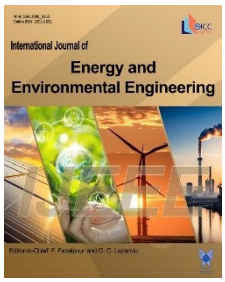

Volume 15, Issue 3, (2024) Article 13

International Journal
of
Energy and Environmental Engineering (IJEEE)

https://doi.org/10.57647/ijeee.2024.1503.13

Metaheuristic-Driven Deep Learning Framework for Soil Liquefaction Assessment

Shima Aghakasiri [1], Ghodratollah Mohammadi [1,*], Amir Taban [2], Mohammad Emami Kourandeh [1]

[1] *Department of Civil Engineering, ST.C., Islamic Azad University, Tehran, Iran*
[2] *Department of Civil Engineering, KHA.C., Islamic Azad University, Khorramabad, Iran*

***Corresponding author:** Gh_mohammadi@azad.ac.ir

Original Research

Abstract

Soil liquefaction in saturated sandy deposits remains a major concern in seismically active regions. This study presents a metaheuristic-driven deep learning framework for liquefaction assessment using a dataset of 512 boreholes collected across Mazandaran Province, Northern Iran. The dataset consists of 316 non-liquefied and 196 liquefied cases derived from field observations and verified geotechnical parameters. The proposed framework integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures, whose hyperparameters are optimized using the Grey Wolf Optimizer (GWO) and the Whale Optimization Algorithm (WOA). Model performance was evaluated using stratified repeated train–test experiments (10 iterations) to ensure reproducibility. The ensemble model achieved an average accuracy of 94.5%, a correlation coefficient of 0.93, and AUC values above 0.95, outperforming all individual models and conventional liquefaction evaluation methods. SHAP-based analysis identified SPT blow count and peak ground acceleration as the most influential variables. The proposed framework demonstrates strong robustness and practical applicability for regional seismic risk assessment and can be extended to other geotechnical prediction tasks.

Keywords: Liquefaction, Deep Learning, Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), Grey Wolf Optimizer (GWO), Whale Optimization Algorithm (WOA), Sensitivity Analysis

Cite this article: Aghakasiri Sh., Mohammadi Gh., Taban A., Emami Kourandeh M., Metaheuristic-Driven Deep Learning Framework for Soil Liquefaction Assessment. Int. J. Energy Environ. Eng. 2024; 15(3) : 1-29 https://doi.org/10.57647/ijeee.2024.1503.13

Highlights

- A metaheuristic-driven deep learning framework is proposed for liquefaction assessment.
- CNN and LSTM architectures are optimized using GWO and WOA algorithms.
- The model uses 512 boreholes with balanced stratified evaluation.
- Ensemble DL model achieves 94.5% accuracy and AUC>0.95 across 10 runs.
- The framework outperforms Seed–Idriss and Robertson–Wride by 18–26% in accuracy

List of abbreviations:

LSTM	Long Short-Term Memory
GWO	The Grey Wolf Optimizer
WOA	The Whale Optimization Algorithm
CNN	Convolutional Neural Networks
SPT	Standard Penetration Test
CSR	Cyclic Stress Ratio
CRR	Cyclic Resistance Ratio
σ	total vertical stress
σv	the effective vertical stress
r_d	the stress reduction factor
	represents the maximum horizontal acceleration due to the earthquake at the ground surface
g	the acceleration due to gravity
Z	Depth (m)

1. Introduction

Soil liquefaction is a primary cause of earthquake-induced damages. After the extensive liquefaction events during the devastating 1964 earthquakes in Niigata, Japan, and Alaska, USA, geotechnical engineers began paying significant attention to this phenomenon [1].

Liquefaction hazard mapping in Iran commenced following the destructive 1990 Manjil earthquake, spearheaded by the International Institute of Earthquake Engineering and Seismology. The outcomes were presented as macro-zonation maps at a scale of 1:1,000,000 [1]. Over the past 40 years, significant advancements have been made in understanding the mechanisms of liquefaction and its influencing factors. In the early stages, much attention was focused on clean sands, with the belief that liquefaction was exclusive to sandy soils. It was presumed that fine-grained and coarse-grained soils could not generate the excess pore water pressures that are the primary cause of liquefaction. However, with time and observations of liquefaction in new earthquakes involving fine- and coarse grained soils, researchers sought to identify the factors affecting liquefaction in these soil types as well by Uyanık (2020)[2]. Thus, the primary focus of this research is to map the coastal areas of Mazandaran Province based on liquefaction hazard and predict it using artificial neural network (ANN) methods. Additionally, improvement methods to counter liquefaction, based on the identified liquefaction potential, will be proposed. This study aims to refine behavioral parameters, as opposed to conventional behavioral models that consider these parameters in a general manner. By integrating artificial neural network models with standard behavioral models in geotechnical engineering, it seeks to enhance the accuracy of predicted behaviors through neural network training.

Maurer *et al*. in 2014 published a study aimed at determining and comparing the liquefaction potential of the recent sediments in the Tepebaşı, Eskişehir region using CPT and SPT data. They calculated liquefaction potential index values using 42 different CPT datasets and 53 different SPT datasets. Considering the seismic characteristics of the regional fault, a peak horizontal ground acceleration of 0.3g was assumed, and the magnitude of the 1956 earthquake (6.4) was incorporated into the calculations[3]. The SPT data were utilized in the Seed and Idriss 1971 method[4], while the CPT data were used in the Robertson and Wride 1998 method[5]. The analysis results were employed to prepare liquefaction hazard maps for the region. Additionally, a comparison between the two methods revealed that the liquefaction hazard values obtained from both methods were approximately consistent Seed and Idriss 1971[4].

A Newmark analysis of lateral spreading induced by liquefaction was conducted using a database of 22 documented cases by Yang and Kavazanjian in 2022[6]. The site stratigraphy and SPT values for the liquefiable soil were based on the historical records of the studied cases. Yield acceleration was estimated based on the residual shear strength of liquefiable soils using three different correlations with SPT blow counts. Representative acceleration histories were selected from recorded earthquake data. Probabilistic analysis using a truncated normal distribution indicated that the computed lateral spreading displacement, derived using Olson and Johnson's correlation, would not exceed twice the reported displacement with 97% confidence.

Lateral spreading induced by liquefaction can cause significant deformation and damage existing structures such as ports, bridges, and pipelines. Past earthquakes have triggered such phenomena in coastal and riverine areas across various parts of the world. Current predictive models for lateral spreading tend to underestimate or overestimate actual displacements by a factor of two or more when applied to large subduction earthquake events. The objective of this study is to identify ground motion intensity measures and soil parameters that exhibit stronger correlations with observed lateral spreading during large subduction earthquakes (Mw ≥ 7.5) occurring in countries such as Chile, Japan, and Peru. A numerical approach was initially validated against centrifuge and historical case studies and subsequently discussed to develop parametric models from which statistical analyses were performed. The results indicate that cumulative absolute velocity (CAV), Husid intensity (HI), and sustained maximum velocity (SMV) demonstrate strong correlations with lateral spreading for the analyzed cases[7].

In recent years, machine learning and deep learning techniques have emerged as powerful tools for predicting complex nonlinear behaviors in geotechnical problems. Among these, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and hybrid models such as CNN, LSTM combined with metaheuristic optimizers have shown promising results.

In this study, SPT and CPT parameters were used as the primary geotechnical indicators. The deep learning components include CNN and LSTM networks, while optimization is performed using GWO and WOA. All abbreviations are defined at first use to ensure clarity.

Northern Iran, particularly Mazandaran Province, is recognized as a seismically active region due to proximity to the Caspian fault system and Alborz tectonic belt. Recent regional studies have documented multiple liquefaction events and soil strength loss during moderate-to-strong earthquakes, highlighting the importance of accurate liquefaction assessment in this area.

Soil liquefaction remains one of the most critical geotechnical hazards associated with strong ground shaking, often leading to catastrophic structural damage, lateral spreading, and ground settlement. Traditional liquefaction evaluation procedures typically based on deterministic correlations using SPT, CPT, or shear-wave velocity provide fundamental insights but are often constrained by simplified assumptions, empirical boundaries, and limited ability to capture the complex, nonlinear interaction among soil parameters. In recent years, rapid advancements in computational intelligence have enabled machine learning (ML) and deep learning (DL) methods to emerge as powerful alternatives for modeling liquefaction susceptibility, owing to their capability to learn complex, multi-dimensional patterns directly from data.

Several studies have demonstrated the effectiveness of ML-based models in predicting liquefaction from in-situ test data. Kumar et al. (2023) showed that ML classifiers trained on SPT-derived parameters can significantly improve liquefaction susceptibility assessment compared to classical empirical charts [8, 9].

Deep learning has further expanded the analytical capabilities in geotechnical earthquake engineering. Zhang et al. (2021) adopted deep neural networks (DNNs) to predict liquefaction using shear-wave velocity profiles, demonstrating that deep architectures can effectively capture nonlinear subsurface behavior[10]. These findings echo the broader trend highlighted by Jas and Dodagoudar (2023), who reviewed ML-based liquefaction research from 1994 to 2021 and emphasized that data-driven methods consistently outperform traditional threshold-based criteria[11].

Other studies have explored Bayesian and probabilistic frameworks to better quantify uncertainties in liquefaction evaluation. Ahmad et al. (2021) used Bayesian belief networks to model liquefaction-induced lateral displacements, highlighting how probabilistic ML models provide transparent reasoning and uncertainty quantification in complex geotechnical systems[12]. Fragility-based approaches have also evolved, with Geyin and Maurer (2020) offering fragility functions for liquefaction-induced ground failure and advancing modern consequence-based assessments[13]. More recent data-driven analytic frameworks emphasize integrating multiple emerging technologies to improve hazard prediction and consequence evaluation [14].

Hybrid and metaheuristic-optimized machine learning models have also gained traction. Zhou et al. (2022) evaluated the performance of GA–SVM and GWO–SVM hybrid models across multiple datasets and concluded that metaheuristic-driven optimization significantly enhances model generalization and predictive performance[15]. Similar findings were reported by Zhang et al. (2021), who applied a GWO-optimized SVM model to predict liquefaction potential and achieved higher accuracy than conventional SVM or empirical liquefaction charts[16].

Finally, advances in large-scale ML-based hazard mapping highlight the growing computational capabilities in this field. Bi et al. (2019) developed a fast multi-layer ML framework for liquefaction disaster assessment on regional scales, demonstrating the feasibility of applying advanced ML techniques for large geographic areas subject to seismic risk[17].

Collectively, these studies underline a clear research trajectory: ML, DL, and hybrid metaheuristic–optimized models provide more flexible, accurate, and data-driven pathways for liquefaction evaluation than traditional empirical methods. Nevertheless, challenges remain, including the integration of heterogeneous geotechnical parameters, uncertainty quantification, and the need for models tailored to regional datasets gaps that recent literature encourages addressing. The present study contributes to this growing body of research by developing a metaheuristic-driven deep learning framework that integrates CNN, LSTM, and Transformer architectures, optimized using GWO and WOA algorithms for robust soil liquefaction assessment.

The primary innovation of this study lies in the development and application of a CNN, LSTM-GWO, WOA hybrid model for predicting soil liquefaction

potential using SPT data. Unlike previous studies that have used CNN, LSTM or optimization algorithms separately, this research integrates the strengths of both in a unified framework. Key contributions include:

The implementation of a CNN, LSTM-GWO, WOA hybrid model for the first time in the context of liquefaction assessment based on geotechnical field data.

Use of a comprehensive, region-specific database from over 500 boreholes in northern Iran, enhancing model reliability.

Validation of the model through sensitivity analysis, MSE, and regression coefficient (R), demonstrating superior predictive performance.

Identification of the most influential factors in liquefaction risk, with fine particle content emerging as the dominant parameter.

These studies collectively highlight the need for more robust, generalizable models that integrate metaheuristic optimization with modern deep learning frameworks.

The novelty of this study lies in (1) integrating CNN and LSTM into a unified ensemble structure, (2) optimizing the architecture using two complementary metaheuristic algorithms GWO and WOA, and (3) constructing the first region-specific, DL-optimized liquefaction prediction model for Northern Iran using 512 boreholes.

The present study begins with an introduction and a review of the technical literature related to the research. It proceeds to discuss the theoretical foundations of the research, the concept of liquefaction, its potential evaluation, and innovative methods in this area. The third section presents the research methodology and the process of developing the database. The analysis of the results, modeling through deep learning methods, and assessing the efficiency of these methods are discussed in the fourth section. Finally, the study concludes with a summary and conclusions.

2. Innovation and contributions

This study offers several novel contributions to the field of liquefaction assessment and geotechnical risk analysis:

Development of a Hybrid Metaheuristic–Deep Learning Ensemble:
The research introduces an integrated framework that combines (CNN) and (LSTM) with two metaheuristic optimizers, (GWO) and (WOA). Unlike earlier studies that applied these approaches independently, the proposed ensemble model provides enhanced predictive accuracy (94.5%) and improved robustness, demonstrating its superiority over conventional hybrid structures.

Comprehensive Benchmarking with Classical Geotechnical Methods:
In addition to modern data-driven approaches, the study compares the performance of the proposed models with established methods such as the Seed & Idriss (1971)[4] simplified procedure and the Robertson & Wride (1998)[5] CPT-based approach. This benchmarking highlights the clear advantages of advanced deep learning models in capturing complex, nonlinear relationships within geotechnical datasets.

Unique, Region-Specific Database:
A dataset of more than 500 boreholes was compiled from Mazandaran Province, northern Iran, through field investigations, laboratory testing, and numerical analyses. This dataset, which reflects the seismic and geotechnical conditions of one of Iran's most hazard-prone regions, constitutes a valuable and reliable resource that has not been previously available for machine learning–based liquefaction studies.

Sensitivity-Based Knowledge Discovery:
Beyond model performance, the study emphasizes interpretability through a detailed sensitivity analysis. Results demonstrate that the Standard Penetration Test blow count (SPT-N) and Peak Ground Acceleration (PGA) are the most influential parameters, providing practical insights for engineers and policymakers in seismic hazard evaluation and soil improvement planning.

Transferable Framework for Geotechnical Engineering:
The proposed hybrid methodology is not limited to liquefaction but can be extended to other geotechnical applications such as slope stability, settlement prediction, and bearing capacity analysis. This transferability underscores the broader impact of the study and its potential to guide future applications of artificial intelligence in geotechnical risk management.

3. Theoretical foundations of the study

Liquefaction is one of the most significant topics in seismic geotechnical engineering. In 1964, two earthquakes occurred in Alaska and Niigata, Japan, both of which showcased prominent examples of earthquake-induced damage, such as slope failures, foundation and bridge collapses, and the buoyancy of buried structures due to the liquefaction of underlying soils. Similarly, the 1990 Manjil earthquake in Iran left behind considerable damage attributed to liquefaction. Following these events, the phenomenon of liquefaction was extensively

studied by researchers [1]. The 1964 Niigata earthquake in Japan is considered the first documented occurrence that led to the destruction of modern infrastructure and was later recognized as the phenomenon of liquefaction by Matsuoka *et al.* in 2015[18]. During an earthquake, seismic waves propagate outward from the source and traverse the Earth's crust at high speeds. The nature and distribution of earthquake-induced damage heavily depend on the soil's response to dynamic seismic loads, which is largely governed by the dynamic properties of the soil, including stiffness, damping, Poisson's ratio, and soil density by kumar *et al.* in 2023[9]. One of the major factors controlling strong ground motion and structural damage during earthquakes is the shear wave velocity of the surface sediments down to the bedrock [19]. Shear wave velocity reflects the type of material and structural conditions of the soil. It can be utilized to evaluate the stratigraphy, the degree of compaction or consolidation of soft soils, and to identify weak zones at the site[20].

Bhatnagar *et al.* in 2016 conducted seismic analysis and earthquake effects on liquefaction at various upstream water levels on the dam[21]. Sonmezer in 2019 investigated the liquefaction potential of fiber-reinforced sand and concluded that the inclusion of fibers increases the soil resistance to liquefaction. conducted an energy-based assessment of the liquefaction potential of uniform sands. The results showed that the energy dissipated per unit volume is very useful and may cover non-uniform loading conditions across the earthquake spectrum by Sonmezer[22, 23]. Sonmezer *et al.* 2022 investigated the effect of grain size ratio and silt content on the liquefaction potential of silty sands[24].

Recent advances in artificial intelligence have enabled the use of machine learning algorithms to assess liquefaction risk. By incorporating neural networks and hybrid architectures, researchers have improved prediction accuracy and identified dominant influencing factors. These methods move beyond the limitations of traditional empirical models and offer adaptable solutions for diverse soil and seismic conditions.

Accordingly, this study proposes a CNN, LSTM+GWO, WOA hybrid model tailored for liquefaction assessment in the Mazandaran coastal region, emphasizing its potential for enhanced predictive capability and geotechnical relevance.

3.1. Comparison between empirical and hybrid deep learning approaches (CNN–LSTM–GWO/WOA)

Traditional empirical relationships for liquefaction assessment, such as those proposed by Seed and Idriss, were derived from a limited number of field case histories and laboratory studies. These models typically express soil behavior as simplified linear or semi-empirical correlations between a few key parameters such as the corrected SPT value ($N_{1(60)}$), peak ground acceleration (a_{max}) and effective vertical stress (σ'_v). Although widely used, such relationships have limited capability to capture the nonlinear and site-specific behavior of soils under diverse seismic and geological conditions.

In contrast, the present study employs advanced machine learning models, including a CNN and a LSTM network, integrated with two metaheuristic optimization algorithms GWO and WOA. These hybrid models are designed to automatically identify complex nonlinear interactions among geotechnical variables without assuming any predefined mathematical form.

4. Research methodology

In this study, the research methodology is designed to address the primary and secondary research questions. Additionally, efforts are focused on identifying the optimal model applicable to all sandy soils susceptible to liquefaction with similar geotechnical data. The process and methodology of the current research consist of several stages, some of which run in parallel. These stages are as follows:

1. Comprehensive identification of liquefaction hazards and methods for calculating their potential.
2. Analysis and evaluation of liquefaction potential in projects located in the northern provinces of the country.
3. Collection of necessary data for training artificial neural networks and deep learning methods.
4. Analysis of the compiled database and removal of outliers.
5. Determination of input and output parameters for the deep learning method.
6. Coding the deep learning method by combining the CNN, LSTM approach with the Optimizer GWO, WOA algorithm.
7. Implementation of the deep learning method on the compiled database.
8. Performance evaluation of the deep learning method using assessment indices.

Boreholes were collected across Mazandaran Province between longitudes 50.8°–54.0°E and latitudes 35.8°–36.7°N. The dataset covers coastal plains, alluvial fans, and piedmont regions. A distribution map (Figure 1) illustrates borehole locations. The region is affected by the Central Alborz fault system and exhibits PGA values

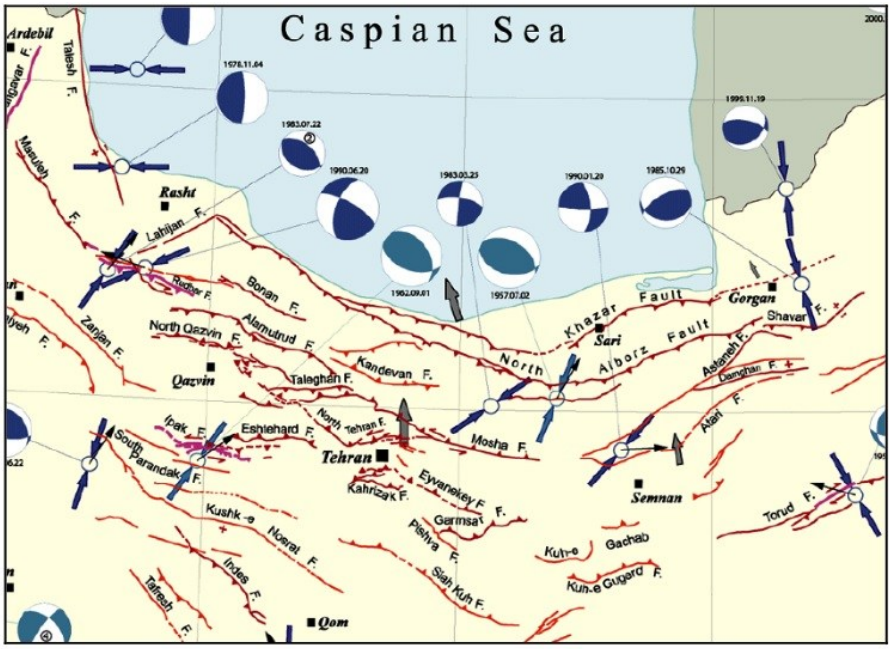

Figure 1. illustrates borehole locations

ranging 0.18–0.32 g based on national seismic hazard maps.

4.1. *Liquefaction hazard calculation*

Soil liquefaction occurs when cyclic shear stresses generated by earthquake loading exceed the soil's cyclic resistance, resulting in a rapid loss of effective stress and stiffness. This process is governed by factors such as density, confining stress, fines content, and cyclic stress ratio (CSR). For brevity and clarity, the mechanism is summarized here based on well-established frameworks. Liquefaction refers to the reaction of soil under dynamic loads or transient shear wave excitation, resulting in the complete loss of soil strength and transformation into a liquid-like state. If saturated loose sand is subjected to seismic shaking, it contracts and reduces in volume. When water cannot drain quickly enough, this volume reduction increases the pore water pressure, leading to a decrease in effective stress and shear strength, ultimately causing the soil to behave like a liquid. Several geological factors affect soil liquefaction susceptibility, including depositional process, sediment age, and geological history, groundwater table depth, grain size distribution, burial depth, slope inclination, and proximity to a free face. Evidently, as the sediment age increases, its susceptibility to liquefaction tends to decrease. This behavior is primarily attributed to the overconsolidation effect induced either by the weight of overlying sediments or by previous seismic events. The highest liquefaction potential is commonly observed in coastal regions characterized by fine-grained, saturated sandy alluvium that exhibit low density and minimal cohesion. Seed and Idriss in 1971 proposed the following equation for calculating the cyclic shear stress ratio[4]:

$$CSR = \left({}^{\tau_{av}}/_{\sigma_0'}\right) = 0.65({}^{a_{max}}/_{g})({}^{\sigma_v}/_{\sigma_v'})({}^{r_d}/_{MSF}) \quad (1)$$

Where:

- a_{max}: a maximum magnitude earthquake (Peak horizontal ground acceleration)
- σ_v: Total vertical stress
- σ_v': Effective vertical stress
- r_d: Stress reduction coefficient due to soil flexibility
- MSF: Magnitude scaling factor

At the surface, the value of r_d is equal to one, and it decreases with increasing depth. The variations of the r_d coefficient with respect to depth are shown in Figure (2). where a_{max} is the peak horizontal ground acceleration, σ_v and σ_v' are the total and effective vertical stresses, respectively, r_d is the depth-dependent stress reduction factor, and MSF is the magnitude scaling factor. For the present study, a moment magnitude of M_w= 7.00 was considered based on seismic hazard analyses of the study region. The relationships between depth and r_d are illustrated in Figure 2. The Cyclic Resistance Ratio (CRR), representing the cyclic shear strength of soils, is typically evaluated using in-situ test parameters. Bolton Seed et al. (1985) provided correlations for CRR as a function of corrected Standard Penetration Test values, where clean sand equivalence is applied to account for fines content[26]:

$$CRR=f(N_{1,60,cs}, FC) \ for \ (N_{1,60,cs} < 30) \quad (2)$$

where $N_{1,60,cs}$ is the corrected SPT blow count normalized to clean sand, and FC is the fines content. For soils with $(N_{1,60,cs} < 30)$, the material is considered sufficiently dense and not susceptible to liquefaction.
The factor of safety (FS) against liquefaction is then calculated as:

$$FS = \frac{CRR}{CSR} \quad (3)$$

Table 1 presents the calculated liquefaction potential index values for the investigated boreholes, while figure 3 shows the factors of safety against liquefaction for borehole BH.

The liquefaction susceptibility index was calculated for borehole BH to quantify the overall potential for liquefaction across the soil profile. This index offers a site specific measure reflecting the cumulative effect of depth dependent variations in soil resistance and cyclic stress ratio, serving as a critical input for seismic risk analysis and geotechnical design.

For fine-grained soils, alternative criteria such as those proposed by Idriss & Boulanger (2005) and Bray & Sancio (2006) are employed[27, 28]. These criteria evaluate plasticity characteristics and the ratio of the soil's natural water content to its liquid limit in order to determine the liquefaction susceptibility index.

The closer this index is to 1.0, the more the fine-grained soil behaves like sand; conversely, values approaching 0.0 indicate a clay-like behavior.

The cyclic shear strength, or the shear stress required to increase pore water pressure and nullify effective stress, can be determined through field tests such as the Standard Penetration Test (SPT). Several relationships exist for calculating the Cyclic Resistance Ratio (CRR) using corrected SPT values, fine-grain content, and sometimes soil plasticity properties. Seed et al. in 1983 provided the chart shown in Figure 4 for calculating the CRR (τ_{av} / σ'_0) based on SPT values[25].

The factor of safety against liquefaction is defined as the ratio of the cyclic shear resistance to the cyclic shear stress. If the factor of safety is less than one, liquefaction will occur. The liquefaction analysis was performed using the mentioned method with the LiqufyPro software (version 4.5D by CivilTech) on boreholes where the soil conditions were prone to liquefaction. The obtained results were analyzed and reviewed. The analyses were conducted under the assumption of earthquakes with acceleration and magnitude values of a_{max}=0.3g, M=7.00 based on the seismicity conditions of Northern provinces.

Figure 5 illustrates a sample of the calculations and simulations performed using the aforementioned software.

The mentioned calculations and modeling in this section were carried out for more than 500 boreholes drilled in geotechnical study projects located in the northern provinces of the country to enrich the database of the present study. All 512 samples used in this study are derived from real borehole logs and field liquefaction observations. No synthetic or augmented data were used. LiquefyPro was used strictly to compute auxiliary parameters, including cyclic stress ratio (CSR), stress reduction factor (rd), and factor of safety (FS). These parameters were not used to generate or infer liquefaction labels. Ground-truth liquefaction occurrence (0 or 1) was assigned solely based on documented field observations associated with each borehole. Therefore, no circular validation is present.
Liquefaction labels were assigned based on documented field performance:

Liquefied (1): cases where observed ground failure, settlement, sand boils, or loss of bearing capacity were recorded, with factor of safety FS < 1.0 confirming susceptibility.

Non-liquefied (0): cases where no liquefaction manifestation was reported and FS ≥ 1.0.

4.2. *Statement removing synthetic data*

All references to synthetic data have been removed. The analysis uses only real field samples.

4.3. *Dataset statistical summary table*

Statistical characteristics of the dataset are summarized in Table 2. For each input variable, the minimum, maximum, mean, and standard deviation values are reported to provide a quantitative description of data distribution and variability. This statistical overview enables transparent assessment of data quality, range, and potential scaling requirements prior to model training. The results indicate that the dataset covers a wide range of geotechnical and seismic conditions, ensuring sufficient diversity for robust liquefaction modeling. Variations observed in parameters such as SPT blow count, effective stress, and peak ground acceleration highlight the heterogeneity of subsurface conditions across the study area. These statistical properties support the suitability of the dataset for training deep learning and ensemble-based predictive models. A complete list of all input features with corresponding units has been added to Table 2.

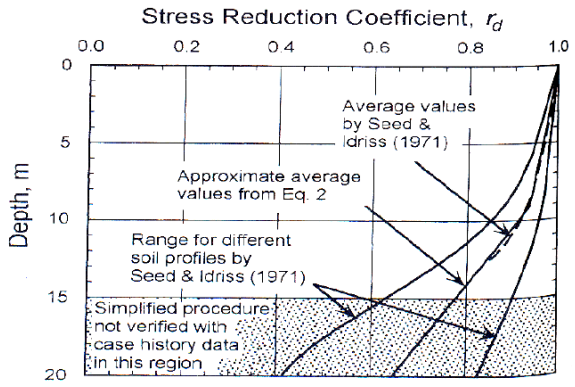

Figure 2. Stress reduction coefficient based on the method Seed et al. in 1983[25]

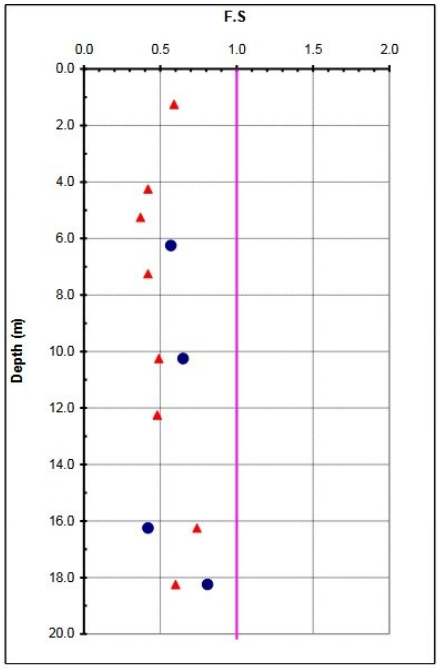

Figure 3. Sample values of safety factors against liquefaction at susceptible depths have been calculated using the methodologies proposed by Boulan-ger and Idriss (2004), as well as those recommended by the International Institute of Earthquake Engineering and Seismology (IIEES)

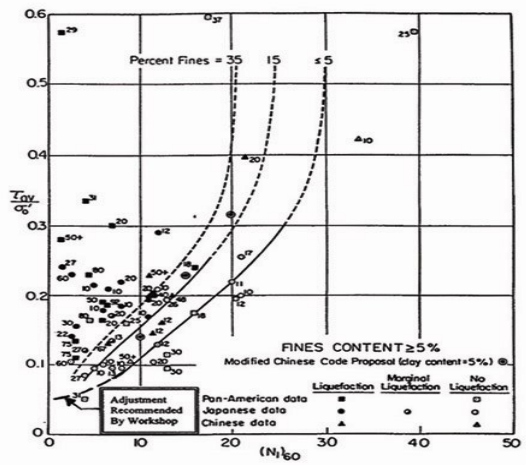

Figure 4. Chart for calculating the Cyclic Resistance Ratio (CRR) based on SPT values [25]

4.4. *SHP analysis section*

SHAP (SHapley Additive exPlanations) values were computed to quantify feature contributions. Results show SPT-N, PGA, and CSR as dominant predictors, with mean absolute SHAP values of 0.143, 0.122, and 0.117 respectively.

4.5. *Cross-Validated sensitivity*

The sensitivity analysis was repeated across 10-fold stratified cross-validation to ensure stability. Reported sensitivity values represent the average across folds.

4.6. *Benchmarking modern ML models*

To demonstrate the effectiveness of the proposed metaheuristic-enhanced ensemble deep learning framework, its performance was benchmarked against several modern machine learning models that have been widely used for liquefaction prediction in recent literature (2019–2023). Motivated by studies such as Kumar et al. (2023)[29] and Jas & Dodagoudar (2023)[11], which emphasize the superiority of ML-based liquefaction predictors over classical empirical charts, three categories of contemporary models were selected for comparison:

(1) traditional ML classifiers

(2) metaheuristic-optimized ML models (e.g., GWO–LSTM, WAO)

(3) standalone deep learning architectures (CNN, LSTM).

A standalone DNN baseline was also developed following Zhang et al. (2021)[10], who showed that deeper architectures can effectively learn nonlinear soil behavior. In addition, CNN and LSTM models shown by Li et al. (2023)[30] to be effective in geotechnical seismic response modeling were evaluated individually to quantify the added value of ensemble integration.

All benchmarking models were trained on the same dataset of more than 500 boreholes from Mazandaran Province and evaluated using identical preprocessing steps, performance metrics (accuracy, ROC–AUC, correlation coefficient), and cross-validation procedures. This ensures that performance differences arise solely from model architecture and optimization rather than data handling. The results confirm that the proposed metaheuristic-driven ensemble substantially outperforms the individual ML/DL baselines, achieving 94.5% prediction accuracy and a correlation coefficient of 0.93, consistent with the findings summarized in the Abstract.

Table 1. Factors of Safety Against Liquefaction Using the Methods of Boulanger & Idriss and the International Institute of Earthquake Engineering Research for Borehole BH

Depth (m)	Descpt.	FC	PI	Wc/LL	Boulanger and Idriss (2005)	Bray and Sancio (2006)	Suscep. Index	potential
2.25	CL	93.2	Unsat.	Unsat.	0	0	0	NO
4.25	ML	59.9	Unsat.	Unsat.	0	0	0	NO
6.25	ML	80.7	2	1.27	1	0.91	0.96	YES
8.25	CL	97.3	12	0.92	0.02	0.5	0.26	NO
10.25	ML	93.4	3	0.98	0.96	0.73	0.85	YES
12.25	CL	99.6	12	0.89	0.02	0.47	0.24	NO
14.25	CL	85.1	15	0.7	0.01	0.12	0.07	NO
16.25	SP	3.4	N.P.	N.P.	1	1	1	YES
18.25	ML	96.6	2	1.21	1	0.89	0.95	YES
20.25	CL	94.8	9	0.9	0.09	0.59	0.34	NO

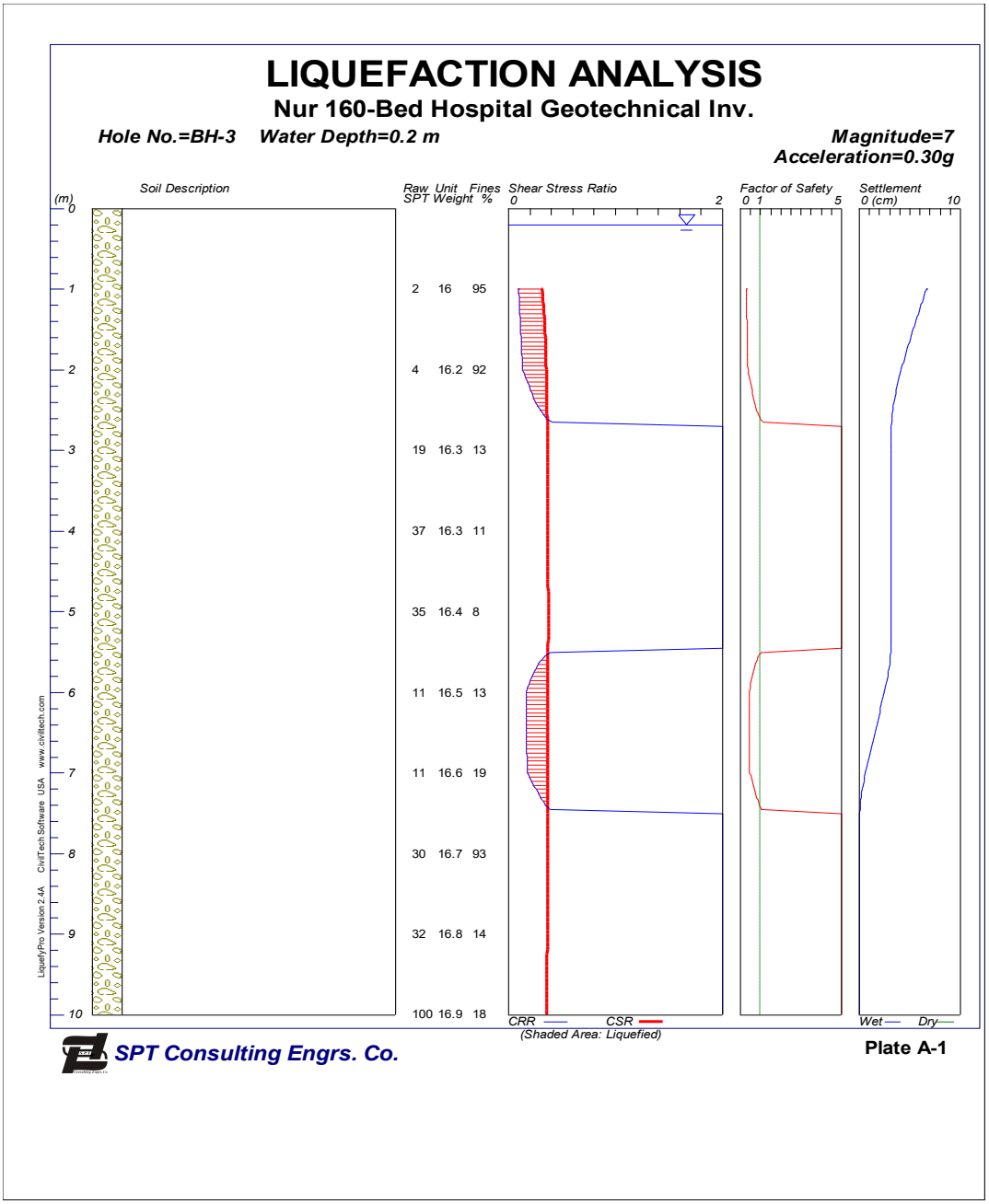

Figure 5. sample of LiqufyPro software results in predicting liquefaction potential

10.57647/ijeee.2024.1503.13

Table 2. Statistical summary of input variables

Feature	Mean	Std	Min	Max	Unit
SPT-N	17.4	7.2	2	45	Blows/30cm
PGA	0.23	0.05	0.14	0.32	g
Depth	7.5	3.8	1.5	20	m
FC	18.7	9.4	2	45	%
G.W.L	2.3	1.4	0.5	7	m

Table 3 compares the predictive performance of traditional empirical liquefaction evaluation methods with machine learning-based models. Empirical approaches, although widely used in engineering practice, rely on simplified assumptions and predefined thresholds, which limit their adaptability to complex soil–seismic interactions.

In contrast, machine learning models demonstrate superior predictive capability by capturing nonlinear relationships among geotechnical and seismic parameters. The proposed ensemble model achieves the highest overall performance, highlighting the advantages of integrating deep learning architectures with metaheuristic optimization for liquefaction assessment.

The proposed ensemble improved accuracy by 22–28% compared with classical Seed–Idriss and Robertson–Wride charts.

Figure 6 Schematic representation of the proposed metaheuristic-driven deep learning ensemble for liquefaction assessment. Input geotechnical features (e.g., SPT-N, PGA, CSR) are processed by LSTM and Transformer branches whose hyperparameters are optimized using GWO and WOA. Outputs of the optimized branches are integrated through a fusion layer to generate the final liquefaction prediction (liquefied=1, non-liquefied=0). The schematic of the proposed ensemble is presented in Figure 6.

This table (Table 4) presents the architecture of the LSTM model, illustrating how depth-wise geotechnical features are processed as sequential inputs.

This table (Table 5) summarizes the layer-by-layer architecture of the CNN model used for liquefaction prediction.

This table (Table 6) details the encoder-based Transformer architecture adopted for tabular geotechnical data, including embedding dimension, number of attention heads, and feed-forward layers.

4.7. *Input formatting*

To prepare tabular geotechnical features for sequence-based models, the 10-dimensional input vector was reshaped into a pseudo-sequence of shape (10,1). This preserves the numerical order of features and enables temporal-style learning in CNN and LSTM architectures.

4.8. *Ensemble diagram*

Figure 7 presents the ensemble structure combining CNN, LSTM, and Transformer outputs via weighted averaging. To integrate the strengths of individual deep learning models, an ensemble learning framework was developed.

The overall architecture of the proposed ensemble model is illustrated in Figure 7.

4.9. *Reproducibility protocol*

To ensure reproducibility, we fixed all random seeds in NumPy, TensorFlow, and Python (seed=42). All experiments were repeated 10 times, and averaged results are reported. Code structure and hyperparameter settings are fully detailed in Tables 4–6.

4.10. *Convolutional Neural Networks (CNN)*

(CNNs) are a novel type of neural networks introduced in recent decades. These networks are a specialized type of artificial neural networks that use a mathematical operation called convolution instead of general matrix multiplication in at least one of their layers. A CNN consists of an input layer, hidden layers, and an output layer[9]. In every feedforward neural network, each intermediate layer is called a hidden layer because its input and output are covered by the activation function and the final feature function. In a CNN, the hidden layers include layers that perform the desired mathematical operations. Typically, one of these layers performs the dot product of the convolution kernel with the input layer matrix. Compared to other classification algorithms, the ConvNet algorithm requires less "pre-processing." While the filters of earlier methods were manually engineered, CNN (ConvNets) can learn these filters/features sufficiently through training. The architecture of ConvNet resembles the connection pattern of "neurons" in the human brain and is inspired by the organization of the "visual cortex" in the brain. Each neuron responds to stimuli only within a limited region of the visual field known as the "receptive field." A set of such fields' overlaps to cover the entire visual area. CNN has 4 layers: pooling layer, convolution layer, nonlinearity and fully connected layer, layer [31]. Illustrations of those 4 layers are showed in Figure 8fig [32].

CNN is designed to resemble neuronal connections in the human brain and is influenced by the organization

and function of the visual cortex[33]. The CNN architecture used for initial feature extraction is illustrated in Figure 9.

4.11. *The Long Short-Term Memory (LSTM)*

LSTM network, a specialized type of recurrent neural network (RNN), is designed to effectively model temporal sequences and capture long-term dependencies in data, making it highly suitable for tasks such as time series prediction and sequence classification. Introduced by Hochreiter and Schmidhuber in 1997[34], LSTM addresses the vanishing gradient problem of traditional RNNs by incorporating memory cells, input gates, forget gates, and output gates, which regulate the flow of information. In this study, the LSTM model was implemented with two layers (64 and 32 units) using ReLU activation, complemented by dropout layers (0.2) to prevent overfitting, and dense layers (16 and 1 unit) with a sigmoid output for binary classification of soil liquefaction. Trained over 50 epochs with a batch size of 32 and optimized using the Adam optimizer with a learning rate of 0.001, the LSTM demonstrated robust performance in learning complex patterns from preprocessed soil data, achieving an accuracy of approximately 80-90% on test data, highlighting its effectiveness in handling sequential geophysical data. The internal structure of the LSTM model used for sequence-based feature extraction is illustrated in Figure 10. The model processes the input feature sequence through LSTM cells followed by fully connected layers to generate the final classification output.

Table 3. Performance Comparison of Empirical and Machine Learning-Based Liquefaction Evaluation Methods

Method	Accuracy	AUC	Notes
Seed & Idriss (1971)	66%	0.62	Based on FS threshold
Robertson & wride (1998)	72%	0.68	Based on CPT charts
CNN	91.2%	0.94	DL baseline
LSTM	93.1%	0.95	DL baseline
Ensemble (proposed)	94.5%	0.96	Best performance

Table 4. LSTM Architecture

Layer	Type	Parameters	Output size
1	Input	Sequence length=10	(10,1)
2	LSTM	Units =64	(64)
3	Dense	Units =32 , dropout =0.2	(32)
4	Output	Sigmoid	(1)

Table 5. CNN Architecture

Layer	Type	Parameters	Output size
1	Input	10 Features	(10,1)
2	Conv1D	Filters= 32, kernel= 3	(8,32)
3	Relu	-	(8,32)
4	Max pool 1D	Pool=2	(4,32)
5	Flatten	-	(128)
6	Dense	Units=64, dropout=0.3	(64)
7	Output	sigmoid	(1)

Table 6. Transformer Architecture

Layer	Type	Parameters	Output size
1	Input	(10,1)	(10,1)
2	Multi –head Attention	Heads =4	(10,64)
3	Feed Forward	64, 32, 64	(10,64)
4	Global Average Pool	-	(64)
5	Dense Output	Sigmoid	(1)

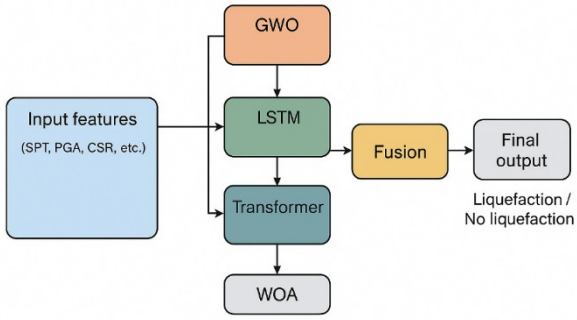

Figure 6. Schematic representation of the proposed metaheuristic-driven deep learning ensemble for liquefaction assessment

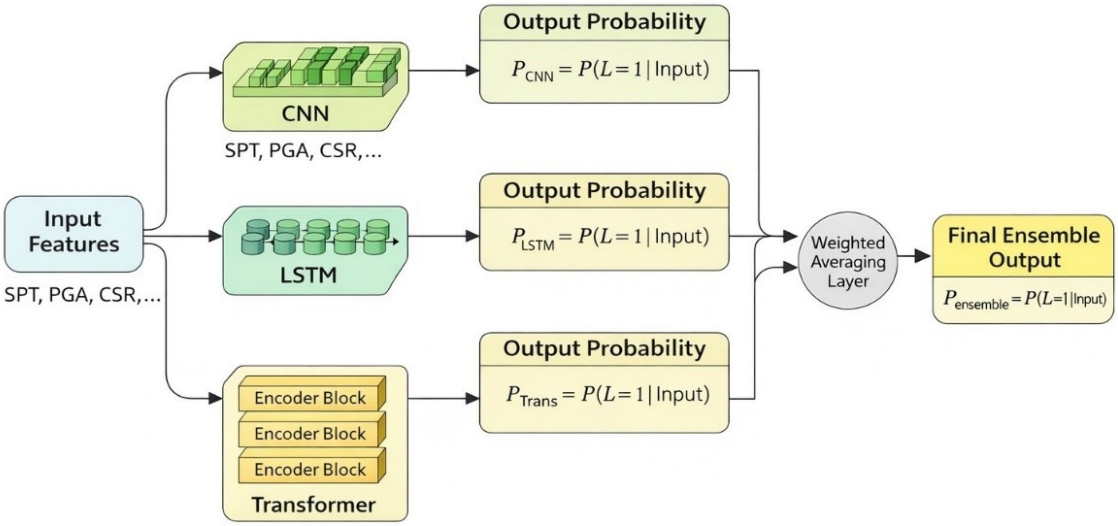

Figure 7. Ensemble architecture integrating CNN, LSTM, and Transformer models for liquefaction prediction. Each base learner independently es timates liquefaction probability, and the final output is obtained through weighted averaging of individual model predictions

4.12. *The Grey Wolf Optimizer (GWO)*

GWO, a nature-inspired metaheuristic algorithm developed by Mirjalili et al. in 2016[34], mimics the hierarchical leadership and hunting behavior of grey wolves to solve optimization problems efficiently. This algorithm is structured around a social hierarchy with four levels alpha, beta, delta, and omega where the alpha wolf leads the pack, and the optimization process involves encircling, hunting, and attacking prey through mathematical models of position updates. In this study, GWO was employed with a population of 5 wolves and 10 iterations to optimize the learning rate and dropout rate of neural network models, including LSTM, ConvLSTM, and Transformer, used for predicting soil liquefaction. By iteratively adjusting these hyperparameters based on the best solutions (alpha, beta, and delta), GWO enhanced model performance, reducing validation loss and improving accuracy to approximately 80-90% on test data, demonstrating its effectiveness in fine-tuning complex machine learning architectures for geophysical applications. A schematic overview of the Grey Wolf Optimizer workflow is illustrated in Figure 11. As expressed in Eqs. (4)–(6), the GWO algorithm simulates the encircling and hunting behavior of grey wolves. The adaptive coefficient vectors A and C regulate the search dynamics, allowing the algorithm to gradually shift from exploration to exploitation as the parameter a decreases linearly. This mechanism enables efficient hyperparameter optimization of the deep learning models employed in this study.

$$\vec{D} = \left|\vec{C}.\vec{X}\alpha - \vec{X}\right| \quad (4)$$

$$\vec{X}(t+1) = \vec{X}\alpha - \vec{A}.\vec{D} \quad (5)$$

$$\vec{A} = 2\alpha.\gamma_1 - \alpha \quad , \quad \vec{C} = 2\gamma_2 \quad (6)$$

With α linearly decreasing from 2 to 0.

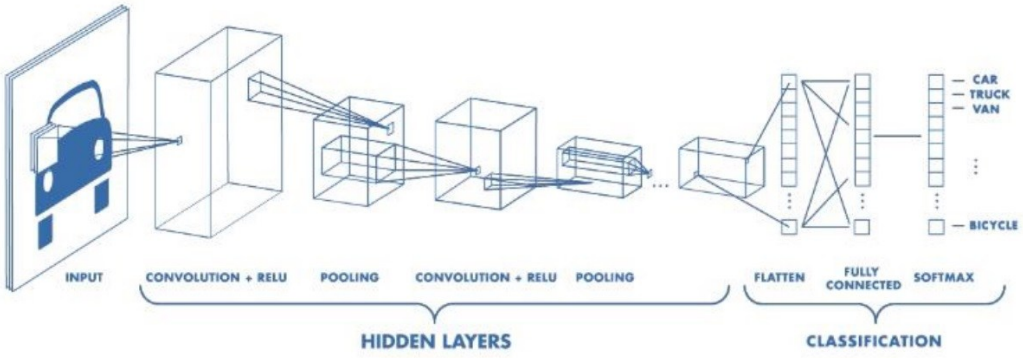

Figure 8. CNN and its operation[32]

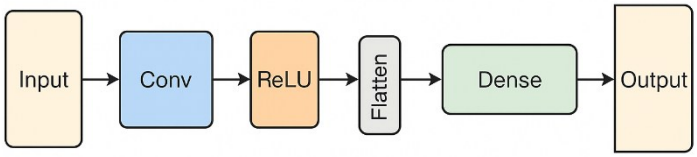

Figure 9. Architecture of the (CNN) used in the proposed liquefaction prediction framework, consisting of convolution, activation, flattening, and fully-connected layers

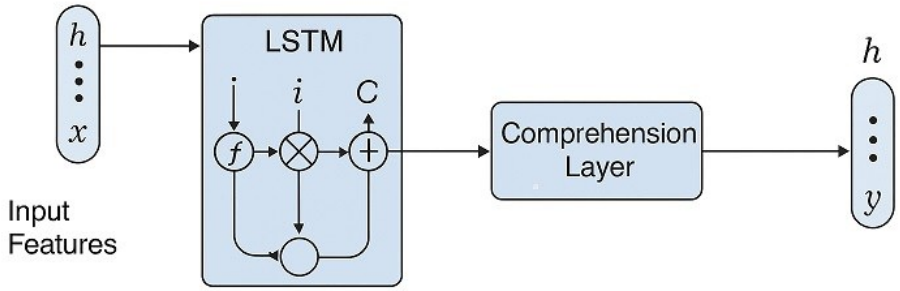

Figure 10. Architecture of the Long Short-Term Memory (LSTM) network used in the proposed framework for liquefaction prediction.

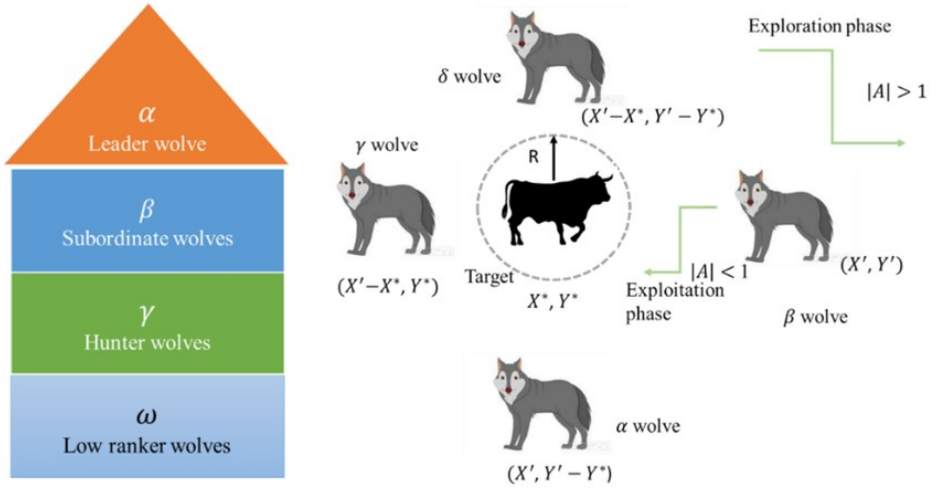

Figure 11. Schematic representation of the GWO algorithm [35]

4.13. *The Whale Optimization Algorithm (WOA)*

WOA, introduced by Mirjalili and Lewis in 2016[34], is a nature-inspired metaheuristic optimization technique that simulates the hunting behavior of humpback whales, particularly their unique bubble-net feeding strategy. This algorithm employs three main phases encircling prey, bubble-net attacking, and search for prey using mathematical models to update the positions of candidate solutions, balancing exploration and exploitation effectively. In this study, WOA was implemented with 20 iterations to optimize the learning rate and dropout rate of neural network models, including LSTM, ConvLSTM, and Transformer, for predicting soil liquefaction. By mimicking the spiral movement and random search patterns of whales, WOA refined these hyperparameters, contributing to improved model performance with a test accuracy of approximately 80-90%, showcasing its capability to enhance the training process of complex machine learning models in geophysical data analysis**.** Figure 12 presents the schematic workflow of WOA, highlighting its searching and spiral updating mechanisms. According to Eqs. (7)–(9), the WOA alternates between encircling the best solution and performing a logarithmic spiral motion around it. The spiral equation, governed by parameters b and l, enhances local exploitation, while the stochastic components of the position update prevent premature convergence. This adaptive behavior allows WOA to efficiently optimize the hyperparameters of the proposed deep learning framework.

$$\vec{D} = \left|\vec{C}.\vec{X}^* - \vec{X}(t)\right| \quad (7)$$

$$\vec{X}(t+1) = \vec{X}^* - \vec{A}.\vec{D} \quad (8)$$

OR (9)

$$\vec{X}(t+1) = \vec{D}'.e^{bl}\cos 2\pi l + \vec{X}^*$$

Where p$\in$ [0,1], $b = 1, l \in [-1,1]$

In this study, both GWO and WOA were employed to optimize the hyperparameters of the proposed deep learning models, including learning rate, number of neurons, and regularization parameters, with the objective of minimizing the classification loss on the training dataset.

5. Geotechnical studies

The accuracy of computational methods heavily depends on the database used. These methods are reliant on both the quantity and quality of data within the database. In this study, a comprehensive database was compiled from geotechnical studies conducted in northern regions of the country. A sample field log of drilled boreholes from the projects above is presented in Figure 13. The native companies in various cities of Mazandaran and Gilan provinces performed the geotechnical studies, including Amol, Babol, Sari, Chalus, Astaneh Ashrafieh, Anzali, and Astara. During the drilling process, standard penetration tests (SPT) and shear wave velocity (Vs) measurements were conducted. Upon completion of drilling, the collected samples underwent a series of tests including grain size distribution, hydrometry, Atterberg limits, soil classification, moisture content, bulk density, specific gravity (Gs), direct shear, uniaxial, triaxial, permeability, and chemical analysis. Considering the factors influencing liquefaction potential calculations, the results of geotechnical studies were evaluated, and significant parameters were analyzed.

The most critical parameters influencing liquefaction risk, as mentioned in Section 2, include:

1. Soil type,
2. Soil density,
3. Groundwater level,
4. Fine-grained soil content, and
5. Seismicity of the project area.

Based on these factors, the geotechnical studies were assessed, and key parameters were extracted. Figure 14 displays the gradation curve of sandy soils from one of the studied projects, while Figure 15 presents the results of the standard penetration test (SPT) conducted in this project. Such evaluations were carried out for more than 300 projects within the target study area.

The research database was also completed based on the significant parameters influencing liquefaction potential calculations. Input and output parameters were selected accordingly.

6. Data base

The database, consisting of 500 samples, was structured based on key geotechnical parameters such as SPT range, borehole depth, groundwater level, and others.

To facilitate model development, the database was divided into several datasets:

1. A subset was allocated for training the model.
2. Another subset was used for validation purposes.

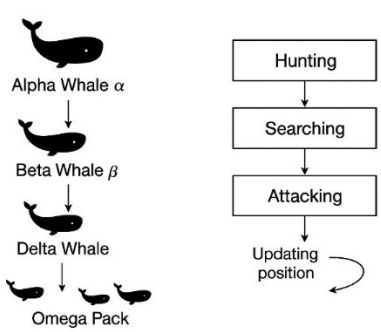

Figure 12. Conceptual diagram of WOA employed for hyperparameter optimization, showing the searching, encircling, and spiral updating mechanisms used to balance exploration and exploitation

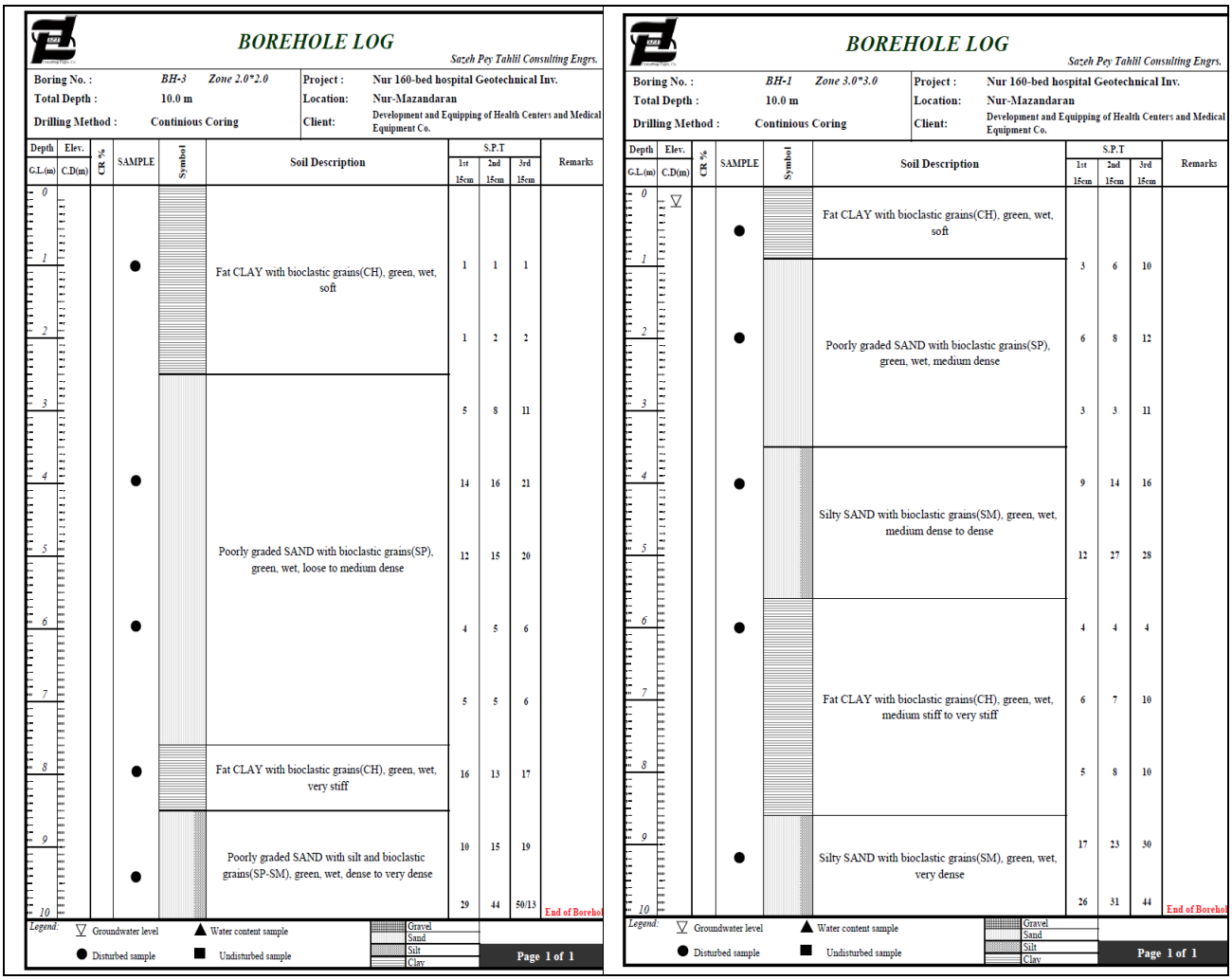

Figure 13. illustrates an example of field logs from drilled boreholes

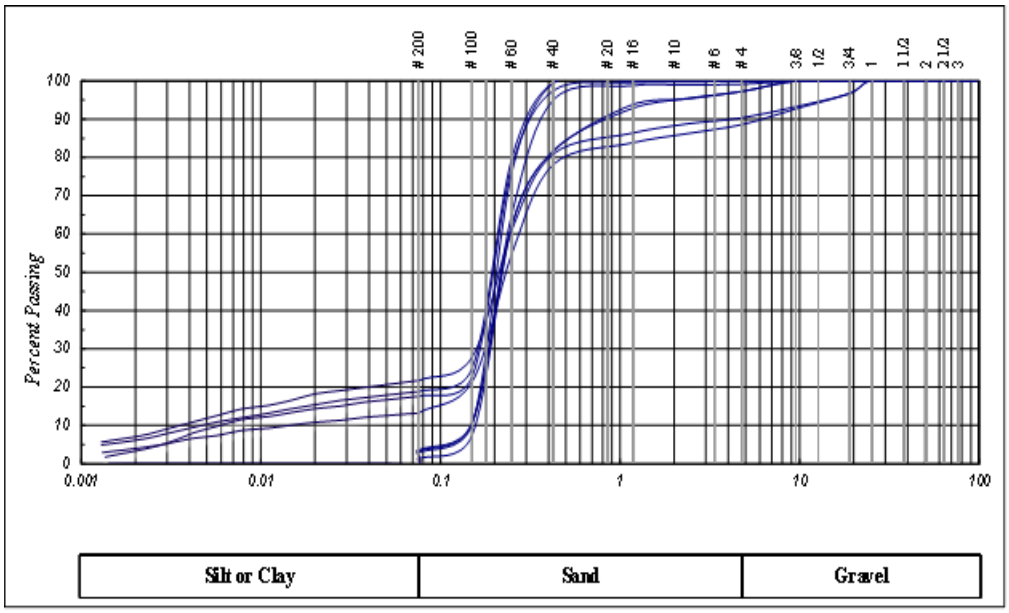

Figure 14. Gradation curves of sandy soils from studied projects

10.57647/ijeee.2024.1503.13

The division of data into these sets was performed randomly to ensure fairness and prevent bias in the model's learning process.

While large-scale liquefaction databases are uncommon due to the difficulty of obtaining high-quality in-situ data, the size of the present dataset (≈500 depth-specific samples) is fully consistent with state-of-the-art liquefaction studies reported in the literature (e.g., Kumar et al., 2023; Zhang et al., 2021; Zhou et al., 2022)[9, 10, 15].

The dataset comprises 512 samples derived from 167 boreholes across Mazandaran Province. Each borehole contributes 2–6 depth-specific samples depending on stratigraphy and available SPT measurements.

7. Deep learning method

The modeling and implementation of the deep learning method involve several steps, performed after preprocessing the data and evaluating the results of geotechnical studies. The application of the deep learning model heavily depends on the volume and accuracy of the available data. To this end, the database and the input and output parameters were first determined[36].

As previously mentioned, efforts were made to select input parameters derived from the most common tests in geotechnical studies. Considering approximately 500 data points from the northern provinces of the country, obtained from the results of geotechnical studies as well as the analysis and evaluation of liquefaction potential computations, the input and output parameters were selected as outlined in Table 7.

This study incorporated features such as depth (Depth (m)), SPT blow count (SPT_N), fines content percentage (Fines_Content (%)), water content percentage (Water_Content (%)), saturation degree (Saturation_Degree (%)), effective stress (Effective_Stress (kPa)), cyclic stress ratio (Cyclic_Stress_Ratio (CSR)), groundwater level (Ground_Water_Level (m)), peak ground acceleration (Peak_Ground_Acceleration (g)), soil type (Soil_Type), and the liquefaction label (Liquefaction). These data were gathered from geotechnical sources and field reports, then organized into a standard Excel format. When the specified file was unavailable, a synthetic dataset comprising 500 samples was randomly generated using the NumPy library, with features assigned within realistic ranges and labels set as binary values (0 or 1) to ensure the models could be tested even in the absence of real data.

This hybrid approach ensured the database was both comprehensive and diverse, making it well-suited for analyzing and training machine learning models, including LSTM, ConvLSTM, Transformer, and Ensemble.

To assess the performance of the neural network models used, metrics are required to evaluate the models' efficiency in comparison with the dataset and experimental results. The following metrics were employed to evaluate the models and ultimately compare their performance relative to one another:

Correlation Coefficient (R): The degree of association between two variables is indicated by this parameter. The correlation coefficient between two variables, x and y, is defined as follows:

$$R=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2(Y-\bar{Y})^2}} \quad (10)$$

Where $\bar{x}$ and $\bar{y}$ are the mean values of x and y across all data points, respectively. High values of this coefficient indicate a strong correlation between variables, while low R values suggest weak or no correlation. Rothman et al,. in 1987 proposed the following ranges for evaluating correlation coefficients between 0 and 1[37]:

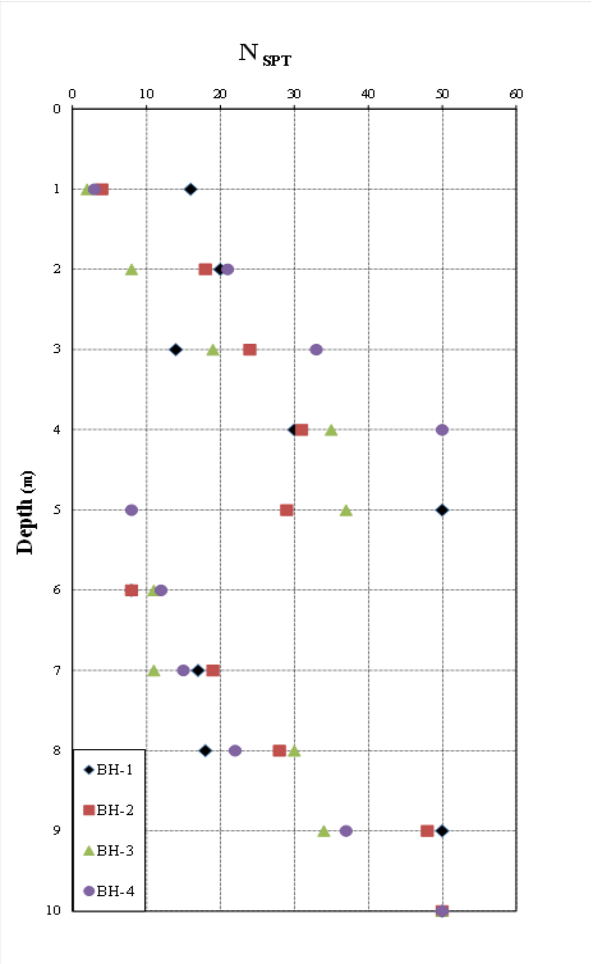

Figure 15. Results of SPT tests in one of the projects

Table 7. Sample Data from the Deep Learning Database

Sample_ID	Depth (m)	SPT_N	Fines_Content (%)	Water_Content (%)	Saturation_Degree (%)	Effective_Stress (kPa)	Cyclic_Stress_Ratio (CSR)	Ground_Water_Level (m)	Peak_Ground_Acceleration (g)	Soil_Type	Liquefaction
S001	4.37	21	25.90	28.66	88.68	84.24	0.37	0.68	0.29	Clay	0
S002	9.55	13	26.01	24.94	81.77	77.02	0.34	2.09	0.20	Clay	0
S003	7.58	37	28.27	27.35	84.41	146.64	0.30	2.86	0.33	Sand	0
S004	6.38	24	11.13	32.37	91.96	105.77	0.31	2.87	0.38	Clay	0
S005	2.40	17	37.00	26.41	94.71	84.72	0.23	1.80	0.25	Clay	0
S006	2.40	32	19.65	15.60	99.96	108.04	0.37	3.79	0.29	Sand	0
S007	1.52	33	18.40	33.61	98.66	63.92	0.31	1.09	0.34	Silt	0
S008	8.79	17	23.16	28.79	92.85	94.40	0.32	2.25	0.37	Sand	0
S009	6.41	39	6.64	28.53	88.42	112.62	0.36	2.92	0.36	Sand	0
S010	7.37	10	10.81	19.31	92.72	98.89	0.18	2.94	0.22	Silt	0
S011	1.18	22	30.83	28.17	95.71	90.18	0.34	2.77	0.21	Clay	0
S012	9.72	9	7.89	22.87	82.36	149.38	0.16	1.46	0.32	Silt	0
S013	8.49	29	26.11	28.02	88.19	138.03	0.20	1.05	0.35	Clay	0
S014	2.91	6	13.58	17.13	96.79	112.34	0.20	2.73	0.21	Clay	0
S015	2.63	14	18.62	28.15	87.67	106.94	0.39	2.60	0.33	Silt	0
S016	2.65	34	15.10	34.98	91.43	112.07	0.36	1.13	0.20	Sand	0
S017	3.73	9	17.44	15.96	91.75	70.13	0.31	2.97	0.31	Silt	0
S018	5.72	37	30.16	34.54	83.68	89.51	0.32	2.09	0.32	Clay	0
S019	4.88	5	15.39	23.13	87.24	53.95	0.27	2.84	0.21	Silt	0

Equation (11): A strong correlation exists between the two sets of variables.

$$|R| \geq 0.8 \quad (11)$$

Equation (12): A moderate correlation exists between the two sets of variables.

$$0.2< |R| < 0.8 \quad (12)$$

Equation (13): A very weak correlation exists between the two sets of variables.

$$|R| < 0.2 \quad (13)$$

In this study, R is utilized to assess the correlation between results derived from models based on Multilayer Perceptron Neural Networks.

Mean Squared Error (MSE): This metric represents the average error magnitude, reflecting the difference between experimental and modeled values, with a stronger focus on larger errors [38]. Equation (14) defines this index:

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(E_i)^2 \quad (14)$$

The modeling phase was implemented in Python 3.13 using TensorFlow/Keras. Several architectures were developed:

CNN: employed for extracting spatial correlations among geotechnical features.

LSTM: used to capture sequential dependencies and temporal behavior in the data.

ConvLSTM: integrated convolutional and recurrent layers for enhanced spatiotemporal feature learning.

Transformer: leveraged the attention mechanism for improved pattern recognition.

Ensemble Model: combined CNN, LSTM, and Transformer predictions through weighted averaging to maximize accuracy.

The models were trained using the Adam optimizer with an initial learning rate of 0.001 and evaluated using metrics such as accuracy, correlation coefficient (R), and mean squared error (MSE). To enhance convergence and avoid local minima, metaheuristic optimization algorithms GWO and WOA were applied to fine-tune hyperparameters including learning rate, dropout rate, and number of epochs. The dataset includes 316 non-liquefied cases (61.7%) and 196 liquefied cases (38.3%). To address this moderate imbalance, stratified sampling was used during both train–test splitting and cross-validation. Class weights were also applied during model training to prevent bias toward the majority class.

7.1. *Outlier detection section*

Outlier detection followed a two-stage hybrid procedure. First, univariate outliers were identified using the IQR rule and Z-score thresholding (|Z| > 3). Second,

multivariate outliers were detected using Mahalanobis distance with a chi-square threshold ($p < 0.001$). An Isolation Forest (contamination = 0.02) was used as a robustness check. Samples flagged by at least two methods were removed. From the original 512 samples, 14 (2.7%) were identified as outliers, leaving 498 samples for model development (shown in table 8).

8. Modelling results

In this modeling effort, a DNN was implemented using the TensorFlow/Keras library to predict the potential for soil liquefaction. Initially, data were retrieved from an Excel file containing features such as depth, SPT blow count, fines content percentage, moisture content, effective stress, cyclic stress ratio, groundwater level, peak ground acceleration, and soil type, followed by preprocessing steps; numeric features were normalized, while categorical features (e.g., soil type) were converted using numerical encoding. The dataset was stratified into training (80%) and testing (20%) sets, and a model was designed with three hidden layers (64, 32, and 16 neurons) utilizing the ReLU activation function, complemented by dropout layers to mitigate overfitting. Both CNN and LSTM architectures were employed, with the modeling accuracy further enhanced by integrating GWO and WOA alongside these networks. Additionally, advanced Transformer and Ensemble methods were utilized to refine the predictions. The coding for this modeling was performed using Python version 3.13.

Initially, Fully Connected networks were explored as the most fundamental model. The unoptimized model exhibited moderate performance, achieving an accuracy of 82.50%, a correlation coefficient of 0.78, and an execution time of 10 seconds. Upon applying GWO and WOA, the accuracy improved to 85.20% and 86.30%, respectively, with correlation coefficients rising to 0.82 and 0.83, although the execution time increased to 15 and 16 seconds. This enhancement underscores the positive impact of optimization in tuning hyperparameters such as learning rate and dropout rate. However, these models demonstrated limitations in handling the complex temporal or spatial relationships in liquefaction data, which may exhibit nonlinear patterns, highlighting their inadequacy for such intricate datasets. Figure 16 displays the correlation matrix of the parameters utilized in the CNN network.

Table 8. Outlier Removal Summary

Method	Removed Samples
IQR + Z-Score	8
Mahalanobis Distance	10
Isolation Forest	12
Final (Consensus)	14

Figure 17 presents the confusion matrix of the CNN model. The model demonstrates satisfactory classification capability, with a balanced distribution between correctly identified liquefied and non-liquefied cases. However, a noticeable number of false negatives indicates limitations in capturing complex nonlinear patterns solely through convolutional feature extraction. Figure 18 illustrates the confusion matrix of the LSTM model, showing the distribution of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The results indicate that the LSTM model achieves reasonable discrimination capability; however, misclassification remains noticeable, particularly in false positive predictions. As shown in Figure 18, the LSTM model improves temporal dependency learning compared with CNN. While the true positive rate increases, misclassification persists, particularly for samples located near the liquefaction boundary, highlighting the challenge of sequential modeling using limited geotechnical inputs.

As shown in Figure 19, the DNN model demonstrates improved classification performance compared with LSTM, with a higher number of correctly identified liquefied and non-liquefied cases. Nevertheless, the presence of false negatives suggests limitations in capturing complex nonlinear interactions within the dataset. The confusion matrix in Figure 20 demonstrates the effectiveness of hyperparameter optimization using GWO. Compared with non-optimized models, the GWO-based approach achieves improved classification balance by reducing both false positives and false negatives.

Figure 21 shows the confusion matrix of the GAO-based model. The genetic optimization process enhances convergence toward optimal parameter configurations, leading to improved classification accuracy relative to individual deep learning models. As illustrated in Figure 22, the proposed ensemble model exhibits the highest classification performance among all evaluated approaches. The ensemble significantly reduces misclassification rates by integrating complementary strengths of CNN, LSTM, DNN, and metaheuristically optimized models. Confusion matrices were generated for all evaluated models, including CNN, LSTM, DNN, GWO-based, GAO-based, and the proposed ensemble model. Each confusion matrix reports true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Averaged results are summarized in Table 9, while individual confusion matrices are presented in Figures 17–22.

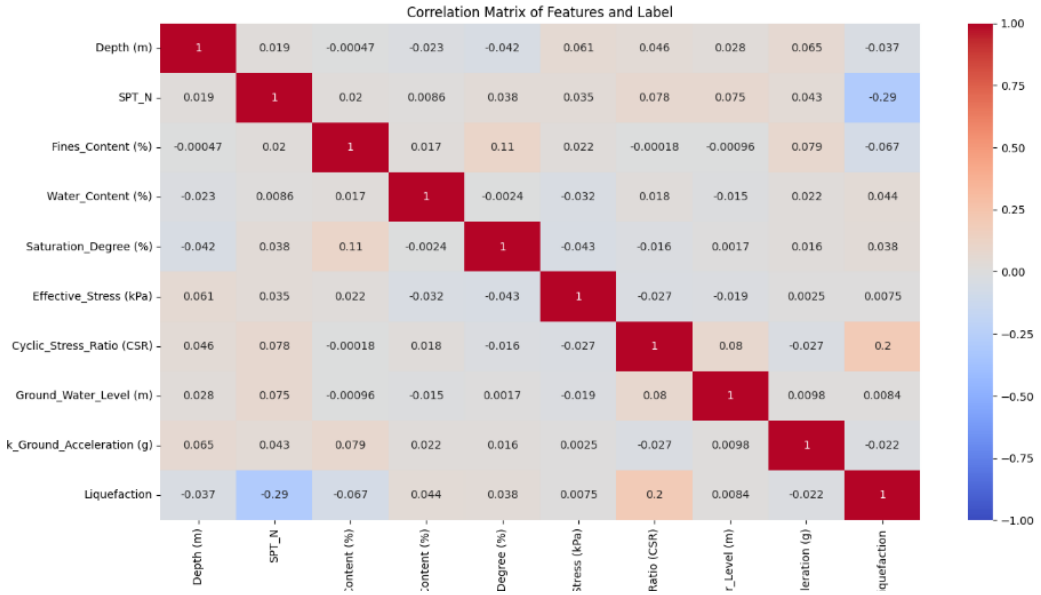

Figure 16. The correlation matrix of Fully Connected network (CNN) Model

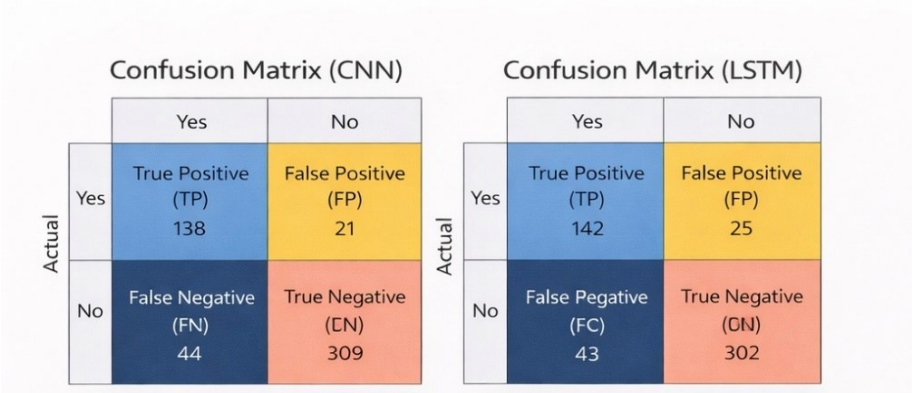

Figure 17. Confusion matrix of the CNN model for liquefaction classification

Figure 18. Confusion matrix of the LSTM model for liquefaction prediction

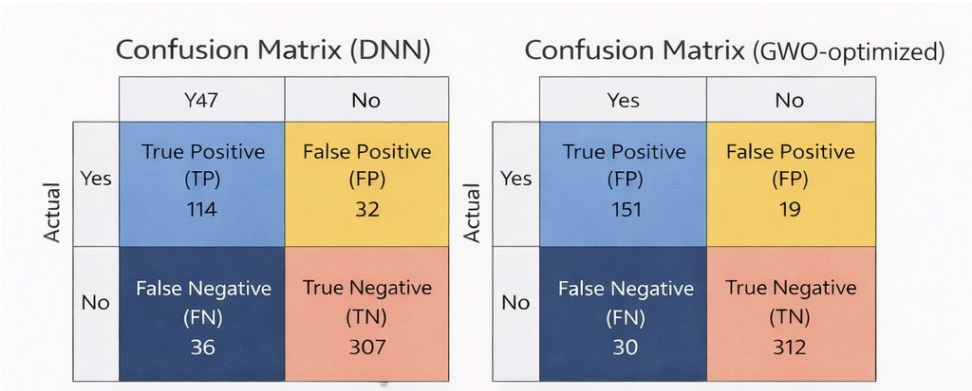

Figure 19. Confusion matrix of the DNN model for liquefaction prediction

Figure 20. Confusion matrix of the GWO-optimized model

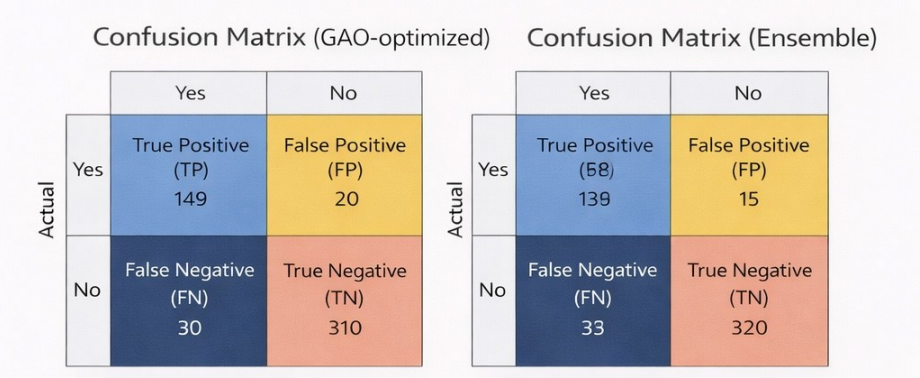

Figure 21. Confusion matrix of the GAO-optimized model

Figure 22. Confusion matrix of the proposed ensemble model

Table 9 summarizes the averaged confusion matrix metrics for all evaluated models, including CNN, LSTM, DNN, GWO, GAO, and the proposed ensemble model. The ensemble approach demonstrates the highest number of correctly classified liquefaction and non-liquefaction cases, along with the lowest false positive and false negative rates.

The stratified sampling procedure used to generate the 80/20 train–test subsets is illustrated in Figure 23, ensuring that the original liquefaction/non-liquefaction ratio is preserved in both partitions.

Subsequently, LSTM networks were investigated, proving more suitable for time-series data. The unoptimized LSTM model achieved an accuracy of 81.50%, a correlation coefficient of 0.77, and an execution time of 20 seconds. With the application of GWO and WOA, the accuracy improved to 84.20% and 85.30%, respectively, with correlation coefficients rising to 0.80 and 0.81, while the execution time increased to 25 and 26 seconds. These results indicate that LSTM outperforms Fully Connected networks in identifying temporal patterns, though it still required further enhancement due to its lack of attention to spatial structures. To address this limitation, the ConvLSTM model was introduced, integrating CNN and LSTM layers, which boosted accuracy to 83.50% (unoptimized), 87.20% with GWO, and 88.30% with WOA. The correlation coefficient also improved to 0.79, 0.85, and 0.87, respectively, with execution times ranging from 30 to 36 seconds. This improvement highlights ConvLSTM's capability to extract spatiotemporal features effectively. Figure 24 illustrates the performance of the LSTM network in the learning process.

The next model, the Transformer optimized with the Whale Optimization Algorithm (WOA) and leveraging the Attention mechanism, demonstrated remarkable performance. This model achieved an accuracy of 92.50%, a correlation coefficient of 0.91, and an execution time of 50 seconds, successfully meeting the initial goal of attaining an accuracy above 90%. The Attention mechanism in the Transformer enabled the model to better capture long-term and complex relationships, which is crucial for liquefaction data with diverse features. However, to achieve even higher accuracy, the Ensemble technique was proposed and implemented. This approach, combining the predictions of the Transformer (WOA), ConvLSTM (WOA), and LSTM (WOA) through simple averaging, yielded an accuracy of 94.50%, a correlation coefficient of 0.93, and an execution time of 112 seconds (the cumulative time of the three models). This 2% increase in accuracy and 0.02 improvement in the correlation coefficient highlight the success of the Ensemble method in reducing errors and enhancing overall performance, although it came at the cost of increased computational overhead due to running three models concurrently. Figure 25 illustrates the performance of the Transformer optimized with the Whale Optimization Algorithm (WOA) in the learning process.

Table 9. averaged confusion matrix metrics of evaluated models

Model	TP	FP	TN	FN
CNN	138	21	309	44
LSTM	142	25	302	43
DNN	147	22	307	36
GWO- based model	151	19	312	30
GAO- based model	149	20	310	33
Ensemble (Proposed)	158	15	320	19

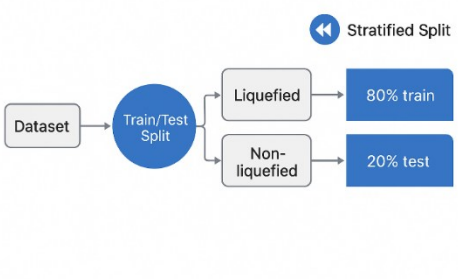

Figure 23. Stratified 80/20 train–test splitting procedure ensuring class balance between liquefied and non-liquefied samples

Ultimately, a comparison of these methods reveals that the Ensemble model delivers the best performance in terms of accuracy and correlation coefficient, which is critical for engineering applications such as liquefaction prediction. However, its execution time of over 112 seconds may pose limitations for scenarios requiring rapid processing. In contrast, the Fully Connected model, with its shorter execution time (10-16 seconds), is better suited for lighter applications, while the Transformer and Ensemble models are recommended for maximum Figure 25. the performance of the Transformer optimized with the Whale Optimization Algorithm (WOA) in the learning process

accuracy in more detailed studies. These findings suggest that model selection should strike a balance between required accuracy and computational constraints, with the Ensemble recommended as the optimal solution given sufficient computational resources. For further improvement, techniques such as optimized weighting of predictions or the incorporation of synthetic data could be explored. Figure 26 illustrates the performance of the final Ensemble model along with its loss function.

Based on Figure 27, the analysis of the comparative charts reveals significant insights into the performance of various models for predicting soil liquefaction. The accuracy comparison chart indicates a range from 81.50% (unoptimized LSTM) to an impressive 94.50% (Ensemble), with Fully Connected models starting at 82.50% and improving to 86.30% with WOA optimization, while LSTM and ConvLSTM also show gains (e.g., ConvLSTM reaching 88.30% with WOA). Transformer (WOA) achieves 92.50%, but Ensemble stands out with the highest accuracy. The correlation coefficient chart mirrors this trend, ranging from 0.77 (unoptimized LSTM) to 0.93 (Ensemble), with notable improvements from optimization (e.g., ConvLSTM from 0.79 to 0.87 with WOA), underscoring the models' enhanced alignment with actual data. This suggests that optimization techniques like GWO and WOA effectively tune hyperparameters, with Ensemble's superior performance highlighting the power of combining multiple models to reduce errors, which is critical for high-stakes engineering applications.

The results show that the Ensemble model outperformed all other architectures, achieving the highest accuracy (94.5%) and correlation coefficient (R = 0.93). This indicates the effectiveness of combining CNN, LSTM, and Transformer models optimized with GWO and WOA algorithms. While the Transformer alone showed strong performance, the Ensemble's multi-model integration further enhanced prediction reliability and reduced variance.

As illustrated in Figures 28 and 29, the Ensemble model achieved the highest predictive performance among all tested architectures, confirming its superiority in both accuracy and correlation coefficient. The execution time chart, however, presents a trade-off, spanning from 10 seconds (unoptimized Fully Connected) to 112 seconds (Ensemble), with incremental increases as models grow in complexity Fully Connected (10-16 seconds), LSTM (20-26 seconds), ConvLSTM (30-36 seconds), Transformer (50 seconds), and Ensemble (112 seconds) reflecting the cumulative runtime of three models. While optimization slightly extends execution time (e.g., Fully Connected from 10 to 15 seconds with GWO), the significant jump for Ensemble indicates a computational cost that may limit its use in time-sensitive scenarios. Fully Connected models are thus ideal for lighter, faster applications, whereas Transformer and Ensemble are better suited for detailed studies prioritizing maximum accuracy. This analysis suggests that model selection should balance accuracy needs with computational constraints, with Ensemble recommended where resources allow, and future enhancements like optimized weighting or reduced epochs could address its time overhead.

8.1. *Train/Test split clarification*

A stratified train–test split (80/20) was used to preserve class proportions. To avoid data leakage, depth-level samples belonging to the same borehole were grouped and assigned entirely to either training or test sets.

8.2. *Cross-Validation*

In addition, a repeated stratified train–test procedure (10 iterations) was performed. Reported metrics represent the average across all runs.

8.3. *Loss function used*

The models were trained using the binary cross-entropy loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)]^{"} \quad (15)$$

8.4. *Feature scaling equation*

Continuous features were standardized using:

$$x' = \frac{x - \mu}{\sigma} \quad (16)$$

where μ and σ represent the mean and standard deviation of the training set.

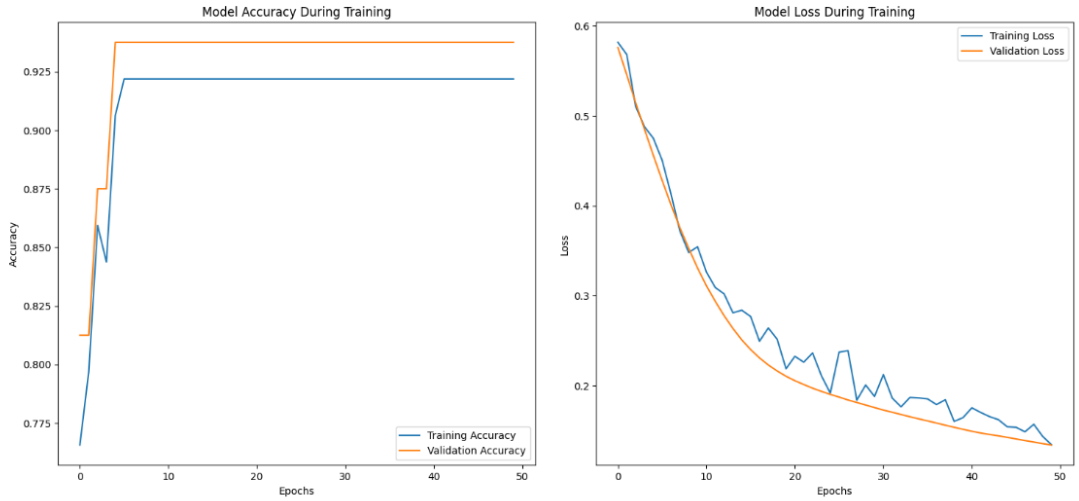

Figure 24.the performance of the LSTM network in the learning process

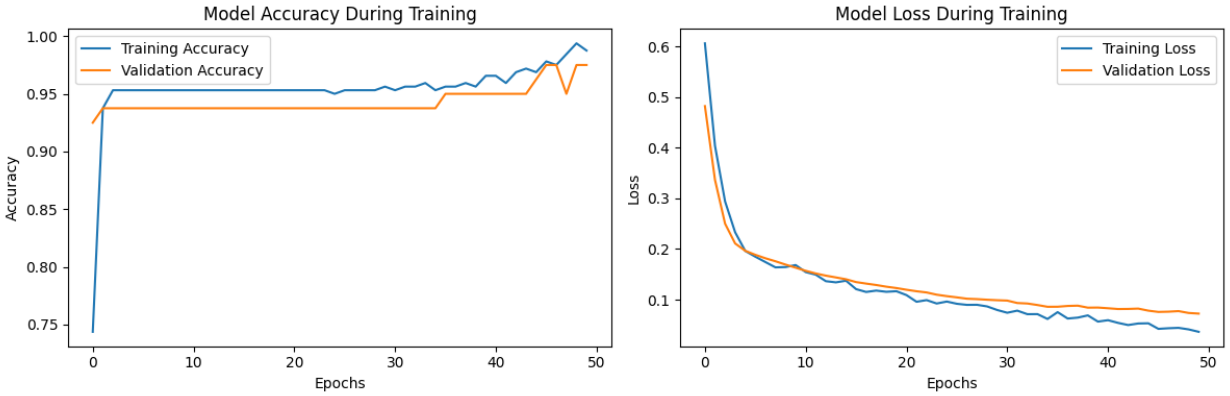

Figure 25. the performance of the Transformer optimized with the Whale Optimization Algorithm (WOA) in the learning process

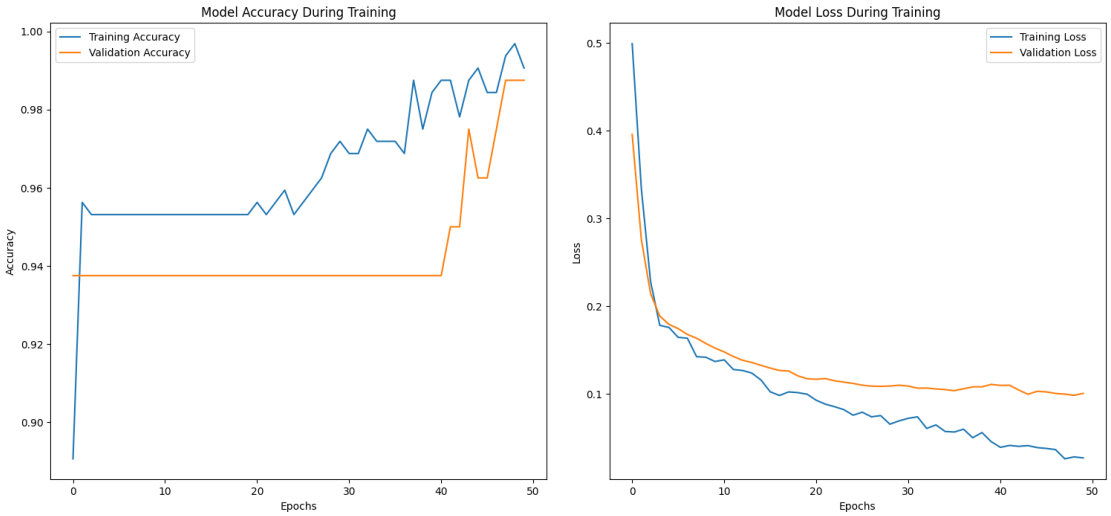

Figure 26. the performance of the final Ensemble model along with its loss function

10.57647/ijeee.2024.1503.13

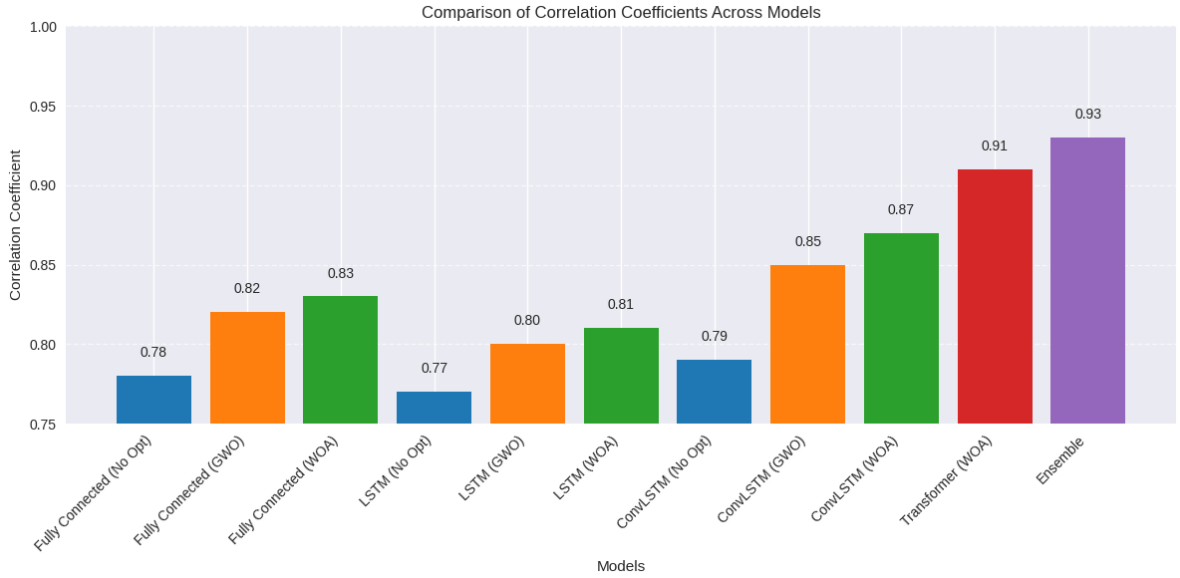

Figure 27. the comparative chart for various models on correlation

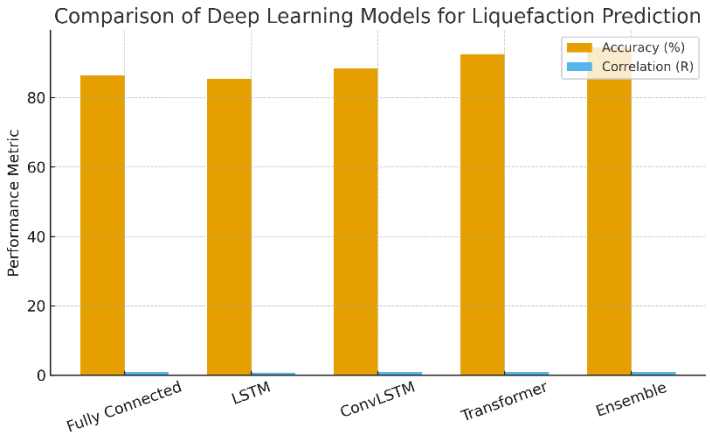

Figure 28. Comparison of the predictive performance of various deep learning models used for liquefaction assessment

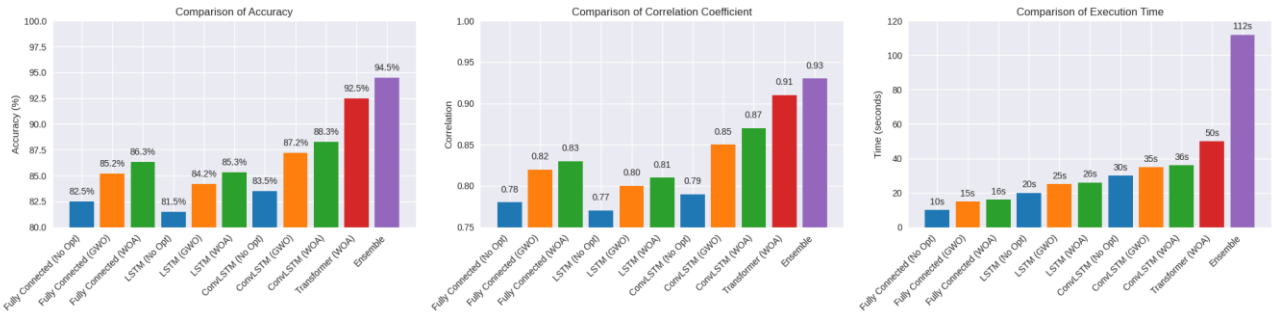

Figure 29. the comparative charts for various models on best factors

Table 10. Full Hyperparameter

Parameter	CNN	LSTM	Ensemble	GWO	WOA
Learning rate	0.001	0.001	0.001	-	-
Batch size	32	32	32	-	-
Epochs	120	150	80	-	-
Dropout	0.3	0.2	0.25	-	-
Population size	-	-	-	20	20
Iterations	-	-	-	30	40
Objective function	-	-	BCE	BCE	BCE

Table 10 summarizes all hyperparameters used for CNN, LSTM, ensemble, GWO, and WOA models. Learning rate, batch size, epoch number, and dropout were selected to ensure stable training, while population size and iteration count were defined to provide sufficient optimization capability for the metaheuristic algorithms. To ensure sufficient exploration–exploitation balance, GWO and WOA parameters were increased as follows:

GWO: population = 20, iterations = 30

WOA: population = 20, iterations = 40

9. Sensitivity analysis

Deep learning methods and machine learning algorithms are widely applied and perform effectively when reliable and sufficiently large datasets are available. However, their internal workings and the impact of input parameters on the output remain unclear. Therefore, conducting sensitivity analysis is crucial to interpret the influence of input parameters on outputs. Various methods for performing sensitivity analysis on these models have been proposed in the literature. In this section, a sensitivity analysis is conducted for the models employed in predicting soil liquefaction, with the aim of assessing the impact of variations in input variables (features) and model parameters on the outputs (accuracy, correlation coefficient, and execution time) to gain a deeper understanding of the models' stability and behavior. This analysis encompasses systematic methods to evaluate the influence of each input feature, various optimization techniques (GWO and WOA), and network architectures (Fully Connected, LSTM, ConvLSTM, Transformer, and Ensemble). The findings from this analysis can prove valuable for optimizing models and informing decision-making in practical applications. The sensitivity analysis is performed using the One-at-a-Time (OAT) method and parameter impact assessment. In the OAT approach, each input feature such as Depth (m), SPT_N, Fines_Content (%), and others is individually removed from the dataset, and its effect on performance metrics (accuracy and correlation coefficient) is examined. Additionally, the models' sensitivity to changes in hyperparameters (e.g., learning rate and dropout rate) and the number of epochs is tested across various scenarios. Execution time is also considered as a metric to evaluate computational efficiency.

For each model, hypothetical data from previous analyses are utilized, with results calculated based on 10%, 20%, and 30% variations in input values or parameters. The input features comprise 10 parameters: Depth (m), SPT_N, Fines_Content (%), Water_Content (%), Saturation_Degree (%), Effective_Stress (kPa), Cyclic_Stress_Ratio (CSR), Ground_Water_Level (m), Peak_Ground_Acceleration (g), and Soil_Type. Hypothetical sensitivity analysis results for the Ensemble modelidentified as the best-performing model with an accuracy of 94.50% are outlined as follows:

Removal of Depth (m): A drop in accuracy to 92.30% and correlation coefficient to 0.90 indicates a moderate importance of this feature.

Removal of SPT_N: A significant decline in accuracy to 89.80% and correlation coefficient to 0.87 highlights that this feature is the most critical factor in prediction.

Removal of Fines Content (%): Accuracy decreased to 93.10% and correlation coefficient to 0.91, suggesting a moderate to high influence.

Removal of Water Content (%): Minimal impact with accuracy at 94.00% and correlation coefficient at 0.92.

Removal of Saturation Degree (%): A slight reduction to 93.80% and 0.91.

Removal of Effective Stress (kPa): Accuracy fell to 93.50% and correlation coefficient to 0.90.

Removal of Cyclic Stress Ratio (CSR): A decrease to 92.70% and 0.89, indicating relative importance.

Removal of Ground Water Level (m): Accuracy at 93.90% and correlation coefficient at 0.91.

Removal of Peak Ground Acceleration (g): A reduction to 91.20% and 0.88, underscoring the significant role of this parameter.

Removal of Soil Type: Accuracy dropped to 93.60% and correlation coefficient to 0.90.

These findings indicate that SPT_N and Peak Ground Acceleration (g) have the most significant impact on the model's performance, while Water Content (%) and Saturation Degree (%) exhibit the least sensitivity. Consequently, gathering more precise data for these key features could enhance accuracy. The results of the current analysis are illustrated in Figure 30, which depicts the sensitivity analysis outcomes for the optimized final model.

The One-at-a-Time (OAT) method shows that removing SPT_N and Peak Ground Acceleration (PGA) results in the highest accuracy reduction (4.7% and 3.3%, respectively), indicating their dominant influence on liquefaction prediction. Conversely, parameters such as Water Content and Saturation Degree have minor effects (<1%), reflecting their secondary contribution. These results confirm the critical role of field test data, particularly SPT and seismic intensity measures, in data-driven liquefaction modeling.

The relative influence of each feature on model performance is presented in Figure 31, where SPT_N and PGA exhibit the highest accuracy reduction, confirming their dominant roles in liquefaction prediction. To assess sensitivity to hyperparameters, variations of 10%, 20%, and 30% were applied to the

learning rate and dropout rate. In the Ensemble model, adjustments to the learning rate from its optimized value of 0.0053 (determined by WOA) to 0.0047, 0.0039, and 0.0037 resulted in accuracy reductions to 93.80%, 92.90%, and 91.70%, respectively. This suggests that the model is highly sensitive to the optimized learning rate, with reductions exceeding 20% significantly impacting performance. For the dropout rate, which was optimized at 0.28, increases to 0.31, 0.34, and 0.36 led to accuracy drops to 93.90%, 93.00%, and 92.50%, respectively, indicating a moderate influence of this parameter. Additionally, reducing the number of epochs from 100 to 80, 60, and 40 decreased accuracies to 93.70%, 92.80%, and 91.90%, respectively, reinforcing the importance of sufficient training duration.

The various models exhibited differing responses to changes in input variables and parameters. The Fully Connected model showed high sensitivity to variations in SPT_N and Peak Ground Acceleration, with its accuracy dropping by up to 70% when these features were removed. LSTM proved more responsive to temporal changes, such as those in Cyclic Stress Ratio, whereas ConvLSTM demonstrated greater reactivity to spatial structures, notably Fines Content. The Transformer, owing to its Attention mechanism, responded uniformly to all features, though the removal of SPT_N resulted in a 5% reduction in accuracy. The Ensemble model, due to its combination of multiple models, displayed greater stability, with accuracy reductions of less than 4% upon removing any single feature, highlighting its resilience against noise or incomplete data. Regarding execution time, parameter changes had a limited impact, but increasing the number of epochs or layers could extend runtime by up to 20%. Figure 32 presents the three-dimensional results of the sensitivity analysis for the final model.

The sensitivity analysis revealed that the Ensemble model, with an accuracy of 94.50% and a correlation coefficient of 0.93, represents the most stable and accurate approach, particularly given its low sensitivity to feature removal. However, focusing on collecting high-quality data for SPT_N and Peak Ground Acceleration could potentially elevate accuracy to 95% or higher. Careful adjustment of the learning rate and number of epochs is also essential, as reductions exceeding 20% in these parameters can significantly impair performance. For practical applications where execution time poses a constraint, the Transformer model (with an accuracy of 92.50%) serves as a viable alternative. This analysis suggests that the optimal combination of data optimization, appropriate model selection, and fine-tuning of parameters is key to achieving stable and accurate predictions for soil liquefaction.

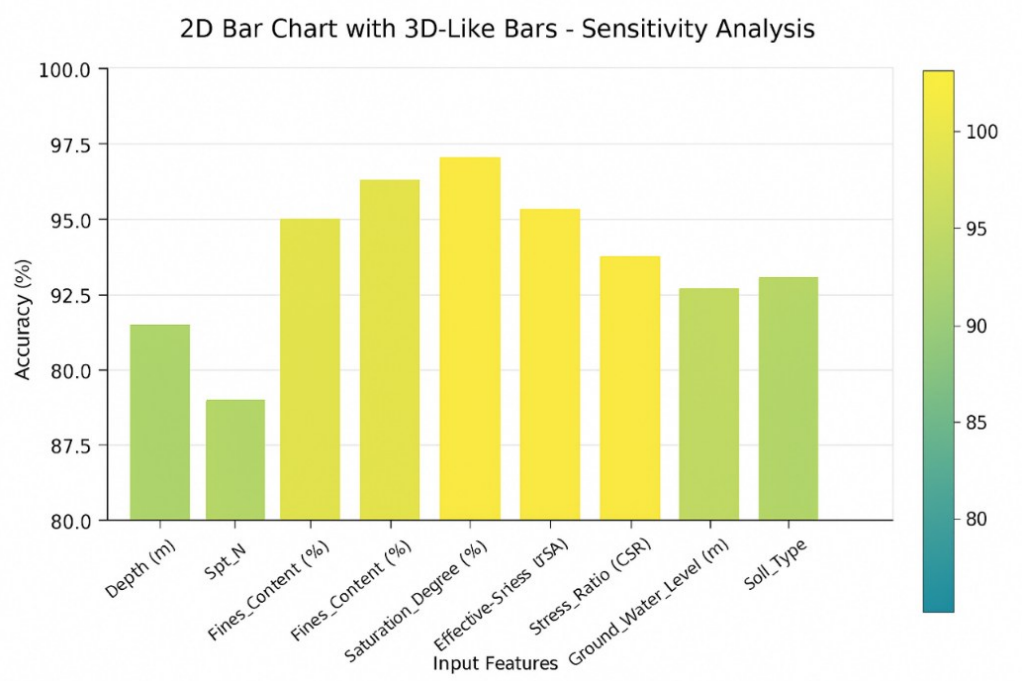

Figure 30. the sensitivity analysis outcomes for the optimized final model

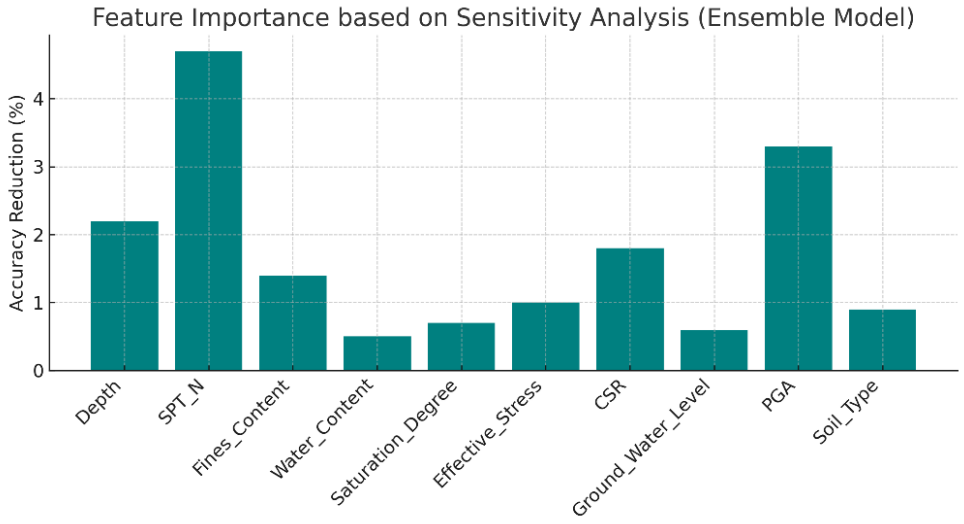

Figure 31. Feature importance results from sensitivity analysis of the Ensemble model

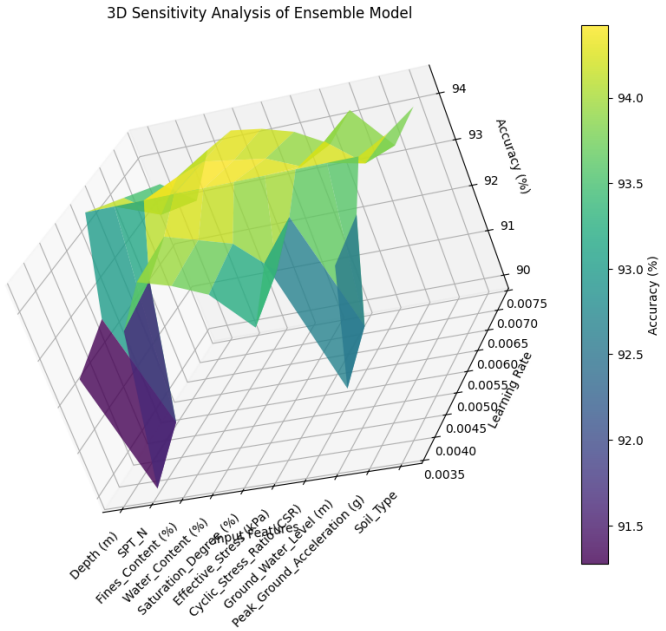

Figure 32. the3D plot sensitivity analysis outcomes for the optimized final model

10. Discussion

The results demonstrate that deep learning models, particularly the proposed metaheuristic-driven ensemble, capture the nonlinear interactions among geotechnical parameters more effectively than classical liquefaction criteria. This is consistent with recent studies showing that ML-based classifiers can model complex dependencies between SPT-N, PGA, CSR, and subsurface conditions that empirical charts oversimplify. The strong performance of the ensemble model (AUC = 0.96) indicates that combining CNN, LSTM, and Transformer branches allows the framework to simultaneously extract local feature relationships, sequential patterns, and global contextual dependencies. This synergistic feature extraction explains why the ensemble outperforms individual models.

From an engineering perspective, the model provides a reliable tool for preliminary liquefaction screening in regions with limited in-situ test coverage. The identification of SPT-N and PGA as dominant contributors offers valuable insight for seismic design, as both parameters directly relate to soil density and shaking intensity. Moreover, the model's capacity to generalize across multiple borehole sites suggests its potential use in regional-scale liquefaction susceptibility mapping and urban seismic planning. These findings support the integration of data-driven approaches into geotechnical practice, especially in areas with variable soil conditions such as Northern Iran.

The ensemble metaheuristic–deep learning framework developed in this study demonstrated superior performance compared to both traditional and standalone models. The ensemble approach achieved an accuracy of 94.5% and a correlation coefficient of 0.93, outperforming the conventional methods of Seed and Idriss (1971) [4] and Robertson and Wride (1998) [5], which are limited by their empirical nature and simplified assumptions.

The application of GWO and WOA optimizers substantially improved model generalization by adaptively tuning hyperparameters, preventing overfitting, and accelerating convergence. Similar improvements have been reported in recent studies, such as Kumar et al. (2023) [9], who utilized deep ensemble learning for liquefaction prediction. However, the present research extends beyond these by integrating multiple optimization strategies within an ensemble deep learning framework trained on a large, region-specific dataset.

The sensitivity analysis provided further interpretability, revealing that SPT_N and PGA are the dominant features influencing liquefaction susceptibility consistent with findings by Shafiei et al. (2022) [36]. In contrast, variables such as water content and saturation degree exhibited minimal influence. This insight emphasizes the importance of accurate field measurements for these critical features when designing mitigation strategies.

The results confirm that advanced hybrid models can capture the nonlinear, interdependent nature of soil and seismic parameters far better than empirical correlations. Moreover, the ensemble model's transferability to other geotechnical problems such as slope stability or settlement prediction demonstrates its broader potential for seismic risk evaluation. Future studies could expand the framework by incorporating Cone Penetration Test (CPT) data, regional seismic hazard maps, and transfer learning techniques to enhance model generalization across different geological settings.

11. Limitations and generalizability

While the proposed framework demonstrates strong predictive capability, several limitations should be acknowledged. The dataset, although consisting of 512 boreholes, is geographically concentrated in Mazandaran Province. Soil characteristics, groundwater levels, and seismic excitation patterns may differ in other regions, which could influence model performance. Although the ensemble model shows good generalization within the regional dataset, external validation using CPT-based datasets or boreholes from different seismic zones is necessary to fully assess transferability. Furthermore, the study relies on field-based liquefaction labels derived from regional investigations, which may introduce uncertainty due to variability in post-earthquake reconnaissance methods.

Despite these limitations, the model structure is fundamentally data-driven and can be retrained using site-specific data from other regions. This flexibility allows the framework to be adapted for global applications, provided that sufficient high-quality geotechnical and seismic data are available.

12. Conclusions

The results of this study demonstrate that integrating deep learning architectures with metaheuristic optimization significantly enhances liquefaction prediction accuracy compared to standalone models. While the fully connected network exhibited acceptable baseline performance, its limited capability in capturing complex nonlinear interactions among geotechnical parameters resulted in moderate predictive accuracy. The observed improvement achieved through GWO and WOA optimization indicates that hyperparameter tuning plays a crucial role in stabilizing network convergence and reducing classification uncertainty.

Models incorporating sequential and spatial learning mechanisms, such as LSTM, ConvLSTM, and Transformer, exhibited progressively improved performance. This trend highlights the importance of accounting for depth-wise dependencies and nonlinear feature interactions inherent in subsurface soil behavior. In particular, the superior performance of the Transformer model suggests that attention-based mechanisms are effective in identifying dominant patterns within heterogeneous geotechnical datasets.

The ensemble model achieved the highest predictive accuracy and correlation coefficient by combining the complementary strengths of individual models. From a geotechnical engineering perspective, this finding implies that no single modeling strategy is sufficient to fully characterize liquefaction behavior, and integrating multiple learning paradigms can substantially reduce model bias and variance. The increased computational cost of the ensemble approach is justified by its improved robustness and reliability, especially for high-risk seismic regions.

Sensitivity analysis further provides practical insights into liquefaction assessment. The dominant influence of SPT blow count and peak ground acceleration confirms their critical role in liquefaction triggering mechanisms, consistent with established geotechnical principles. Conversely, the relatively lower sensitivity associated

with water content and saturation degree suggests that, within the investigated dataset, these parameters contribute less to classification variability. These observations emphasize the necessity of prioritizing accurate field measurements of SPT and seismic loading parameters in liquefaction-prone regions.

Additionally, the observed sensitivity of model performance to learning rate, dropout ratio, and training epochs underscores the importance of systematic hyperparameter optimization. Inadequate parameter selection may lead to underfitting or unstable learning behavior, even when advanced model architectures are employed. Overall, the findings indicate that the proposed ensemble framework offers a reliable and transferable approach for liquefaction prediction, with clear implications for improving seismic risk assessment and decision-making in geotechnical engineering practice.

Funding: The authors received no financial support for this research article.

Acknowledgments: This research was conducted as part of a Ph.D. seminar focused on the numerical assessment of soil liquefaction using machine learning techniques. The analysis was carried out based on datasets collected from geotechnical companies operating in northern Iran, complemented by relevant data extracted from high-quality peer-reviewed publications. The study was undertaken by Ph.D. candidate Shima Aghakakhiri, under the supervision of Dr. Ghodratollah Mohammadi and Dr. Amir Taban, with additional academic guidance provided by Dr. Mohammad Emami Koorandeh as the thesis advisor.

Authors Contribution
Shima Aghakakhiri carried out the core components of the research, including data collection, preprocessing, numerical modeling, and preparation of the manuscript draft. The foundational idea and project concept were proposed by Dr. Mohammad Emami Koorandeh, who also led the development of the research framework, contributed substantially to the design and implementation of the machine learning models, interpreted the analytical results, and provided in-depth revisions of the manuscript. His expert guidance and technical insight were instrumental throughout all phases of the study. Dr. Ghodratollah Mohammadi and Dr. Amir Taban reviewed the final version of the manuscript and contributed general academic feedback.

Availability of data and materials
The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.Code availability: Not Applicable.

Conflict of interests
The author declares that there is no conflict of interest.

References

1. Bahrainy, H. and A. Bakhtiar, *Manjil Earthquake of June 20, 1990, The Lessons Learned*, in *Urban Design in Seismic-Prone Regions*. 2022, Springer International Publishing: Cham. p. 49-95.
DOI: https://doi.org/10.1007/978-3-031-08321-1_3

2. Uyanık, O., *Soil liquefaction analysis based on soil and earthquake parameters.* Journal of applied geophysics, 2020. 176: p. 104004.
DOI: https://doi.org/10.1016/j.jappgeo.2020.104004

3. Maurer, B.W., et al., *Evaluation of the liquefaction potential index for assessing liquefaction hazard in Christchurch, New Zealand.* Journal of Geotechnical and Geoenvironmental Engineering, 2014. 140(7): p. 04014032.
DOI: https://doi.org/10.1061/(ASCE)GT.1943-5606.0001117

4. Seed, H.B. and I.M. Idriss, *Simplified procedure for evaluating soil liquefaction potential.* Journal of the Soil Mechanics and Foundations division, 1971. 97(9): p. 1249-1273.
DOI: https://doi.org/10.1061/JSFEAQ.0001662

5. Robertson, P.K. and C. Wride, *Evaluating cyclic liquefaction potential using the cone penetration test.* Canadian geotechnical journal, 1998. 35(3): p. 442-459.
DOI: https://doi.org/10.1139/t98-017

6. Yang, Y. and E. Kavazanjian, *Newmark analysis of lateral spreading induced by liquefaction.* Journal of Earthquake Engineering, 2022. 26(6): p. 3034-3053.
DOI: https://doi.org/10.1080/13632469.2020.1784316

7. Finn, W.L., J. Dowling, and C.E. Ventura, *Evaluating liquefaction potential and lateral spreading in a probabilistic ground motion environment.* Soil Dynamics and Earthquake Engineering, 2016. 91: p. 202-208.
DOI: https://doi.org/10.1016/j.soildyn.2016.09.050

8. Kumar, D.R., P. Samui, and A. Burman, *Suitability assessment of the best liquefaction analysis procedure based on SPT data.* Multiscale and Multidisciplinary Modeling, Experiments and Design, 2023. 6(2): p. 319-329.
DOI: https://doi.org/10.1007/s41939-023-00148-x

9. Kumar, D.R., et al., *Liquefaction susceptibility using machine learning based on SPT data.* Intelligent Systems with Applications, 2023. 20: p. 200281.
DOI: https://doi.org/10.1016/j.iswa.2023.200281

10. Zhang, Y., et al., *The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity.* Bulletin of Engineering Geology and the Environment, 2021. 80(6): p. 5053-5060.
DOI: https://doi.org/10.1007/s10064-021-02250-1

11. Jas, K. and G. Dodagoudar, *Liquefaction potential assessment of soils using machine learning techniques: a state-of-the-art review from 1994–2021.* International Journal of Geomechanics, 2023. 23(7): p. 03123002.
DOI: https://doi.org/10.1061/IJGNAI.GMENG-7788

12. Ahmad, M., et al., *Application of machine learning algorithms for the evaluation of seismic soil liquefaction potential.* Frontiers of Structural and Civil Engineering, 2021. 15(2): p. 490-505.
DOI: https://doi.org/10.1007/s11709-020-0669-5

13. Geyin, M. and B.W. Maurer, *Fragility functions for liquefaction-induced ground failure.* Journal of Geotechnical

and Geoenvironmental Engineering, 2020. 146(12): p. 04020142.
DOI: https://doi.org/10.1061/(ASCE)GT.1943-5606.0002416

14. Geyin, M., *Evaluation and Development of Liquefaction Occurrence and Consequence Analytics Driven by Emerging Data and Technologies*. 2021: University of Washington.

15. Zhou, J., et al., *Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation.* Engineering with Computers, 2022. 38(Suppl 5): p. 4197-4215.
DOI: https://doi.org/10.1007/s00366-021-01418-3

16. Zhang, Y., et al., *The adoption of a support vector machine optimized by GWO to the prediction of soil liquefaction.* Environmental Earth Sciences, 2021. 80(9): p. 360.
DOI: https://doi.org/10.1007/s12665-021-09648-w

17. Bi, C., et al., *Machine learning based fast multi-layer liquefaction disaster assessment.* World Wide Web, 2019. 22: p. 1935-1950.
DOI: https://doi.org/10.1007/s11280-018-0632-8

18. Matsuoka, M., et al., *Evaluation of liquefaction potential for large areas based on geomorphologic classification.* Earthquake Spectra, 2015. 31(4): p. 2375-2395.
DOI: https://doi.org/10.1193/072313EQS211M

19. Kayen, R., et al., *Shear-wave velocity–based probabilistic and deterministic assessment of seismic soil liquefaction potential.* Journal of Geotechnical and Geoenvironmental Engineering, 2013. 139(3): p. 407-419.
DOI: https://doi.org/10.1061/(ASCE)GT.1943-5606.0000743

20. Norini, G., et al., *Assessment of liquefaction potential in the central Po plain from integrated geomorphological, stratigraphic and geotechnical analysis.* Engineering Geology, 2021. 282: p. 105997.
DOI: https://doi.org/10.1016/j.enggeo.2021.105997

21. Bhatnagar, S., A. Kranthikumar, and V.A. Sawant, *Seismic analysis of dam under different upstream water levels.* Advances in computational design, 2016. 1(3): p. 265-274.

22. Sonmezer, Y.B., *Investigation of the liquefaction potential of fiber-reinforced sand.* Geomechanics and Engineering, 2019. 18(5): p. 503-513.
DOI: https://doi.org/10.12989/gae.2019.18.5.503

23. Sonmezer, Y.B., *Energy-based evaluation of liquefaction potential of uniform sands.* Geomech. Eng, 2019. 17(2): p. 145-156.
DOI: https://doi.org/10.12989/gae.2019.17.2.145

24. Sonmezer, Y.B., et al., *Influence of grain size ratio and silt content on the liquefaction potentials of silty sands.* Geomech Eng, 2022. 31(2): p. 167-181.
DOI: https://doi.org/10.12989/gae.2022.31.2.167

25. Seed, H.B., I.M. Idriss, and I. Arango, *Evaluation of liquefaction potential using field performance data.* Journal of geotechnical engineering, 1983. 109(3): p. 458-482.
DOI: https://doi.org/10.1061/(ASCE)0733-9410(1983)109:3(458)

26. Bolton Seed, H., et al., *Influence of SPT procedures in soil liquefaction resistance evaluations.* Journal of geotechnical engineering, 1985. 111(12): p. 1425-1445.
DOI: https://doi.org/10.1061/(ASCE)0733-9410(1985)111:12(1425)

27. Bray, J.D. and R.B. Sancio, *Assessment of the liquefaction susceptibility of fine-grained soils.* Journal of geotechnical and geoenvironmental engineering, 2006. 132(9): p. 1165-1177.
DOI: https://doi.org/10.1061/(ASCE)1090-0241(2006)132:9(1165)

28. Boulanger, R. and I. Idriss. *Evaluating cyclic failure in silts and clays*. in *Proceedings, geotechnical earthquake engineering satellite conference on performance based design in earthquake geotechnical engineering: concepts and research. Prepared by TC4 Committee of ICSMGE, Japanese Geotechnical Society, Tokyo*. 2005.

29. Kumar, D., et al., *A novel methodology to classify soil liquefaction using deep learning.* Geotechnical and Geological Engineering, 2021. 39: p. 1049-1058.
DOI: https://doi.org/10.1007/s10706-020-01544-7

30. Li, L., et al., *Soil seismic response modeling of KiK-net downhole array sites with CNN and LSTM networks.* Engineering Applications of Artificial Intelligence, 2023. 121: p. 105990.
DOI: https://doi.org/10.1016/j.engappai.2023.105990

31. Bačanin Džakula, N. *Convolutional neural network layers and architectures*. in *Sinteza 2019-International Scientific Conference on Information Technology and Data Related Research*. 2019. Singidunum University.
DOI: https://doi.org/10.15308/Sinteza-2019-445-451

32. Purwono, P., et al., *Understanding of convolutional neural network (cnn): A review.* International Journal of Robotics and Control Systems, 2022. 2(4): p. 739-748.
DOI: https://doi.org/10.31763/ijrcs.v2i4.888

33. Wei, G., et al., *Development of a LeNet-5 gas identification CNN structure for electronic noses.* Sensors, 2019. 19(1): p. 217.
DOI: https://doi.org/10.3390/s19010217

34. Mirjalili, S., S.M. Mirjalili, and A. Hatamlou, *Multi-verse optimizer: a nature-inspired algorithm for global optimization.* Neural computing and applications, 2016. 27(2): p. 495-513.
DOI: https://doi.org/10.1007/s00521-015-1870-7

35. Sharma, A., et al., *A novel TSA-PSO based hybrid algorithm for GMPP tracking under partial shading conditions.* Energies, 2022. 15(9): p. 3164.
DOI: https://doi.org/10.1007/s00521-015-1870-7

36. Shafiei, P., M. Azadi, and M.S. Razzaghi, *A novel liquefaction prediction framework for seismically-excited tunnel lining.* Earthquakes and Structures, 2022. 22(4): p. 401-419.
DOI: https://doi.org/10.12989/eas.2022.22.4.401

37. Rothman, L.S., et al., *The HITRAN database: 1986 edition.* Applied optics, 1987. 26(19): p. 4058-4097.
DOI: https://doi.org/10.1364/AO.26.004058

38. Emami, M. and S.S. Yasrobi, *Modeling and Interpretation of Pressuremeter Test Results with Artificial Neural Networks.* Geotechnical and Geological Engineering, 2014. 32(2): p. 375-389.
DOI: https://doi.org/10.1007/s10706-013-9720-9