Earthquake Magnitude Prediction Deep Learning Algorithm

Chloe Sturgeon '24, Curtis Sloley '24
Professor Lin Cheng, Professor Deborah A. Fixel
Trinity College Engineering Department

PROBLEM DEFINITION AND BACKGROUND

An increase in incidences of earthquakes in the global context of climate change underscores an urgent need for effective earthquake mitigation strategies. **Earthquakes pose significant risks to infrastructure and lives, demanding more precise and reliable methods for understanding and predicting their impact.** Despite advances in seismic technology and geological studies, the ability to predict earthquake magnitudes accurately and in a timely manner remains a profound challenge in the field of earthquake engineering and seismology. **Current technologies and models focus predominantly on the detection and recording of seismic activities, leaving a gap in the post-event analysis and real-time predictive capabilities.** Most existing models are either too specialized to handle diverse data types or not robust enough to provide precise magnitude estimations, which are crucial for effective emergency responses and preparedness planning.

In response to this, our project aims to enhance the analytical capabilities of seismic data analysis through the application of machine learning techniques.

DESIGN REQUIREMENTS

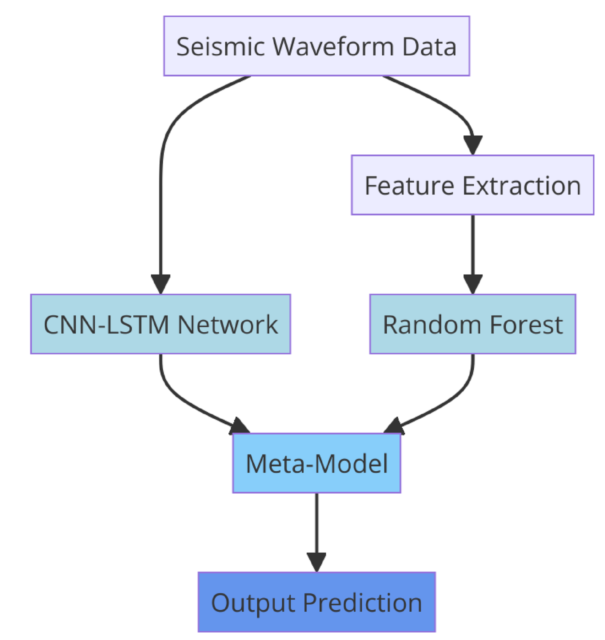
Accuracy: The model must achieve a mean absolute error (MAE) of less than 0.25 when predicting earthquake magnitudes, ensuring high reliability in its predictions.
Data Handling: Capable of processing both raw time series waveform data and structured features extracted from the waveforms.
IEEE Standard for Floating-Point Arithmetic (IEEE 754): Ensure all computational operations adhere to this standard for precision and consistency in calculations.
ISO/IEC 25012:2008 (Data Quality): Comply with data quality standards for accuracy, completeness, consistency, and credibility, essential for the reliability of earthquake predictions.
ISO/IEC 27001: Information Security Management: Adhere to information security management standards to safeguard seismic data.

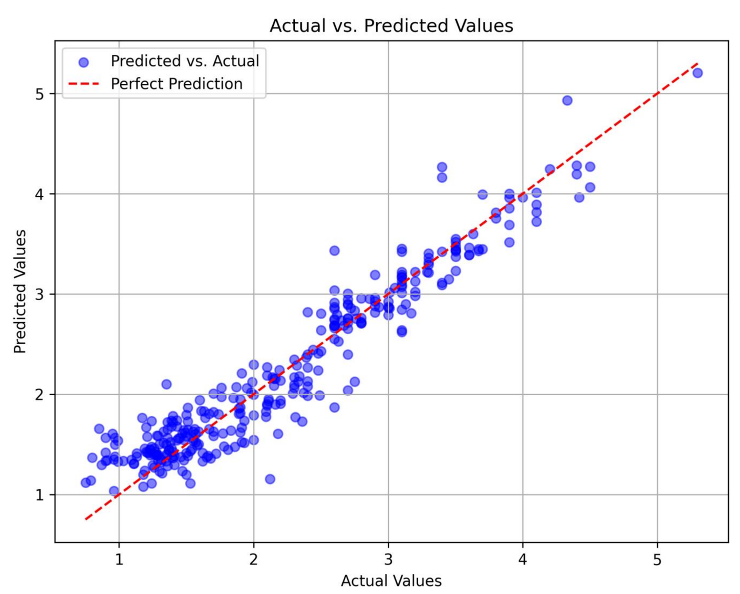
DESIGN ALTERNATIVES

Algorithm Selection: We evaluated four different algorithms based on their suitability for handling seismic data:
1.Random Forest (RF): Known for its effectiveness in regression tasks and ability to handle high-dimensional data. It provides good interpretability and robust performance without extensive parameter tuning.
2.Convolutional Neural Network (CNN): Highly effective in spatial and temporal data analysis, making it suitable for waveform data which has inherent spatial-temporal characteristics.
3.Support Vector Machine (SVM): Renowned for its robustness in classification tasks and its effectiveness in high-dimensional spaces, though typically less used for regression in complex, noisy data environments like seismic data.
4.Long Short-Term Memory (LSTM): An advanced type of recurrent neural network, ideal for time series data due to its ability to remember long-term dependencies, crucial for the temporal nature of seismic waveforms.

Data Input Selection: We considered two types of input data:
1.Time Series Waveform Data: Direct use of raw seismic waveforms, which preserves all original temporal and amplitude information, potentially enabling more nuanced detection of seismic features.
2.Extracted Features: Using statistical and spectral features extracted from the waveform data, which could simplify the model training and focus on the most informative attributes of the data.

FINAL DESIGN AND IMPLEMENTATION





Model Architecture: We adopted a hybrid approach combining a Long Short-Term Memory (LSTM) network and a Random Forest regression model. This ensemble strategy allowed us to leverage the strengths of both models: the LSTM for capturing temporal dependencies in raw time series waveform data, and the Random Forest for handling extracted features from the same data, enhancing prediction accuracy.
Data Preprocessing: We preprocessed the seismic waveform data by cleaning and filtering out noise, and then extracting relevant features such as spectral and statistical characteristics. This preprocessing step was crucial for preparing the data for input into both the LSTM and Random Forest models.
Data Splitting: We adopted a 70-15-15 training, testing, validation data split for a balanced distribution of data that could be used for unbiased model evaluation.
Final MAE on test data set: 0.21171509433962274

DESIGN EVALUATION AND ITERATIVE PROCESS

Evaluation Criteria: For evaluation of which machine learning models best fit our project requirements we primarily focused on two metrics:
1.Mean Average Error (MAE): Measures the average magnitude of the errors in a set of predictions, without considering their direction.
2. Model Timing: This involved measuring how long it takes for each model to train and then to predict new data points. This metric is crucial for our application since the ability to process data rapidly is essential for timely earthquake warnings.

Iterative Process:

- Data selection: Decided on raw seismic waveform data. Considered multiple seismic data sources, based on data format and quality decided on Sanford Earthquake Dataset (STEAD)
- Goal Adjustment: Initially aimed at predicting future earthquakes using historical seismic data, we found this goal too ambitious given our time and expertise constraints and refined our project to predicting the magnitude of recorded seismic waves.
- Model Selection: Trained all algorithm options and compared MAE and training/testing times. Decided a hybrid CNN-LSTM deep learning model was most suitable for this data format and task.
- Model Tuning: Fine tuned hyperparameters and added/removed layers based on their effect on MAE and model timing. Then measured MAE of the final CNN-LSTM model.
- Re-evaluation: Considered other options for improving the accuracy of our model. We decided that the nature of the raw seismic waveforms didn't give us enough information.
- Data selection: To get more important information out of the seismic data and improve accuracy we explored features that could be extracted from the waveforms and compiled them in a CSV.
- Algorithm alteration: To train on the new data format we again trained each of the algorithm options and decided on the most accurate and fastest one for this data, the RF model.
- Feature evaluation: Using permutation importance with training the RF model we sorted features by importance and dropped the least important ones.
- Data refinement: To further increase accuracy and computational efficiency we filtered training data to the Los Angeles area. The intention is that with transfer learning this trained model can be trained on data from more regions.
- Ensemble learning hybridization: Wanting to retain insights from both models and forms of data, we applied an ensemble learning stacking technique to generate the final output predictions.

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

The Hybrid CNN-LSTM Random Forest Model developed in this project significantly advances seismic data analysis by accurately determining earthquake magnitudes from recorded events, laying a crucial foundation for enhancing real-time earthquake monitoring and predictive modeling. The LSTM model effectively harnessed the temporal dynamics inherent in time series waveform data, while the Random Forest model leveraged structured features extracted from the same data to enhance prediction accuracy. This hybrid approach capitalized on the strengths of both models, addressing the complex nature of seismic data, which includes non-linear patterns and significant noise levels.
With more time, this project can be expanded to predict future earthquakes by incorporating real-time data and historical patterns and using adaptive learning to refine predictions based on new seismic activity.

REFERENCES

[1] STEAD (Stanford Earthquake Dataset), https://web.stanford.edu/group/riskin/docs/STEAD/
[2] IEEE Standard for Floating-Point Arithmetic (IEEE 754), IEEE Std 754-2008, IEEE, New York, NY, USA, 2008.
[3] ISO/IEC 25012:2008 (Data Quality), International Organization for Standardization, Geneva, Switzerland, 2008.
[4] ISO/IEC 27001 (Information Security Management), International Organization for Standardization, Geneva, Switzerland.