Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Hybrid CNN-LSTM Approach for Geolocation-based Earthquake Risk Prediction using USGS Data

Sneka M.[1], Kanchana K.[2]

Department of Computer Science and Engineering, Kathir College of Engineering, Coimbatore, India

Email: [1] snekamurugesan83@gmail.com, [2] kanchana@kathir.ac.in

Abstract

Predicting earthquake risk accurately is essential for reducing the devastating effects of seismic activity. To improve earthquake risk prediction by geolocation, this study presents a hybrid learning strategy that combines Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. The research uses CNNs to extract spatial characteristics from geophysical and seismic data and LSTM networks to capture temporal dependencies in the sequence of events using the United States Geological Survey (USGS) dataset. The model creates a strong foundation for estimating earthquake risk at the local and regional levels by combining these complementary approaches. By integrating these complementary approaches and utilizing geolocation data such as latitude, longitude, depth, and proximity to fault lines, the model provides a robust framework for local and regional earthquake risk estimation, providing granular insights into vulnerable areas. According to experimental data, the hybrid CNN-LSTM model works better than conventional machine learning techniques, resulting in reduced false positives and increased prediction accuracy. Additionally, the model demonstrates flexibility and scalability, enabling real-time updates through the use of streaming data from IoT-enabled sensors and seismometers.

Keywords: Hybrid Learning Approach, Earthquake Risk Prediction, Geolocation-based Prediction, CNN-LSTM Model, United States Geological Survey Dataset, Spatiotemporal Data Analysis, Seismic Risk Modeling, Machine Learning for Earthquake Forecasting

DOI: https://doi.org/10.36548/jaicn.2025.1.005
Received: 06.03.2025, received in revised form: 05.04.2025, accepted: 21.04.2025, published: 30.04.2025

1. Introduction

One of the most destructive natural catastrophes, earthquakes continue to cause enormous destruction and a high death toll all over the globe [1]. More accurate earthquake risk prediction is essential for emergency preparation, disaster mitigation, and lessening the negative effects on the economy and society. The complex interaction of geological, environmental, and temporal elements makes earthquake prediction difficult, even with advances in seismology. The dynamic and nonlinear character of seismic occurrences may be missed by traditional approaches, which often depend on static statistics and basic models.

More advanced methods for predicting earthquake risk have been made possible by recent advancements in deep learning and machine learning [2]. Specifically, hybrid models combining multiple learning approaches provide new possibilities for achieving unprecedented levels of geophysical data analysis and interpretation. Long Short-Term Memory (LSTM) networks are excellent at simulating temporal relationships in sequential data, whereas Convolutional Neural Networks (CNNs) are well known for their capacity to extract spatial characteristics from structured data. A hybrid learning strategy that combines these two architectures can handle temporal and spatial information more efficiently, improving prediction accuracy.

In this study, the United States Geological Survey (USGS) dataset is used to forecast earthquake hazards using a CNN-LSTM hybrid learning model. The dataset is a useful resource for training and assessing prediction models as it provides extensive geophysical data, such as earthquake locations, magnitudes, depths, and timestamps. The suggested method seeks to give more precise and detailed risk evaluations by using geolocation data, especially for seismic zones that are considered high risk.

The rest of the work demonstrates how the CNN-LSTM hybrid model was designed, how USGS geophysical data was pre-processed, and how geolocation elements were included. Results from experiments are shown to compare the model's performance to more conventional methods. The study concludes by discussing the research's implications for disaster management, urban planning, and the creation of real-time early warning systems and suggesting possible future improvements for wider application.

2. Related Work

Precursory anomaly analysis and statistical approaches are part of the current earthquake prediction techniques. Statistical approaches focus on past earthquakes and the structure of possible sources, and they employ geological and statistical theories to examine the connection between past earthquakes and future earthquakes.

The principle that there will be some abnormalities behind an earthquake is the foundation of the precursory anomalies analysis approach. To identify each kind of earthquake precursor and investigate the suggested physical explanations to explain each one, Cicerone reviewed the available scientific literature [3]. Rikitake examines the likelihood that an anomalous signal of different geophysical elements will be associated with an impending earthquake. These odds were calculated using the available precursor data [4]. It was found that as the mainshock magnitude increases, precursors can be found farther from the epicentre [5]. Yun-Tai discovers that these gravity changes seem to be closely linked to earthquake occurrences [6]. According to Brodsky, there may be detectable preludes to major earthquakes, as shown by the Chilean incident on April 1, 2014 [7]. In contrast to a long-term analysis based on the Gutenberg–Richter law or other statistical laws, these works demonstrated that there are typically some precursor signals prior to a earthquake rupture, such as land deformation reported by geodetic survey and tide-gauge observation, ground tilt observed by a water-tube tiltmeter, anomalous seismic activity, and geomagnetic field change. However, the analysis of earthquake precursory abnormalities is challenging because of their complexity and unpredictability. The study uses PCA, a popular technique, to extract the data since it performs well in feature extraction. A kind of robust machine learning model called LightGBM is based on the decision tree is used as it is fast, stable, and has a high degree of accuracy and prediction ability [8].

According to Colombelli et al. [9], the P (primary) wave peak displacement changes over time for various earthquakes in the early stages of the rupture process. They found that the peak displacement of small earthquakes increases quickly in the initial phase, whereas the peak displacement of large earthquakes grows slowly. In contrast, Rydelek and Horiuchi [10] contested the notion of a deterministic assumption, arguing that the rupture process is ultimately unexpected and that the earthquake nucleation process is universal and independent of the final magnitude. According to rupture unpredictability, early rupture behaviour alone cannot be used to forecast an earthquake's ultimate size.

Rupture weak determinism, which addresses the capacity to infer the ultimate earthquake magnitude after the nucleation, has replaced deterministic rupture nucleation as the main emphasis of earthquake research in recent years. According to Melgar and Hayes' [11] weakly determined model of rupture development, big earthquakes rupture by self-similar slip pulses. They created an average source–time function with 0.5 magnitude units within the magnitude range of Mw 7.0 to 8.5 in order to examine the change in the average seismic moment rate over time. According to the data, the average seismic moment rate was much lower in the first 10 seconds than it was throughout the earthquake's length. This indicates that shortly after the rupture begins, a self-similar slip pulse forms. Therefore, it may be possible to rule out entirely deterministic rupture processes. Nevertheless, a thorough examination of seismic or geodetic data might sometimes still reveal weak or probabilistic types of deterministic assumption.

3. Proposed Work

In order to create and evaluate a hybrid learning strategy for earthquake risk prediction, this study adheres to a systematic methodology. Before combining them into a hybrid CNN-LSTM framework, the procedure includes data gathering, preprocessing, and the creation of separate CNN and LSTM models [12]. Each stage is explained in depth in the sections that follow.

3.1 Dataset Collection

The dataset utilized in this investigation was taken from USGS [15], which maintains an extensive global database of earthquake records. This study employed the USGS Earthquakes 2024 dataset, comprising 14,150 data samples with 22 characteristics. The dataset includes essential features such as latitude, longitude, depth, magnitude, time of occurrence, and location specifics. To enhance forecast accuracy, it also incorporates geophysical and environmental characteristics like past seismic activity, tectonic plate boundaries, and proximity to fault lines. Figure 1 illustrates sample data from this dataset Multiple decades' worth of earthquake recordings are included, with an emphasis on seismically active places, to guarantee a solid dataset. The USGS Earthquake Catalogue API, which provides both historical and real-time earthquake data, is used to retrieve the information. Other geographic data sources, such as satellite images and topographic maps, are included in the dataset to further improve it by capturing structural and environmental elements that affect earthquake hazards.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	net	id	updated
2	2013-04-11T	20.7915	122.226	8.29	4.6	mb	46	115	2.28	1.21	us	usb000g50m	2013-04-11T
3	2013-04-11T	-17.3579	175.0663	9.88	5.3	mb		50	7.802	1.44	us	usb000g4z6	2013-04-11T
4	2013-04-11T	-17.4508	-178.7735	535.92	4.9	mb	21	128	8.56	0.71	us	usb000g4xf	2013-04-11T
5	2013-04-11T	-16.9546	-179.1921	528.46	4.5	mb	45	75	9.06	0.84	us	usb000g4ug	2013-04-11T
6	2013-04-11T	-10.6708	166.0755	167.21	4.7	mb	42	102	6.16	0.8	us	usb000g4su	2013-04-11T
7	2013-04-11T	2.8643	125.4971	64.96	4.6	mb		72	2.794	0.78	us	usb000g4rg	2013-04-11T
8	2013-04-11T	20.9199	122.1061	12.94	4.5	mb	32	130	2.12	0.69	us	usb000g4qu	2013-04-11T
9	2013-04-11T	-2.7939	148.1628	9.39	4.6	mb	23	147	1.09	1.12	us	usb000g4q9	2013-04-11T
10	2013-04-11T	19.2629	95.6948	10.07	5.2	Mwp	75	43	2.14	0.98	us	usb000g4nw	2013-04-11T
11	2013-04-11T	41.6165	141.9924	55.27	4.6	mb	43	123	0.96	1.09	us	usb000g4ni	2013-04-11T
12	2013-04-11T	28.5074	51.6758	10.07	4.8	mb		64	10.83	1.19	us	usb000g4mt	2013-04-11T
13	2013-04-10T	18.854	97.5096	8.27	4.7	mb	29	75	0.63	0.6	us	usb000g4i5	2013-04-11T
14	2013-04-10T	20.8187	122.1203	4.2	5.8	Mww	115	31	2.21	1.28	us	usb000g4ca	2013-04-10T
15	2013-04-10T	2.6017	127.2174	66.02	5	mb	60	105	1.82	1.03	us	usb000g4br	2013-04-10T
16	2013-04-10T	15.5366	-87.228	10	5.5	mb		37	1.471	0.85	us	usb000g4a2	2013-04-11T
17	2013-04-10T	-10.7302	-75.2622	99.62	5.2	mb		72	1.99	0.77	us	usb000g43v	2013-04-10T
18	2013-04-10T	-17.7569	167.7868	10	4.6	mb	20	160	3.71171	0.87	us	us2013nvap	2013-04-10T
19	2013-04-10T	28.5135	51.5523	9.93	4.6			94	10.877	0.97	us	usb000g3y3	2013-04-10T
20	2013-04-10T	28.438	51.738	9.87	5.2	mb	76	58	10.87	1.03	us	usb000g3ts	2013-04-10T
21	2013-04-10T	28.309	51.7514	10.06	4.8	mb		75	10.974	0.95	us	usb000g3t2	2013-04-10T
22	2013-04-10T	37.4728	142.0723	27.79	4.6	mb	40	132	3.23	1.17	us	usb000g3qe	2013-04-10T
23	2013-04-10T	28.45	51.6075	10.02	5.6	mb	76	25	10.91	1.14	us	usb000g3p7	2013-04-10T
24	2013-04-10T	-2.9729	139.0662	55.19	4.8	mb	33	61	6.81	1.6	us	usb000g3ns	2013-04-10T
25	2013-04-10T	28.4814	51.604	10	4.9	mb		139	10.883	0.83	us	usb000g3nn	2013-04-10T
26	2013-04-10T	-2.0824	-79.5666	103.34	4.5	mb	35	113	2.56	0.5	us	usb000g3ng	2013-04-11T
27	2013-04-09T	-22.7541	69.1376	10.2	4.6	mb	17	110	11.11	0.55	us	usb000g3ls	2013-04-09T
28	2013-04-09T	28.2759	51.6754	9.88	4.8	mb		86	11.034	0.68	us	usb000g3ft	2013-04-09T
29	2013-04-09T	5.6129	93.3101	31.21	4.7	mb		139	3.641	0.61	us	usb000g3fe	2013-04-09T
30	2013-04-09T	28.4201	51.6408	19.93	4.6	mb	94	94	10.92	0.72	us	usb000g3dn	2013-04-09T

Figure 1. USGS Earthquakes Dataset Sample

3.2 Dataset Preprocessing

Extensive preprocessing is necessary to eliminate discrepancies and get raw seismic data ready for deep learning models [13]. Before training the models, the noise, missing values, and inconsistencies in the raw seismic data must be fixed. Data cleaning is the first step in the preparation stage. Imputation methods, such as mean replacement for numerical characteristics, are used to manage missing values. One-hot encoding is used to transform categorical information, such as earthquake site descriptions, into numerical representations, and duplicate records are eliminated to avoid model bias.

Numerical characteristics like magnitude, depth, and geographical coordinates are scaled to a similar range through normalisation, which guarantees that each input variable contributes proportionately to model training. Timestamps are transformed into numerical sequences since time-series analysis is essential for earthquake prediction. This enables the LSTM model to efficiently capture temporal relationships. The model can recognise patterns in earthquake-prone areas by converting spatial data, such as latitude and longitude, into grid-based representations for CNN processing.

3.3 Building the CNN Model

Regional seismic patterns are captured by the CNN model, which is used to interpret spatial earthquake data and aid in the prediction of earthquake risk [14]. An organised representation of geolocation data in the form of a grid picture, with each cell encoding earthquake characteristics like depth and magnitude, makes up the CNN's input.

Each grid cell in the CNN model's input layer represents a distinct geographic area with seismic characteristics recorded in it. The grid representation of earthquake sites is organised. The model captures spatial correlations between earthquake events by using several convolutional layers with ReLU activation functions. After every convolutional layer, a max-pooling layer is used to minimise dimensionality while maintaining important features. The overview of the CNN architecture used in this research is illustrated in Figure 2.

In order to ensure robust feature learning and avoid overfitting, dropout layers are used to randomly deactivate a portion of neurons during training. After being flattened, the final output from the CNN layers is sent to a dense layer that is completely linked and uses learnt spatial patterns to make predictions. Mean squared error is used for regression-based risk prediction, while categorical cross-entropy loss is used for classification tasks when training the CNN model.

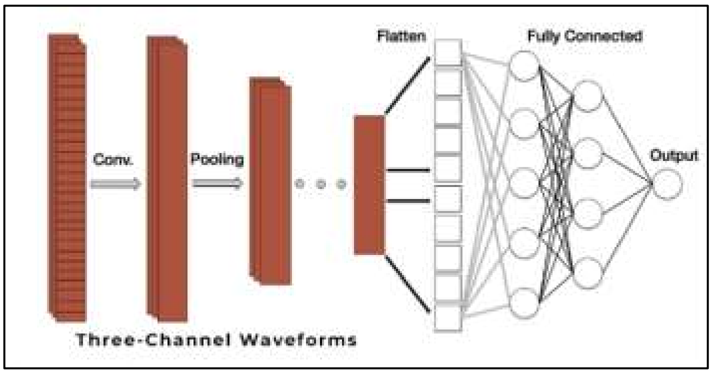

Figure 2. CNN Architecture to Determine Seismic Risk Zones

3.4 Building the LSTM Model

The Long Short-Term Memory (LSTM) model uses past earthquake events to increase prediction accuracy by capturing temporal relationships in seismic activity. LSTM is a good option for modelling time-series earthquake patterns since it can handle sequential data, and earthquake occurrences show temporal relationships.

A series of historical earthquake data with characteristics like magnitude, depth, and timestamps make up the input to the LSTM. Tanh activation functions are used by each of the system's many LSTM layers to capture long-term interdependence in seismic activity. Between LSTM layers, batch normalisation layers are used to stabilise training, guaranteeing that gradients stay constant during backpropagation.

Fully linked dense layers that translate the learnt temporal patterns to seismic risk forecasts are applied to the final output from the LSTM layers. Batch normalisation between bidirectional LSTM layers is used to stabilise learning and speed up convergence in order to enhance model generalization, as demonstrated in Figure 3. Using learnt temporal information, a dense output layer with a SoftMax or linear activation function is utilised to provide earthquake risk predictions.

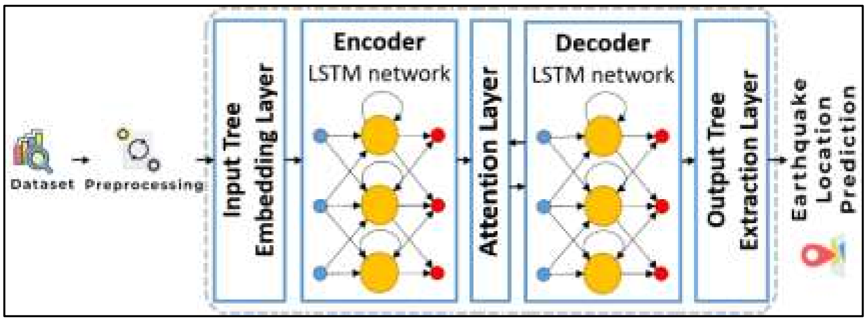

Figure 3. LSTM Architecture to Determine Seismic Risk Zones

3.5 Building the CNN-LSTM Hybrid Model

The CNN-LSTM hybrid model combines LSTM's temporal sequence learning advantages with CNN's spatial feature extraction capabilities. By processing both geolocation information and time-series seismic records, the hybrid architecture provides a more thorough framework for risk prediction.

Sequential earthquake events processed by the LSTM layers and spatial earthquake data handled by the CNN layers are the two input streams that the model uses. The LSTM component records patterns that change over time, while the CNN component retrieves geographical characteristics from the seismic dataset. To learn joint feature representations, the model concatenates the outputs from both networks into a common dense layer as shown in Figure 4.

An output layer that produces earthquake risk probability or magnitude forecasts comes after the last dense layers improve the characteristics that were extracted. To balance the learning of spatial and temporal features, a multi-objective loss function is used to train the hybrid model. The network architecture is optimised by hyperparameter tuning, and early stopping is used to avoid overfitting in order to increase model resilience.

The performance of the CNN-LSTM model is compared to that of solo CNN and LSTM models after extensive testing. Model efficacy is measured using performance indicators including mean absolute error, F1-score, accuracy, and recall. The hybrid model uses both geographical and temporal data to forecast earthquake risk with more accuracy, according to the results.

The potential of deep learning in seismic risk assessment is shown in this research by combining CNN and LSTM into a hybrid framework. For better earthquake prediction, the results emphasise the value of integrating geospatial and time-series analysis, opening the door for further developments in disaster preparation and mitigation techniques.

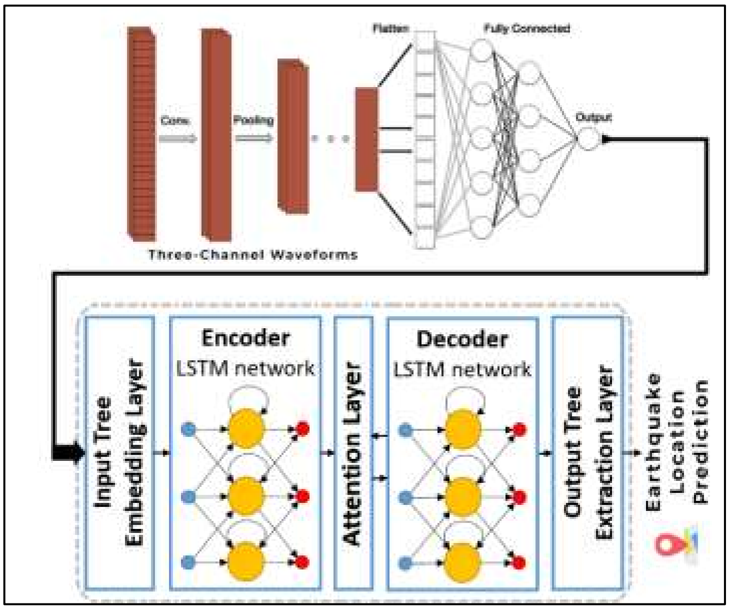

Figure 4. Hybrid CNN-LSTM Architecture to Determine Seismic Risk Zones

4. Results and Discussion

Utilising geolocation data from the United States Geological Survey (USGS), the study's findings show how well the hybrid CNN-LSTM model predicts earthquake risk. To evaluate its performance gains, the CNN-LSTM hybrid model is also contrasted with solo CNN and LSTM models. The findings demonstrate how well the suggested method captures temporal and geographical trends in earthquake occurrences.

The testing dataset includes earthquake occurrences from several geographical locations, enabling a thorough assessment of the model's capacity for generalisation. By examining geolocation variables like latitude, longitude, and closeness to fault lines, the CNN model, which is specifically designed to interpret spatial data, can identify high-risk seismic zones. Its forecasts, however, are restricted to geographical patterns and do not take into account the time dependencies of seismic activity. Because of this, it does a good job of identifying seismic hotspots but is not very accurate at forecasting future occurrences based on past patterns, and achieved an accuracy of 94.35% in training and 94.01% in testing, as seen in Figure 5.

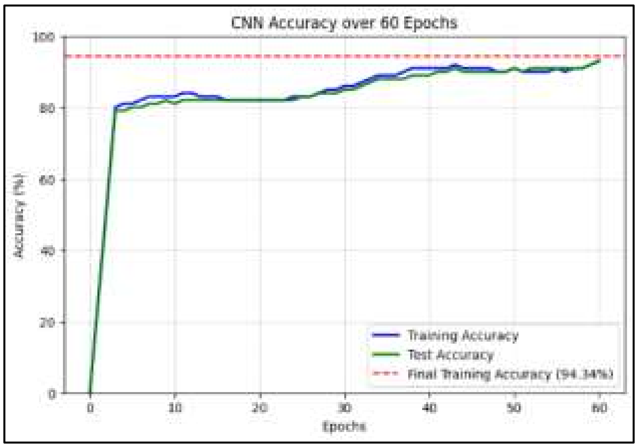

Figure 5. CNN Performance Graph

The temporal relationships between previous and upcoming seismic occurrences are captured by the LSTM model, which was trained only on sequential earthquake data. It effectively learns from past earthquake sequences, seeing trends that suggest the probability of future events. Its geographical precision is decreased by its dependence on time-series data alone, which restricts its capacity to identify regional differences in earthquake risk. The findings demonstrate that while the LSTM model performs very well in time-dependent forecasting, it has trouble successfully integrating geographical data, and achieved an accuracy of 95.17% in training and 94.54% in testing, as seen in the graph in Figure 6.

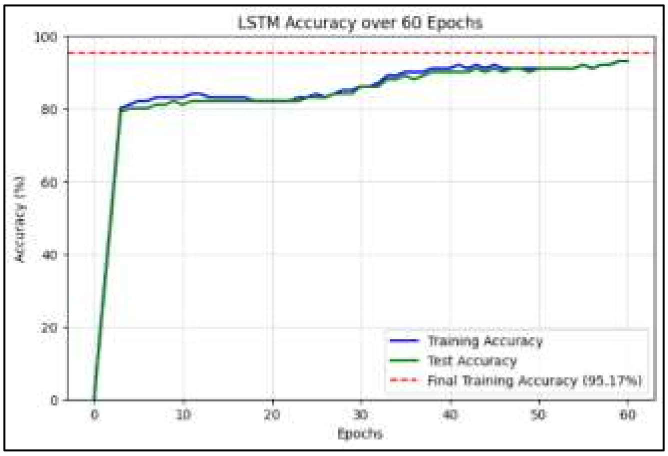

Figure 6. LSTM Performance Graph

These separate drawbacks are addressed by the hybrid CNN-LSTM model, which combines temporal and spatial information into a single prediction framework. While the LSTM component analyses the sequence of seismic events to forecast the probability of future occurrences, the CNN component processes the geographical features of earthquakes to identify high-risk locations. When these two systems are combined, earthquake risk prediction accuracy and dependability are greatly increased.

The hybrid model outperforms the solo CNN and LSTM models in terms of accuracy with 98.45% accuracy in training and 97.61% in testing as shown in Figure 7, according to quantitative performance study. According to the accuracy and recall values, the hybrid technique produces more reliable seismic risk assessments by lowering false positives and false negatives. The model's robustness is further supported by the RMSE and MAE scores, which show reduced error rates in magnitude predictions when compared to single deep learning and conventional machine learning models.

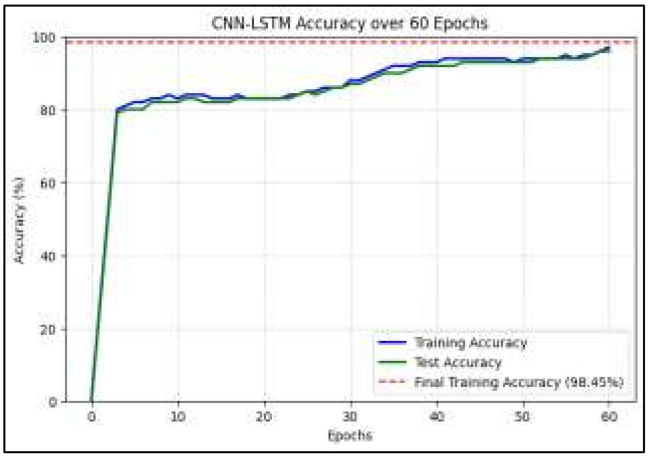

Figure 7. CNN-LSTM Performance Graph

The CNN-LSTM model provides more accurate risk evaluations, especially in areas with high seismic activity, according to a comparison of projected and real earthquake occurrences. The model's capacity to examine both temporal and spatial trends guarantees that forecasts are grounded in the fundamental geological variables affecting earthquake hazards as well as past data patterns. Furthermore, the model's adaptability is validated by real-time testing using recent seismic events, showing its efficacy in dynamic earthquake forecasting.

The results also demonstrate the model's scalability, since tests reveal that it can retain a high level of predicted accuracy even when trained on bigger datasets that span many decades. The CNN-LSTM model is a useful tool for assessing the risk of earthquakes worldwide since it generalises well across various geographical areas. Additionally, the way it integrates with real-time data sources, such IoT-based seismic sensors, shows how it may be used in early warning systems to help with disaster preparation and mitigation.

Overall, the findings demonstrate that by using the advantages of both spatial and temporal data, the hybrid CNN-LSTM model greatly improves earthquake risk prediction. It is a potent instrument for seismic risk assessment because to its exceptional accuracy, precision, memory, and error reduction capabilities. According to the research, combining deep learning with geolocation data has the potential to completely transform earthquake forecasting and provide insightful information for disaster risk management, emergency response plans, and urban planning.

5. Conclusion

This study offers a hybrid learning strategy that combines Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs) to improve earthquake risk prediction using geolocation data. The research effectively illustrates how merging geographical and temporal characteristics may increase forecast accuracy by using the United States Geological Survey (USGS) dataset. According to the findings, the hybrid CNN-LSTM model offers a more thorough and trustworthy framework for evaluating earthquake risk than solo CNN and LSTM designs. The model improves forecasting skills by using CNN to record regional seismic patterns and LSTM to learn previous trends, providing an accuracy of 98.45% in training. This work emphasises how hybrid deep learning models may revolutionise seismic risk assessment and how useful they are for real-time decision-making, urban planning, and catastrophe preparation. To increase forecast reliability, future initiatives include adding worldwide datasets to the model, including more environmental parameters, and resolving issues with data quality.

References

[1] Krichen, Moez, Mohamed S. Abdalzaher, Mohamed Elwekeil, and Mostafa M. Fouda. "Managing natural disasters: An analysis of technological advancements, opportunities, and challenges." Internet of Things and Cyber-Physical Systems 4 (2024): 99-109.

[2] Gürsoy, Güneş, Asaf Varol, and Ahad Nasab. "Importance of machine learning and deep learning algorithms in earthquake prediction: a review." In 2023 11th International Symposium on Digital Forensics and Security (ISDFS), pp. 1-6. IEEE, 2023.

[3] Cicerone, Robert D., John E. Ebel, and James Britton. "A systematic compilation of earthquake precursors." Tectonophysics 476, no. 3-4 (2009): 371-396.

[4] Rikitake, Tsuneji. "Earthquake precursors." Bulletin of the Seismological Society of America 65, no. 5 (1975): 1133-1162.

[5] Rikitake, Tsuneji. "Earthquake precursors in Japan: precursor time and detectability." Tectonophysics 136, no. 3-4 (1987): 265-282.

[6] Chen, Yun-Tai, Hao-Ding Gu, and Zao-Xun Lu. "Variations of gravity before and after the Haicheng earthquake, 1975, and the Tangshan earthquake, 1976." Physics of the Earth and Planetary Interiors 18, no. 4 (1979): 330-338.

[7] Brodsky, Emily E., and Thorne Lay. "Recognizing foreshocks from the 1 April 2014 Chile earthquake." Science 344, no. 6185 (2014): 700-702.

[8] Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in neural information processing systems 30 (2017).

[9] Colombelli, Simona, Aldo Zollo, Gaetano Festa, and Matteo Picozzi. "Evidence for a difference in rupture initiation between small and large earthquakes." Nature Communications 5, no. 1 (2014): 3958.

[10] Rydelek, Paul, and Shigeki Horiuchi. "Is earthquake rupture deterministic?." Nature 442, no. 7100 (2006): E5-E6.

[11] Melgar, Diego, and Gavin P. Hayes. "Systematic observations of the slip pulse properties of large earthquake ruptures." Geophysical Research Letters 44, no. 19 (2017): 9691-9698.

[12] Toumi, Abdelmalek, Jean-Christophe Cexus, Ali Khenchaf, and Mahdi Abid. "A Combined CNN-LSTM Network for Ship Classification on SAR Images." Sensors 24, no. 24 (2024): 7954.

[13] Taye, Mohammad Mustafa. "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions." Computers 12, no. 5 (2023): 91.

[14] S, Siddharth. (2023). Analysis and Prediction of Earthquake Impact using Convolutional Neural Networks. International Journal for Research in Applied Science and Engineering Technology. 11(09). 750-753.

[15] Magrini, Fabrizio, Dario Jozinović, Fabio Cammarano, Alberto Michelini, and Lapo Boschi. "Local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale." Artificial Intelligence in Geosciences 1 (2020): 1-10.