

# ML Lab Research Proposal

## Quake-GNN: Multimodal Graph Neural Network for High-Reliability Earthquake Prediction

**Team:** Team B - Arya Addagarla & Anubhav Chatterjee  
**Date:** April 24, 2026  
**Project:** Quake-GNN: Multimodal Graph Neural Network for High-Reliability Earthquake Prediction  
**Target Compute:** 1x AMD Instinct MI300X GPU, approx. 2 weeks

---

### 1. PROBLEM STATEMENT

We are addressing the unpredictability of earthquakes and how incorporating sensor data and analyzing it with neural network-based predictive models can yield high-reliability predictions. This matters because living in California, a place with relatively high seismic activity, keeps civilians at greater risk of the next deadly earthquake.

### 2. PROPOSED APPROACH

We are using a Multimodal Graph Neural Network. This treats every seismic station, GPS point, and fiber-optic cable as a connected node in a massive web, similar to how the brain connects different senses. We chose this over traditional methods because it doesn't just see a line from data; instead, it tracks how that vibration moves across a map.

By requiring the signal to travel at the actual speed of a tectonic wave, we can instantly tell the difference between a real earthquake and a truck driving past a single sensor.

### 3. DATASET

We are using a multi-stream physics data repository incorporating explicit spatial relationships:

- Raw vibrations from the **Stanford Earthquake Dataset**.
- Satellite movement logs from active **GPS stations**.
- High-speed telemetry from localized **fiber-optic cable arrays**.

#### Preprocessing

- Use the R programming language to explore and clean raw data streams.
- Enforce strict temporal synchronization to ensure every data point matches a shared universal clock.
- Construct graph adjacency matrices based on geographical coordinates of sensor nodes.

### 4. EVALUATION METRICS

Experiment Category	Metric	Target / Interpretation
Predictive Accuracy	Overall Percentage Accuracy	Success threshold established at > 90% accuracy
Anomalous Identification	F1-Score	Balances system sensitivity to small quakes while avoiding false alarms ('crying wolf')
Spatio-Temporal Localization	Epicenter distance error & Warning lead time	Quantifies the extra warning time delivered before shaking starts

## 5. COMPUTE REQUIREMENTS

To handle this massive amount of data, we utilize a high-end accelerator configuration. This specialized hardware allows the AI to think about multiple data streams at the same time rather than one by one.

Phase	Hardware	Estimated Wall-Clock	Notes / Purpose
Data Engineering & Exploration	CPU Core Allocation	~4-6 hours	Math exploration and matrix tuning using R
Multimodal GNN Training	1x AMD Instinct MI300X GPU	~48 hours	Heavy AI lifting using PyTorch graph loaders

### Software Dependencies

- R environment for initial data cleaning and cleaning analytics.
- PyTorch framework core + PyTorch Geometric or related extensions for graph operations.
- ROCm execution environment configured for AMD MI300X hardware allocation (minimum 24GB VRAM target).

## 6. EXPECTED DELIVERABLES

- A functional Multimodal Graph Neural Network model optimized for seismic time-series analysis.
- An exploratory data analysis pipeline implemented in R for parsing structural sensor feeds.
- An end-to-end training and inference pipeline mapping live sensor streaming to real-time warning indices.

## 7. RISK & MITIGATION

Risk	Likelihood	Mitigation
Discovery of illusory correlations when tracking patterns in highly noisy data streams	Medium	Run the AI to isolate real predictors, then perform cross-validation against historic real data blocks to ensure correlation validity.
Regular ambient tremors (e.g., localized vehicle traffic noise) contaminating input sensors	High	Incorporate physical speed propagation constraints; require signal tracking to match true tectonic wave velocities to filter out non-seismic anomalies.