earthquake magnitude prediction in hindukush region using machine learning techniques," *Natural Hazards*, vol. 85, no. 1, p. 471–486, Jan 2017. (Online). Available: http://link.springer.com/10.1007/s11069-016-2579-3

[17] S. Odintsov, K. Boyarchuk, K. Georgieva, B. Kirov, and D. Atanasov, "Long-period trends in global seismic and geomagnetic activity and their relation to solar activity," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 31, no. 1-3, pp. 88–93, 2006.

[18] V. Novikov, Y. Ruzhin, V. Sorokin, and A. Yaschenko, "Space weather and earthquakes: possible triggering of seismic activity by strong solar flares," *Annals of Geophysics*, vol. 63, no. 5, pp. PA554–PA554, 2020.

[19] V. Marchitelli, P. Harabaglia, C. Troise, and G. De Natale, "On the correlation between solar activity and large earthquakes worldwide," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[20] R. Nishii, P. Qin, and R. Kikuyama, "Solar activity is one of triggers of earthquakes with magnitudes less than 6," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa, HI, USA: IEEE, Sep 2020, p. 377–380. (Online). Available: https://ieeexplore.ieee.org/document/9323381/

[21] S. Asaly, L.-A. Gottlieb, N. Inbar, and Y. Reuveni, "Using support vector machine (svm) with gps ionospheric tec estimations to potentially predict earthquake events," *Remote Sensing*, vol. 14, no. 12, p. 2822, 2022.

[22] "Earthquakes." (Online). Available: https://www.usgs.gov/programs/earthquake-hazards

[23] "Silso: World data center for the production, preservation and dissemination of the international sunspot number." (Online). Available: https://www.sidc.be/silso/

[24] B. E. Wood, R. A. Howard, A. Thernisien, and D. G. Socker, "The three-dimensional morphology of a corotating interaction region in the inner heliosphere," *The Astrophysical Journal*, vol. 708, no. 2, p. L89–L94, Jan 2010. (Online). Available: https://iopscience.iop.org/article/10.1088/2041-8205/708/2/L89

[25] "Spdf-omniweb service." (Online). Available: https://omniweb.gsfc.nasa.gov/

[26] E. R. Priest, *Magnetohydrodynamics of the Sun*. New York, NY: Cambridge University Press, 2014.

[27] N. G. D. Center, "Solar flare data — solar terrestrial physics." (Online). Available: https://www.ngdc.noaa.gov/stp/solar/solarflares.html

[28] N. C. for Environmental Information (NCEI), "Goes-r series: Ncei." (Online). Available: https://www.ngdc.noaa.gov/stp/satellite/goes-r.html

[29] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. Tirunelveli, India: IEEE, Aug 2020, p. 729–735. (Online). Available: https://ieeexplore.ieee.org/document/9214160/

[30] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, p. 1247–1250, Jun 2014. (Online). Available: https://gmd.copernicus.org/articles/7/1247/2014/

[31] M. V. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. P. Tyukov, T. A. Janovsky, V. A. Kamaev *et al.*, "A survey of forecast error measures," *World applied sciences journal*, vol. 24, no. 24, pp. 171–176, 2013.

[32] "Supervised learning." (Online). Available: https://scikit-learn.org/stable/supervised learning.html (accessed Feb. 2, 2021)

[33] "Keras: the python deep learning api." (Online). Available: https://keras.io/ (accessed: Aug 29 2021)

[34] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?" in *International workshop on machine learning and data mining in pattern recognition*. Springer, 2012, pp. 154–168.

[35] V. Verdhan and E. Y. Kling, *Supervised learning with Python: concepts and practical implementation using Python*, ser. For professionals by professionals. New York, NY: Apress, 2020.

[36] J. Kim, S. Kim, H. Wimmer, and H. Liu, "A cryptocurrency prediction model using LSTM and GRU algorithms," in *2021 IEEE/ACIS 6th International Conference on Big Data, Cloud Computing, and Data Science (BCD)*. Zhuhai, China: IEEE, Sep 2021, p. 37–44. (Online). Available: https://ieeexplore.ieee.org/document/9581397/

[37] M. A. Istiake Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using lstm and bi-directional lstm model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. Giza, Egypt: IEEE, Oct 2020, p. 87–92. (Online). Available: https://ieeexplore.ieee.org/document/9257950/