DATA DRIVEN EARTHQUAKE PREDICTION USING INTELLIGENT TECHNIQUES

Muhammad Ibrahim Lodhi
01-249241-007

Dr Saba Mahmood

A thesis submitted in fulfilment of the requirements for the award of degree of Masters of Science (Computer Science)

Department of Computer Science

BAHRIA UNIVERSITY ISLAMABAD

MARCH 2026

Approval of Examination

Scholar Name: Muhammad Ibrahim Lodhi
Registration Number: 44556
Enrollment: 01-249241-007
Program of Study: MS-Data Science
Thesis Title: Data Driven Earthquake Prediction using Intelligent Techniques

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index 9 %. that is within the permissible limit set by the HEC for the MS/M.Phil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/M.Phil thesis.

Principal Supervisor Name: Dr Saba Mahmood
Principal Supervisor Signature:
Date:

Author's Declaration

I, Ibrahim Lodhi hereby state that my MS/M.Phil thesis titled is my own work and has not been submitted previously by me for taking any degree from Bahria university or anywhere else in the country/world. At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS/M.Phil degree.

Name of Scholar:Ibrahim Lodhi

Date: 07/01/2026

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled DATA DRIVEN EARTHQUAKE PREDICTION USING INTELLIGENT TECHNIQUES is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me. I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.
I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS/M.Phil degree, the university reserves the right to withdraw / revoke my MS/M.Phil degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Name of Scholar: Ibrahim Lodhi

Date: 07/01/2026

Acknowledgements

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Professor Dr. SABA MAHMOOD, for encouragement, guidance, critics and friendship.. Without her continued support and interest, this thesis would not have been the same as presented here.
Librarians at Bahria University also deserve special thanks for their assistance in supplying the relevant literatures. My fellow postgraduate students should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members.

Abstract

Earthquakes represent one of the most devastating natural hazards, claiming thousands of lives and causing trillions in economic damage annually, particularly in seismically active regions like the Himalayan arc and Pacific Ring of Fire . Despite advances in seismic monitoring, accurate prediction of earthquake magnitude the key determinant of potential destruction remains elusive due to the nonlinear, chaotic dynamics of fault systems and the rarity of large events in historical catalogs.

In conventional probabilistic seismic hazard analysis, long term risk predictions can be made, but reliable short term predictions are not possible, which makes society prone to abrupt ruptures. This thesis works to fill this crucial research gap by proposing and developing a new deep learning model to predict magnitude based on multi modal precursor

Although there have been some recent attempts to apply machine learning to seismicity forecasting, there have been several models still using simple regression models or simple neural networks, which are not capable of handling spatiotemporal correlations effectively. Although models such as random forest approaches have had moderate success at the regional scale, they also tend to neglect long term temporal correlations and lack generability across different tectonic settings.

Although there have been some promising approaches using transformers for time series data, there have been no attempts to apply these models to seismicity forecasting, mainly due to their high complexity and the need for specific adaptations. These models, although successful, also emphasize the need to develop models that can integrate heterogeneous features of seismicity, such as b value anomaly, ionospheric TEC, and waveform magnitude, effectively, which can also provide probabilistic forecasting.

In light of these challenges, this thesis develops the approach for earthquake magnitude prediction using modern deep learning and making a particular focus on: revealing temporal dynamics and predictive uncertainty. It focuses on an enhanced Temporal Fusion Transformer (TFT) architecture combining bidirectional LSTMs, GRUs, multi head attention, and gating mechanisms that combine past observations with simple future covariates and static spatial contextual information into a single unified forecasting model.

To overcome the small size and imbalance in conventional seismic catalogs, this study constructs an augmented dataset with a model based synthetic data generator and relies on the assumption that the added records preserve the statistical and spatial properties of the original events. The original data and augmented are subjected to extensive feature engineering comprising lagged magnitudes, rolling statistics, inter-

action terms, and cluster based spatial descriptors, followed by their transformation into supervised sequences to model time series.
The TFT model is trained using a quantile loss function to produce probabilistic forecasts at multiple quantiles, allowing the derivation of both point estimates and prediction intervals for earthquake magnitude. The training procedure incorporates regularization techniques such as dropout, weight decay, and gradient clipping, together with adaptive learning rate scheduling and validation based check pointing to promote stable convergence and generalization. Model performance is evaluated on a temporally held out test set using root mean squared error (RMSE), mean absolute error (MAE), and average quantile loss, alongside visual diagnostics such as prediction truth scatter plots, residual analysis, and time series overlays.
In addition, Monte Carlo dropout is employed at inference time to form an ensemble of stochastic predictions, providing an estimate of epistemic uncertainty and enabling a comparison between single model and ensemble behavior.
The results indicate that the proposed TFT architecture, combined with carefully designed augmentation and feature engineering, can achieve low prediction errors while generating well calibrated prediction intervals for normalized earthquake magnitudes. Spatial comparison plots show that the synthetic events closely follow the geographic distribution of the original Catalog, suggesting that the augmented dataset is suitable for training without distorting the underlying seismic patterns.
Overall, the thesis demonstrates that attention based sequence models, when paired with realistic data augmentation and rigorous evaluation, offer a promising direction for probabilistic earthquake magnitude forecasting and for quantifying the uncertainty inherent in such predictions

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE