

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

AUTHOR'S DECLARATION	ii
PLAGIARISM UNDERTAKING	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF TABLES	xi
LIST OF FIGURES	xii
1 INTRODUCTION	2
1.1 Context and Motivation	2
1.2 Temporal Fusion Transformers for Earthquakes	4
1.3 Current State and Research Gaps	5
1.4 Thesis Objectives and Research Scope	7
1.4.1 Problem Statement	7
1.4.2 Research Questions	7
1.4.3 Primary Objectives	8
1.5 Scope and Limitations	8
1.6 Broader Significance and Implications	10
1.7 Chapter Summary	11
2 RELATED WORK	13
2.1 Review Protocol	13
2.2 Introduction to Earthquake Prediction Research	14
2.2.1 Shift to data driven deep learning	14
2.3 Seismological Foundations and Statistical Models	16
2.3.1 Tectonic Processes and Point Process Frameworks	16
2.3.2 Magnitude Scales and Contemporary Measurement Chal- lenges	18
2.4 Machine Learning and Deep Learning Approaches	20

2.4.1	Feature Engineering and Classical Machine Learning . .	20
2.4.2	Hybrid CNN LSTM and RNN Architectures	21
2.4.3	Transformer Architectures for Seismic Applications . . .	23
2.5	Temporal Fusion Transformers: Architecture, Applications, and Seismological Potential	25
2.5.1	Core Architecture and Design Principles	25
2.6	Applications to Induced Seismicity and Structural Health Mon- itoring	27
2.7	Potential for Natural Tectonic Earthquake Magnitude Prediction	28
2.7.1	Temporal dependencies	28
2.7.2	Feature selection process in TFT	28
2.8	Critical Analysis and Research Gaps	29
2.8.1	Insufficient Systematic Application to Tectonic Magni- tude Prediction	29
2.8.2	Limited Integration of Multimodal Geophysical Data . .	30
2.8.3	Inadequate Emphasis on Probabilistic Multi Horizon Fore- casting and Calibration	31
2.8.4	Need for Interpretable, Seismologically Consistent Models	32
2.8.5	Conclusion and Motivation for TFT Based Magnitude Prediction	33
2.9	Chapter Summary	36
3	METHODOLOGY	41
3.1	Introduction	41
3.2	DATASET	44
3.2.1	Data Attributes	44
3.2.2	Spatial Distribution of Original and Augmented Earth- quakes	44
3.3	Data Augmentation	47
3.3.1	Overview	47
3.3.2	Synthetic Data Generation with SDV	47
3.4	Role of Augmented Data in the Overall Pipeline	48
3.5	Data Preprocessing and Advanced Feature Engineering	51
3.5.1	Data Ingestion and Cleaning Protocols	54
3.5.2	Spatial Clustering and Dimensionality Reduction	54
3.5.3	Temporal Decomposition and Normalization	55
3.5.4	Global Feature Scaling	55
3.5.5	Enhanced Feature Engineering	56
3.5.6	Sequence Construction for Temporal Supervision	56
3.6	Temporal Fusion Transformer Architecture	56

3.6.1	Static Feature Encoder	59
3.6.2	Past Feature Encoder (Bidirectional LSTM)	60
3.6.3	Future Feature Encoder (GRU)	61
3.6.4	Multi Head Attention Mechanism	62
3.6.5	Static Context Expansion and First Gating Mechanism	64
3.6.6	Second Gating Mechanism and Layer Normalization	65
3.6.7	Output Projection Network	65
3.6.8	Quantile Regression Output Layers	66
3.6.9	Summary of Component Roles	67
3.7	Training Methodology and Optimization	67
3.7.1	Quantile Loss Function	68
3.7.2	Optimization Algorithm: AdamW	69
3.7.3	Learning Rate Scheduling: ReduceLROnPlateau	70
3.7.4	Gradient Clipping and Numerical Stability	71
3.7.5	Regularization Strategies	71
3.7.6	The Complete Training Loop	71
3.7.7	Batch Processing and Device Management	72
3.7.8	Convergence Criteria and Early Stopping	72
3.7.9	Summary of Training Configuration	72
3.8	Evaluation Metrics and Model Assessment	72
3.8.1	Root Mean Squared Error (RMSE)	74
3.8.2	Mean Absolute Error (MAE)	75
3.8.3	Quantile Loss (Pinball Loss)	76
3.8.4	Summary of Metrics Strategy	77
3.9	Ensemble Prediction Strategy via Monte Carlo Dropout	77
3.9.1	Theoretical Foundation of Monte Carlo Dropout	78
3.9.2	Implementation Details	79
3.9.3	Ensemble Statistics Computation	79
3.9.4	Comparison: Single Model vs. Ensemble Predictions	80
3.9.5	Computational Cost and Efficiency	80
3.9.6	Empirical Validation of Ensemble Uncertainty	80
3.9.7	Interpretation in Seismic Context	80
3.9.8	Summary of Ensemble Strategy	81
3.10	Data Partitioning and Time Series Cross Validation	81
3.10.1	The Problem with Standard Cross Validation on Time Series	81
3.10.2	Appropriate Time Series Cross Validation: Walk Forward Validation	82
3.10.3	Single Hold Out Test Set (Simplified Strategy)	82

3.10.4	Validation Set: Purpose and Usage	83
3.10.5	Test Set: Final and Rigorous Assessment	83
3.10.6	Handling Sequence Overlap and Temporal Coherence . .	84
3.10.7	Time Series Specific Considerations	84
3.10.8	Potential Limitations and Mitigations	84
3.10.9	Summary of Partitioning Strategy	84
3.11	Summary of the Chapter	84
3.11.1	Data Augmentation Strategy	85
3.11.2	Feature Engineering Process	85
3.11.3	Model Architecture	85
3.11.4	Uncertainty Quantification	85
3.11.5	Evaluation Protocol	86
4	RESULTS AND DISCUSSION	87
4.1	Introduction	87
4.2	Explicit Analysis of Experimental Results	87
4.2.1	Performance Summary	87
4.2.2	Ensemble vs. Single Model Analysis	88
4.3	Regional Data Context and Scalability	89
4.3.1	Geographical Scope of Comparative Baselines	89
4.4	Comparative Analysis of Evaluation Metrics	89
4.4.1	Discussion of Findings	90
4.5	Technical Justification for Superiority	90
4.6	Summary	91
5	CONCLUSION & FUTURE WORK	92
5.1	Overall Contributions	92
5.2	Practical Implications	93
5.3	Limitations	93
5.4	Future Work	94
5.4.1	Multi Region and Transfer Learning	94
5.4.2	Integration with Real Time Early Warning Constraints .	95
5.4.3	Hybrid Architectures and Spatial Information	95
5.4.4	Uncertainty Quantification and Risk Aware Metrics . . .	95
5.4.5	Domain Informed Features and Physical Constraints . .	96
5.5	Closing Remarks	96
	REFERENCES	98

LIST OF TABLE

2.1	Comprehensive Comparison of Earthquake Magnitude Prediction Studies	37
3.1	Summary of Training Hyperparameters and Configuration . . .	73
3.2	Summary of Ensemble Prediction Configuration	81
3.3	Model State and Configuration across Data Splits	84
3.4	Summary of Data Partitioning Strategy	85
4.1	TFT single and ensemble model performance.	88
4.2	Geographical Scope and Methodological Focus of Comparative Baselines	89
4.3	Comparative performance across evaluation metrics on earthquake magnitude prediction.	89

LIST OF FIGURE

1.1	Title: Earthquakes mechanism, Source: Maximum, Inc. (2023). Earthquakes explained: Types, causes, and aftereffects. Maximum Learning Center	2
3.1	The whole overview of the methodology	42
3.2	Data Augmentation pipeline	43
3.3	Earthquake Locations - Real Data	45
3.4	Earthquake Locations - Augmented Data	46
3.5	The synthetic dataset accurately models human perceived earthquake intensity, preserving key statistical properties of the CDI feature.	49
3.6	The synthetic data preserves the original time based characteristics of earthquake events, indicating high quality temporal augmentation. Updated simply refers to last updated timestamp for the event.	49
3.7	KDE plot illustrating the distribution alignment between real and synthetic "time" features, showing the model's ability to capture the bimodal characteristics of the original dataset. . . .	50
3.8	KDE plot comparing the distribution of real and synthetic "magnitude" values, demonstrating the model's ability to replicate the bimodal frequency of seismic events.	50
3.9	Magnitude distribution comparison between real and synthetic datasets via histogram KDE overlay and box plot, highlighting consistent medians and the model's performance in replicating seismic event intensity clusters	51
3.10	Pearson correlation matrices comparing inter feature relationships in real and synthetic seismic datasets	52
3.11	Pearson correlation matrices comparing inter feature relationships in real and synthetic seismic datasets	53

3.12	Proposed TFT inspired architecture for earthquake magnitude prediction. The model integrates static features with past and known future seismic sequences using recurrent encoders, multi head attention, and gated fusion, producing probabilistic forecasts via quantile regression.	58
3.13	Multi head attention mechanism	63
5.1	Plagiarism report	97