Capstone Project Concept Note and Implementation Plan

Project Title: Machine Learning for Assessing Soil Liquefaction Risk in Seismic Zones

Team Members

1. Wisam Elkamali Dafaalla Mohamed

Concept Note

1. Project Overview

Liquefaction, a process in which soil loses its strength and behaves like a liquid, is a secondary effect of earthquake shaking. It can cause structural settlement, sinking, or floating, posing a significant threat to people's lives and causing substantial financial losses. Given the damage to infrastructure caused by liquefaction, which threatens lives and causes economic losses, it is imperative to conduct more research to mitigate its disastrous consequences.

This project is directly relevant to SDG 11 (Sustainable Cities and Communities) as it contributes to creating safer and more resilient urban environments. By utilizing machine learning techniques to predict soil liquefaction susceptibility after earthquakes, this research aims to mitigate the risks associated with liquefaction, thus contributing to disaster risk reduction and creating more sustainable and resilient cities and communities.

2. Objectives

This research aims to address the challenges of liquefaction geotechnically by leveraging machine learning (ML) techniques. By analyzing historical earthquake data and geotechnical parameters, ML can predict liquefaction susceptibility with greater accuracy, facilitating informed decision-making in foundation design and soil improvement. This approach holds promise for minimizing the detrimental effects of liquefaction-induced damage on infrastructure.

3. Background

While the term "liquefaction" was coined relatively recently, its conceptual roots trace back to early attempts to understand soil behavior under shear stress. However, the complexity of the phenomenon, coupled with variations in critical void ratios and loading conditions, necessitates further investigation.

The urgency of studying liquefaction became apparent after catastrophic earthquakes, such as those in Niigata, Alaska, Loma Prieta, Kobe, Kocaeli, and Chi-Chi. These events resulted in widespread damage to infrastructure, including buildings, embankments, slopes, and bridges, highlighting the dire need for research to mitigate liquefaction-induced hazards.

The recent Kahramanmaraş-Türkiye Earthquakes (Figure 1) further underscored the destructive potential of liquefaction, with observed manifestations including soil boils, lateral spreading, excessive settlements, bearing capacity failures, and building collapses. Given the substantial threats posed by liquefaction to both lives and infrastructure, intensified research efforts are imperative to develop effective mitigation strategies.[1]

Figure 1 An extreme case of liquefaction-induced bearing capacity failure and toppling of a residential building [1]

The evaluation of liquefaction potential is a critical aspect of earthquake geotechnical engineering, involving a systematic approach to address three fundamental questions: Is the soil susceptible to liquefaction? If so, will liquefaction be triggered? And if triggered, will damage occur? If the soil is found to be non-susceptible, the evaluation concludes that liquefaction hazards do not exist. However, if the soil is deemed susceptible, further assessment is necessary.

The evaluation of liquefaction is a complex process dependent on various factors such as soil properties, geological and historical considerations, the type of loading, and the frequency and duration of shaking. Depending on these factors, several methods are employed for evaluation, including field methods, laboratory methods, analytical methods, numerical methods, GIS methods, and AI methods [2].

Field methods typically involve techniques such as Standard Penetration Test (SPT), Cone Penetration Test (CPT), or shear wave velocity measurements. Laboratory methods include tests such as the cyclic triaxial test, cyclic direct simple shear test, centrifuge modeling, and shake table tests. Analytical methods utilize models for analysis and propose empirical equations incorporating different parameters. On the other hand, numerical methods are based on numerical simulations.

Once earthquake loading and liquefaction resistance have been characterized, the evaluation of liquefaction potential involves comparing earthquake loading with liquefaction resistance throughout the soil deposit of interest. Liquefaction can be expected at depths where the loading exceeds the resistance or when the factor of safety against liquefaction is less than 1.

The liquefaction factor of safety (FS) can be expressed as the ratio of the cyclic stress required to cause liquefaction (τcycL) to the equivalent cyclic shear stress induced by an earthquake (τcyc), which is also termed as the cyclic resistance ratio (CRR) to the cyclic stress ratio (CSR).

$FS=\frac{\tau cycL}{\tau cyc}=\frac{\mathrm{CRR}}{\mathrm{CSR}}$	(1)

Geotechnical engineering, which focuses on the behavior of Earth materials and their interaction with structures, traditionally relies on theoretical models and empirical methods. However, the inherent complexity and variability of geological formations often limit the accuracy of predictions. The integration of artificial intelligence (AI), particularly ML and deep learning (DL), presents a transformative opportunity to overcome these challenges.

In recent decades, AI has become increasingly indispensable across various engineering disciplines, including geotechnical engineering. Its applications span diverse sub-disciplines such as frozen soils, rock mechanics, slope stability, foundations, tunnels, dams, and unsaturated soils. By harnessing the power of AI, geotechnical engineers can enhance their understanding of soil behavior and improve the resilience of infrastructure against seismic hazards like liquefaction.[3]

4. Methodology

Deep learning, a subset of artificial intelligence, utilizes neural networks with multiple layers to extract intricate patterns from large datasets. Its purpose in soil liquefaction prediction is to analyze complex geotechnical data, including soil properties and seismic parameters, to accurately assess liquefaction susceptibility. Key features of DL include its ability to capture non-linear relationships within data and its capacity to learn from diverse datasets. DL is commonly used in geotechnical engineering for tasks such as soil classification, landslide prediction, and liquefaction assessment.

DL techniques are relevant to our project as they offer a promising approach to addressing the challenges associated with soil liquefaction prediction. By leveraging large datasets and complex data patterns, DL models can enhance prediction accuracy and improve risk management strategies. Incorporating DL into our research can contribute to the success of our project by providing more reliable liquefaction susceptibility assessments and informing engineering decisions in earthquake-prone regions.

5. Architecture Design Diagram

a comprehensive overview of the system is seen in Figure 2, illustrating the flow from data collection to model deployment and risk assessment, while highlighting the roles and functionalities of each component within the project.

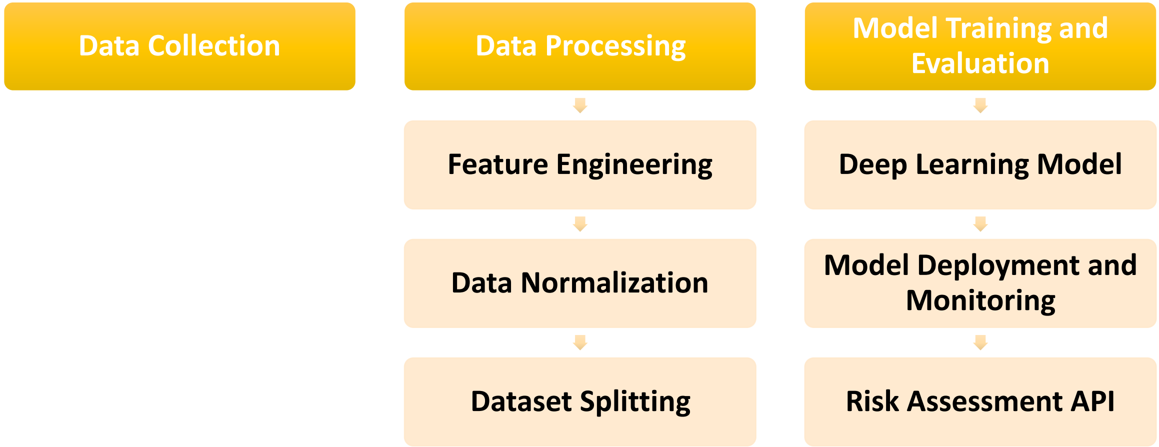

Figure 2 :Architecture Design Diagram

1. **Data Collection**
    - **Role**: To gather and aggregate all necessary data for the project.
    - **Functionality**: Collects data from various sources including seismic event records, soil properties, and historical case studies of liquefaction.
2. **Data Processing**
    - **Role**: To prepare the raw data for analysis and model training.
    - **Functionality**: Includes steps such as cleaning, formatting, and transforming the data into a usable format.
3. **Model Training and Evaluation**
    - **Role**: To develop and refine the deep learning model.
    - **Functionality**: Involves training the model on the processed data, evaluating its performance, and making necessary adjustments.
4. **Feature Engineering**
    - **Role**: To create new features from the raw data that will improve model performance.
    - **Functionality**: Involves selecting and transforming variables to capture important patterns.
5. **Data Normalization**
    - **Role**: To standardize and enhance the dataset.
    - **Functionality**: Normalizes data to ensure consistency.
6. **Dataset Splitting**
    - **Role**: To divide the data into training, validation, and test sets.

- **Functionality**: Ensures the model is trained and evaluated on different subsets of data to prevent overfitting and ensure generalizability.

7. **Deep Learning Model**
    - **Role**: To analyze the data and predict soil liquefaction risk.
    - **Functionality**: Uses neural networks to learn patterns and make predictions based on the input data.
8. **Model Deployment and Monitoring**
    - **Role**: To implement the trained model in a real-world setting.
    - **Functionality**: Involves deploying the model as an API and continuously monitoring its performance.
9. **Risk Assessment API**
    - **Role**: To provide a user interface for accessing the model's predictions.
    - **Functionality**: Allows users to input new data and receive an assessment of soil liquefaction risk.

6. Data Sources

Data sources include geotechnical reports, seismic surveys, and academic research papers. Data formats are primarily tabular datasets and research documents with liquefied and non-liquefied data cases, with data sizes varying based on the specific seismic events and regions analyzed. The selection of this data is due to its relevance in assessing liquefaction potential, providing a comprehensive understanding of the contributing factors: Those data sets- which underwent preprocessing before merging based on calculated CRR7.5 values include:

- Toprak (1999)[4] : CPT data set, SPT data set
- Moss(2006)[5] : CPT data set
- Hanna (2007)[6]: CPT and Vs data set
- Hanna (2007)[7]:SPT and Vs data set
- Kayen (2013) [8]: Vs data set
- Boulanger(2014)[9] : CPT data set, SPT data set
- Zhao (2022) [10]: CPT data set

7. Literature Review

Although the implementation of ML techniques in the evaluation of liquefaction has started since the beginning of the 1990’s the prediction models using DL techniques are not common; since DL models need large data sets to function properly and getting huge data sets can be a problem referring to the required data type. despite this problem, some researchers have successfully implemented DL like the study by : (Table 1)

1. Zhang et al. (2021):[11]

- Proposed a deep neural network (DNN)-based model for predicting soil liquefaction.
- Incorporated shear wave velocity (Vs) and standard penetration test (SPT) data to enhance prediction accuracy.
- Demonstrated high accuracy in liquefaction prediction, especially with the inclusion of Vs data.
- Recommended further studies to improve model performance by incorporating additional parameters and increasing dataset size.

2. **Kumar et al. (2021):**[12]
    - Introduced a novel DL methodology for classifying soil liquefaction.
    - Compared DL models with emotional backpropagation neural network (EmBP).
    - Utilized cone penetration test data, with cone resistance (qc) and peak ground acceleration (PGA) as inputs.
    - Achieved excellent performance compared to EmBP, validated using global earthquake data.
    - Suggested further exploration of DL and EmBP techniques in various civil engineering applications.
3. **Kumar (2023):**[13]
    - Utilized advanced ML models (DNN, CNN, RNN, LSTM, BILSTM) to assess liquefaction triggering.
    - Trained and tested models using 834 case histories of post-liquefaction data.
    - Used two independent variables, N1,60 and CSR, to predict the probability of liquefaction.
    - RNN model outperformed others, achieving high coefficient of determination (R2 = 0.906) during testing.
    - Recommended further steps to enhance model accuracy, including training on larger datasets, comparison with standard models, and validation through laboratory experiments.

Each study contributes to the advancement of soil liquefaction prediction models, addressing challenges such as data availability and model accuracy. Additionally, they highlight the importance of incorporating various parameters and employing advanced ML and DL techniques for more reliable predictions in civil engineering applications.
In this project, a DL approach will be used to determine liquefaction susceptibility using SPT , CPT, and Vs data sets while incorporating more parameters.

Table 1 comparison of the existing DL models

Study by	Model	Test Type	Number of Inputs	Inputs	Number of Layers	Activation function	Loss function	optimizer	Performance Metric	
Zhang et al. (2021)[11]	DL: ML-FCN	SPT & Vs	8	(*z*), (*FC*), (*N*1,60), σv), ($\sigma' v$), (*amax*), Vs and (*Mw*)	9 Dropout (4)	ReLu	BCE	adam	Accuracy 89% 92% 93%	
Kumar et al. (2021)[12]	DL	CPT	2	qc, PGA	-	sigmoid	BCE	adam	Precision–Recall F measure 0.73, 0.65	
D.R.Kumar et al (2023)[13]	DNN	SPT	7	(*z*), (*FC*), (*N*1,60), σv), ($\sigma' v$), (*amax*), and (*Mw*)	2 hidden layers	ReLu	MAE	adam	Total rank	52
	CNN									55
	RNN									70
	LSTM									16
	BILSTM									47

- Cone Resistance(qc),
- Peak Ground Acceleration (PGA)
- depth (*z*),
- fine content (*FC*),
- corrected SPT count (*N*1,60),
- total stress (σv),
- effective stress ($\sigma' v$),
- peak horizontal acceleration (*amax*),

- Shear wave velocity (Vs)
- magnitude of the earthquake (*Mw*)
- BCE: Binary Crossentropy
- MAE**:** Mean Absolute Error

Implementation Plan

1. **Technology Stack**

Programming Languages:

- **Python**: The primary language for developing machine learning models, data processing, and overall implementation due to its extensive libraries and community support.

Libraries and Frameworks:

- **TensorFlow / Keras**: For building and training deep learning models. TensorFlow provides a flexible ecosystem for model training, while Keras offers a user-friendly API for building neural networks.
- **scikit-learn**: For traditional machine learning algorithms, data preprocessing, and model evaluation.
- **Pandas**: For data manipulation and analysis. It's essential for handling large datasets and performing complex operations.
- **NumPy**: For numerical computing and array operations. It's foundational for many data science tasks.
- **Matplotlib / Seaborn**: For data visualization. These libraries are used to create plots, graphs, and visualizations to understand data distributions and model performance.

APIs and Web Frameworks:

- **Flask / FastAPI**: For developing RESTful APIs to deploy the machine learning models and provide access to risk assessment services.
- **Docker**: For containerizing applications to ensure consistency across different environments and facilitate deployment.

Development and Collaboration Tools:

- **Jupyter Notebook**: For interactive development, data exploration, and sharing insights.
- **Git**: For version control, managing code changes, and collaboration.
- **GitHub / GitLab**: Platforms for hosting repositories and collaborating with team members.

2. Timeline

Table 2 work-time table

	Stage	W1	W2	W3	W4	W5	W6	W7	W8
1	Data Collection and Preprocessing								
2	Feature Engineering and Selection								
3	Deep Learning Model Development								
4	Training and Validation of the Model								
5	Evaluation and Fine-tuning								
6	Model Deployment								

3. Milestones

1. Milestone 1: Completion of Data Collection and Preprocessing (End of Week 2)
- All relevant data (seismic event records, soil properties, historical case studies) is collected, cleaned, and formatted.
- Data is free of missing values and outliers.

2. Milestone 2: Completion of Feature Engineering and Selection (End of Week 3)
- All relevant features are identified and engineered.
- The final dataset is prepared and ready for model training.

3. Milestone 3: Development of Initial Deep Learning Model (End of Week 4)
- The deep learning model architecture is defined and implemented.
- Initial scripts for training the model are ready.

4. Milestone 4: Completion of Model Training and Validation (End of Week 6)
- The model is trained on the dataset and validated for performance.
- Cross-validation is performed to ensure model robustness.
- The best performing model is saved.

5. Milestone 5: Model Evaluation and Fine-tuning (End of Week 6)
- The model's performance is evaluated on the test set.
- Necessary fine-tuning and adjustments are made for optimal performance.

6. Milestone 6: Successful Model Deployment (End of Week 8)
- The model is deployed using an API and containerized with Docker.
- The model is tested in a real-world environment.
- Final documentation and user guides are prepared.

4. Challenges and Mitigations

Data Quality

- Challenge: Incomplete or noisy data can lead to inaccurate predictions.
- Mitigation:
    - Performing thorough data cleaning and preprocessing.
    - Using imputation techniques to handle missing values.
    - Collecting additional data if necessary to fill gaps.

Model Performance

- Challenge: The model may not perform well on unseen data or may overfit to the training data.
- Mitigation:
    - Using techniques like cross-validation to ensure model robustness.
    - Regularize the model using techniques such as dropout or L2 regularization.
    - Continuously monitor and adjust hyperparameters.

Technical Constraints

- Challenge: Limited computational resources may slow down model training and deployment.
- Mitigation:
    - Optimizing the code and using efficient algorithms to reduce computational load.
    - Using GPU acceleration for model training.

5. Ethical Considerations

Data Privacy

Consideration: Ensuring the responsible use of data collected from literature, even if it doesn't include personal information.

Approach:

- Respecting Copyrights: Ensuring that all data collected from literature sources are used in accordance with copyright laws and proper attribution is given to original authors.
- Transparency: the source of data will be clearly documented
- Compliance: any specific usage guidelines or restrictions associated with the literature sources from which data is collected will be adhered to.

Impact on Target Community

Consideration: The outcomes of the model could significantly impact communities in seismic zones, influencing safety measures and urban planning.

Approach:

- Experts in geotechnical engineering will be engaged to ensure the model's outputs are reliable and actionable.
- A clear explanation and documentation of the model's predictions and limitations will be provided.
- It will be ensured that the model is used as a decision-support tool rather than a sole determinant for safety measures.

6. References

[1] O. Cetin, "Preliminary Reconnaissance Report on February 6, 2023, Pazarcık Mw=7.7 and Elbistan Mw=7.6, Kahramanmaraş-Türkiye Earthquakes," 2023.

[2] Abbas Daftari, "New Approach in Prediction of Soil Liquefaction," 2015.

[3] A. Baghbani, T. Choudhury, S. Costa, and J. Reiner, "Application of artificial intelligence in geotechnical engineering: A state-of-the-art review," *Earth-Science Rev.*, vol. 228, no. February, p. 103991, 2022.

[4] S. Toprak, T. L. Holzer, M. J. Bennett, and J. J. Tinsley, "CPT-and SPT-based probabilistic assessment of liquefaction potential," *7th US–Japan Work. Earthq. Resist. Des. Lifeline Facil. Countermeas. against Liq. Seattle, WA.*, no. August 1999, pp. 69–86, 1999.

[5] R. E. S. Moss, "CPT-based probabilistic assessment of seismic soil liquefaction initiation," no. April, p. 528, 2003.

[6] A. M. Hanna, D. Ural, and G. Saygili, "Evaluation of liquefaction potential of soil deposits using artificial neural networks," *Eng. Comput. (Swansea, Wales)*, vol. 24, no. 1, pp. 5–16, 2007.

[7] A. M. Hanna, D. Ural, and G. Saygili, "Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data," *Soil Dyn. Earthq. Eng.*, vol. 27, no. 6, pp. 521–540, 2007.

[8] R. Kayen *et al.*, "Shear-Wave Velocity–Based Probabilistic and Deterministic Assessment of Seismic Soil Liquefaction Potential," *J. Geotech. Geoenvironmental Eng.*, vol. 139, no. 3, pp. 407–419, 2013.

[9] R. W. Boulanger and I. M. Idriss, "CPT and SPT based liquefaction triggering procedures, Report UCD/CGM-10/2," *Cent. Geotech. Model.*, no. April, pp. 1–138, 2014.

[10] Z. Zhao, W. Duan, G. Cai, M. Wu, and S. Liu, "CPT-based fully probabilistic seismic liquefaction potential assessment to reduce uncertainty: Integrating XGBoost algorithm with Bayesian theorem," *Comput. Geotech.*, vol. 149, no. February, p. 104868, 2022.

[11] Y. Zhang, Y. Xie, Y. Zhang, J. Qiu, and S. Wu, "The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity," *Bull. Eng. Geol. Environ.*, vol. 80, no. 6, pp. 5053–5060, 2021.

[12] D. Kumar, P. Samui, D. Kim, and A. Singh, “A Novel Methodology to Classify Soil Liquefaction Using Deep Learning,” *Geotech. Geol. Eng.*, vol. 39, no. 2, pp. 1049–1058, 2021.

[13] D. R. Kumar, P. Samui, A. Burman, W. Wipulanusat, and S. Keawsawasvong, “Liquefaction susceptibility using machine learning based on SPT data,” *Intell. Syst. with Appl.*, vol. 20, no. June, p. 200281, 2023.