Journal of Rock Mechanics and Geotechnical Engineering 18 (2026) 3966–3990

Contents lists available at ScienceDirect

Journal of Rock Mechanics and Geotechnical Engineering

journal homepage: www.jrmge.cn

Full Length Article

An automatic parameterization tool for the hardening soil model

Kyrillos Ebrahim [a,b,*], Tarek Zayed [a], Ridwan Taiwo [a,**], Ashraf El-Hamalawi [c]

[a] Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, China
[b] Structural Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt
[c] School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, LE11 3TU, UK

ARTICLE INFO

Article history:
Received 5 February 2025
Received in revised form
9 June 2025
Accepted 10 July 2025
Available online 12 December 2025

Keywords:
Automated parametrization
Hardening soil finite element model
Deep learning
Long short-term memory (LSTM)

ABSTRACT

This research introduces a powerful tool, the automatic parametrization of hardening soil (HS) model (APHS), designed to make the HS model parameterization process easier and faster than conventional methods while maintaining high accuracy. Traditional parameterizations rely on oedometer tests, unloading-reloading data, or domain-specific assumptions. Existing optimization-based models often assume uniform parameter weighting, potentially overlooking the distinct sensitivity of each parameter. APHS addresses these limitations as a standalone tool that relies exclusively on conventional triaxial loading test data. To achieve this goal and address the scarcity of labeled datasets, this study integrates numerical modeling with deep learning. The study focuses on a typical shallow Hong Kong soil with parameter ranges derived from field data and relevant literature. Latin hypercube sampling generated diverse parameter values within theoretical bounds for reliable input, while a two-dimensional (2D) axisymmetric finite element model (SIGMA/W) simulated laboratory tests to create a comprehensive, labeled dataset. Seven novel multi-parallel deep long short-term memory (LSTM) networks were trained and validated, achieving an accuracy of 99.4 %. Validation against a conventionally parameterized reference case confirmed 99.6 % accuracy, while an experimental laboratory case study demonstrated strong agreement between simulated and measured results. APHS accelerates HS model parameterization, delivering accurate results in seconds. It can seamlessly integrate with finite element models for automated laboratory data processing and physically informed models to refine calibration parameter ranges. Future work will expand its applicability to various conditions and parameters.

1. Introduction

The preference for numerical analysis over conventional approaches arises from various factors. One significant advantage is the ability to obtain more detailed information, especially when dealing with geotechnical engineering issues (Brinkgreve, 2019). Several factors influence the performance of numerical studies, including the accurate calculation of constitutive model parameters. Soil constitutive models have evolved significantly over time. Advanced models and accurate input parameters can reflect soil behavior better than simpler ones (Marzouk et al., 2024). Traditionally, these parameters are derived from laboratory tests, such as triaxial and oedometer tests, which demand significant domain expertise for accurate parametrization. As advancements in deep learning (DL) continue to transform the geotechnical field (Nguyen-Minh et al., 2025; Xu et al., 2025a), the development of automated tools that can streamline this process, delivering higher accuracy in a shorter time, has become increasingly critical.

Automated tools for parameterizing soil constitutive models are relatively rare in the existing literature. For instance, Marzouk et al. (2024) introduced an automated parameter determination (APD) system designed for in situ tests, proving valuable when only limited soil data are available before conducting comprehensive laboratory testing. Similarly, Machaček et al. (2022) developed an automatic parameter calibration (AC) method tailored to hypoplastic soil models. Brosz et al. (2024) proposed a monotonic and cyclic loading conditions calibration approach,

* Corresponding author. Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hung Hom, China.
** Corresponding author.
E-mail addresses: kyrillos.ebrahim@connect.polyu.hk (K. Ebrahim), ridwan-a.taiwo@connect.polyu.hk (R. Taiwo).
Peer review under responsibility of Institute of Rock and Soil Mechanics, Chinese Academy of Sciences.

https://doi.org/10.1016/j.jrmge.2025.07.041

focusing on hypoplasticity with intergranular strain and SANISAND models. Liu et al. (2024) contributed by developing an automatic parameter calibration method using the critical state-based clay and sand model (CASM) as a case study. These methods rely on optimization algorithms, including the Nelder–Mead method (Nelder and Mead, 1965), differential evolution algorithms (Storn and Price, 1997), genetic algorithms, and particle swarm optimization algorithms (Shi and Eberhart, 1998), to identify optimal parameter values. Despite the advancements in these methodologies, the integration of advanced DL algorithms remains rare. Furthermore, these methodologies often use a loss function (which could be physically informed (Li et al., 2024)) with uniform weights across all target parameters, focusing on minimizing errors between simulated and laboratory results. This approach can neglect each parameter's unique sensitivity and contribution to the model's accuracy. Therefore, there is a need for a fully automated tool that leverages DL capabilities, accounts for varying loss weights, and validates its accuracy against known parameters, ensuring independence from numerical errors.

Developing an automated and accurate tool capable of accounting for a wide range of boundary conditions (such as saturation level, normal and over-consolidation states, varying parameter ranges, etc.) requires extensive laboratory datasets, which are often not readily available. To address this limitation, two key techniques are proposed: (1) to focus on a specific soil type and a set of conditions where the tool can be reliably applied to provide accurate results; and (2) to develop a finite element model to generate a large synthetic dataset that can replicate laboratory test results (i.e. data augmentation). The tool can generate accurate soil model parameters by employing these strategies while minimizing uncertainties related to insufficient experimental data.

With that in mind, the primary aim of this study is to automate the hardening soil (HS) model parameterization, acknowledging the temporal dependencies inherent in the stress–strain relationship. To address this, DL is employed, specifically long short-term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997), which have proven effective in modeling time-series problems (Xing et al., 2019; Nguyen et al., 2022) and soil stress-strain behavior, as demonstrated by Zhang et al. (2020, 2021a) and Li et al. (2024). LSTM is particularly effective in capturing complex temporal dependencies and automatically extracting relevant features from data (Liu et al., 2025). Unlike traditional methods, DL does not require prior hypotheses or domain-specific knowledge, allowing it to model intricate mechanisms efficiently (Nguyen-Minh et al., 2023). Additionally, it can handle large datasets, eliminating the need for manual feature engineering while maintaining high predictive accuracy (Schmidhuber, 2015).

However, the application of DL in geotechnical parameterization is often hindered by the scarcity of real-world laboratory datasets, which are costly and time-intensive. To address this limitation, various techniques such as data augmentation (Shorten and Khoshgoftaar, 2019), physics-informed neural networks (PINNs) (Li et al., 2024; Wong et al., 2025), active learning (Di Fiore et al., 2024), finite element integrated neural network (FEINN) (Xu et al., 2025b; Zhang et al., 2025), and transfer learning (Pan and Yang, 2010) have been applied. As employed in this study, data augmentation offers a robust solution by generating synthetic datasets replicating laboratory test results. This approach facilitates reliable model training and validation by creating comprehensive datasets with labeled inputs and outputs, effectively minimizing numerical bias and ensuring accuracy.

Given the foregoing analysis, this study integrates field data with literature insights to define a predetermined parameter range. Latin hypercube sampling (LHS) systematically varies parameters within the defined range, with theoretical bounds for reliable and comprehensive input space coverage (McKay et al., 1979). Subsequently, a finite element SIGMA/W model (GeoStudio, 2023) was developed to simulate the soil behavior within the predefined boundary conditions, as detailed in Section 3. The stress–strain relationships extracted from these simulations provided a rich dataset for training and validating the LSTM networks. Finally, the proposed model's performance was validated with actual laboratory data. This approach ensures that the automated tool leverages the strengths of augmented datasets while remaining grounded in real-world geotechnical scenarios.

This user-friendly tool is designed to automatically determine seven critical parameters for the HS model using data derived from conventional consolidated drained triaxial tests. These include the strength parameters (i.e. effective cohesion c', effective internal friction angle ϕ', the failure ratio R_f) and stiffness parameters (the tangent stiffness for primary oedometer loading E_{oed}^{ref}, the secant stiffness in drained triaxial tests E_{50}^{ref}, the unloading/reloading stiffness E_{ur}^{ref}, and the power for the stress level dependency of stiffness m). It is important to emphasize that this work aims to demonstrate the potential of an automated tool that simplifies and accelerates the process while maintaining high precision and accuracy. However, the selection of models and boundary conditions is guided by the specific requirements of the engineering problem. Future research will expand the tool's capabilities to include additional boundary conditions and a broader range of strength and stiffness parameters.

The structure of this paper is organized as follows. Section 2 presents an overview of the research background, providing the necessary context for the study. Section 3 presents the methodology, emphasizing the parameterization workflow, boundary conditions, modeling approaches, and analysis techniques. Section 4 discusses the results in detail, offering an in-depth analysis of their implications and proposing potential directions for future research. Finally, Section 5 summarizes the key findings and emphasizes their practical significance.

2. Research background

This section establishes the foundation for the study by first introducing the practical applicability of the selected case study. It then discusses the HS model, emphasizing its suitability for capturing the complex stress–strain relationships of soils. The discussion then transitions to applying the LSTM model as a cutting-edge approach to automating parametrization, leveraging its ability to handle temporal dependencies and complex datasets.

Ensuring practical engineering applicability, this study leverages Hong Kong's unique geological conditions, in which the presence of colluvium and completely decomposed granitic, volcanic, and sedimentary rocks (a shallow layer with thickness ranges from 1 m to 36 m) is a significant problem related to slope instability, especially with its topographic nature where 60 % of its land is natural slopes (GCO, 1982; Ng and Shi, 1998; Ng et al., 2001; Maunsell Geotechnical Services Ltd., 2007). The urban expansion further complicates the situation as developments increasingly encroach on slopes and natural terrain (Lam, 2018), necessitating accurate geotechnical modeling in this densely urbanized region (Chan et al., 2021) and underscoring the need for tools that simplify and enhance the accuracy of this process.

Since numerous soil constitutive models exist, each catering to specific geotechnical applications, selecting an appropriate model depends on the nature of the project and the conditions to be simulated (Ng et al., 2015; Li and Yin, 2025). Common models include the Mohr–Coulomb model, which is widely used for its

simplicity but has limitations in accurately capturing complex stress–strain behaviors (Dong et al., 2016). Advanced models such as the HS model, which is preferred for their ability to simulate nonlinear elastic and plastic behaviors under both compression and shear conditions, make them ideal for a wide range of soils and loading scenarios (Schanz et al., 1999). The HS model has proven effective in applications such as deep excavations, slope stability, slope nailing, and tunneling (Surarak, 2010; Surarak et al., 2012; Truty and Obrzud, 2015; Lu et al., 2018; Bangun et al., 2020; Chan et al., 2021; Bhadiyadra and Ong, 2024; Bui et al., 2024). Its versatility and relevance to Hong Kong's geotechnical challenges make it the optimal choice for this study (Chan et al., 2021). The HS model is used as an example to evaluate the proposed method's efficiency while acknowledging that the selection of models should align with the specific requirements of the engineering problem and soil type at hand.

2.1. The HS model

The HS model is an advanced constitutive model designed to represent the nonlinear stress–strain behavior of soils (Schanz et al., 1999). Unlike linear elastic models, the HS model incorporates stress-dependent stiffness, considering variations in stiffness for primary loading, unloading/reloading, and oedometer conditions. Parameters such as E_{50}^{ref}, $E_{\text{oed}}^{\text{ref}}$, and $E_{\text{ur}}^{\text{ref}}$ are pivotal in capturing soil behavior under different stress paths. This model also integrates critical concepts such as R_{f} and m, making it particularly suitable for simulating complex soil–structure interactions. Table 1 describes the HS model parameters. The formulation details are comprehensively outlined in the reference book by (GEOSLOPE International Ltd., 2020). A concise summary of the HS model's formulation is provided below.

The stress–strain behavior observed during triaxial testing, as illustrated in Fig. 1, highlights that the HS model formulation integrates the Mohr–Coulomb failure criterion, leading the stress path (q,p') to intersect with the Mohr–Coulomb failure surface. The stress–strain relationship, represented as (q, ε_1), follows a hyperbolic curve. This curve's behavior is influenced by stress-dependent stiffness parameters, namely the initial tangent stiffness (E_{i}), E_{50}, and E_{ur}, which are crucial in defining its characteristics.

As for the stress–strain behavior during 1D compression under oedometer loading, as visualized in Fig. 2, the cap yield surface appears as an elliptical shape in the deviatoric stress vs. mean stress ($q - p'$) space, with a specific aspect ratio (M) defining its geometry. During 1D (oedometer) compression, the stress path intersects the cap yield surface, causing it to expand as the soil transitions from elastic to elastoplastic behavior. E_{oed} characterizes the stress–strain relationship in oedometer tests. This stiffness is stress-dependent, varying with the sample's stress state.

2.2. The DL model selection and development

The relationship between strain and stress can be effectively modeled as a time-series problem because these variables evolve

Table 1
Parameters of the HS model.

Parameter	Description
Effective internal friction angle (ϕ')	The slope of the failure envelope in the Mohr–Coulomb failure criterion
Effective cohesion (c')	The shear strength intercept of the failure envelope
Failure ratio (R_{f})	The ratio of mobilized deviatoric stress to ultimate stress at failure
Dilatancy angle (ψ)	The variable governing the volumetric expansion or contraction during shearing
Secant stiffness (E_{50}^{ref})	Secant stiffness at 50 % of the peak deviatoric stress under drained triaxial conditions
Tangent stiffness ($E_{\text{oed}}^{\text{ref}}$)	Tangent stiffness derived from one-dimensional (1D) compression in an oedometer test
Unloading/reloading stiffness ($E_{\text{ur}}^{\text{ref}}$)	The parameter reflecting the soil's stiffness during unloading/reloading cycles
Exponential power (m)	The parameter describing how stiffness depends on confining pressure
Poisson's ratio (ν_{ur})	The ratio of lateral to axial strain under unloading and reloading conditions
Coefficient of earth pressure at rest (K_0^{NC})	The parameter representing the lateral earth pressure in normally consolidated soils

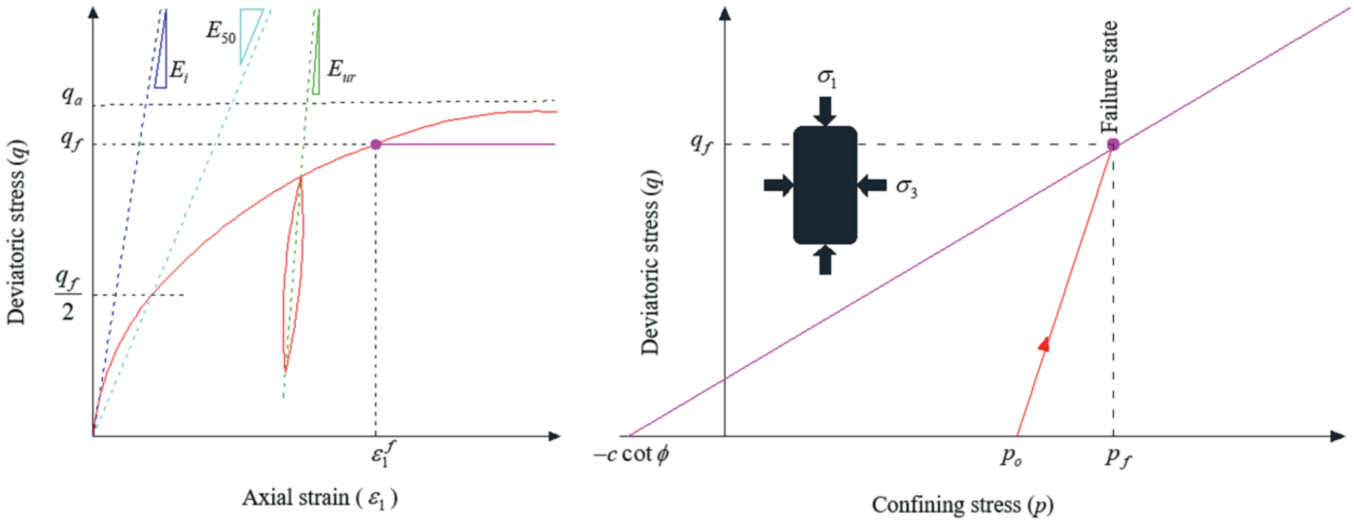


Fig. 1. Stress–strain behavior observed during triaxial testing.

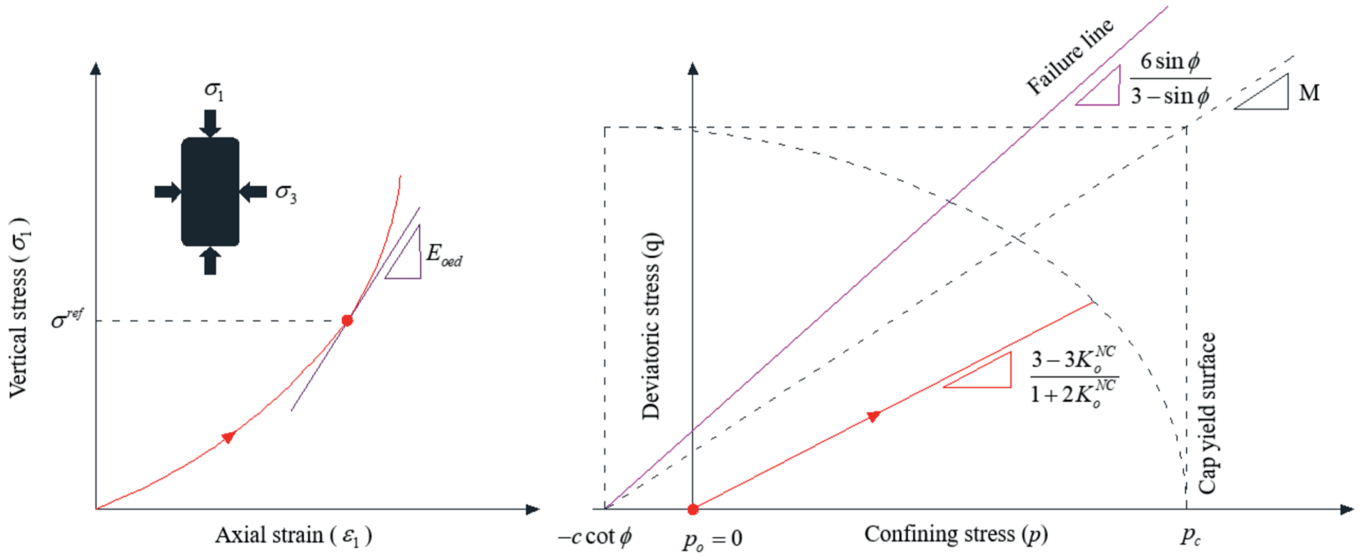


Fig. 2. Stress–strain behavior during 1D compression under oedometer loading.

sequentially under applied loads, reflecting dynamic and dependent interactions over time. As external forces are applied, the stress–strain behavior changes progressively, influencing volumetric strain in a temporally correlated manner, making this problem a time-series application.

Time-series applications rely on inputs and outputs that vary over time, necessitating models capable of capturing temporal relationships. Static models such as linear regression or basic neural networks often fail to establish these connections in many cases. DL techniques, particularly recurrent neural networks (RNNs), are better suited for such tasks. However, RNNs are hindered by vanishing and exploding gradient issues. While gradient clipping can manage exploding gradients, vanishing gradients remain a challenge. Advanced models such as gated recurrent units (GRUs) and LSTM networks address these limitations (Hochreiter and Schmidhuber, 1997).

Relevant studies on geotechnical time-series applications reveal that LSTM outperforms RNN, GRU, and other architectures (Xing et al., 2019; Liu et al., 2020; Varangaonkar and Rode, 2023; Wang et al., 2023). Recent studies have highlighted the superiority of LSTM in various applications. Zhang et al. (2021a) demonstrated that LSTM outperforms feedforward and feedback neural networks in modeling soil stress–strain behavior. Additionally, Zhang et al. (2020) found that LSTM is effective in modeling the cyclic behavior of granular soils. Li et al. (2024) further showed that the PINN (LSTM) model excels in simulating the hydromechanical behavior of unsaturated expansive soils, making the LSTM the preferred model for this study.

Recent findings (Murray et al., 2022, 2023) indicate that while models such as transformer networks, LSTM, and convolutional neural networks (CNNs) each possess distinct strengths, their performance differences are often marginal. At the same time, the no free lunch theorem (Wolpert, 1996) reminds us that no single algorithm is universally optimal. DL remains an empirical process that requires careful optimization of model architecture, input features, loss functions, training strategies, and more. Therefore, although our current focus on LSTM is supported by literature, exploring alternative architectures in future work may further enhance predictive performance while reducing computational burdens.

The LSTM model, developed by Hochreiter and Schmidhuber (1997), is an advanced type of RNN designed to overcome vanishing gradient issues in long-term time-series prediction. Unlike traditional RNNs, LSTMs use memory blocks comprising memory cells and three gates: input ($\boldsymbol{i}_t$), output ($\boldsymbol{O}_t$), and forget ($\boldsymbol{f}_t$) gates. These gates control data flow and retention (Eqs. (1) and (2)). Each LSTM cell processes three inputs: current data ($\boldsymbol{x}_t$), previous output ($\boldsymbol{h}_{t-1}$), and prior memory state ($\boldsymbol{C}_{t-1}$), as shown in Fig. 3a and

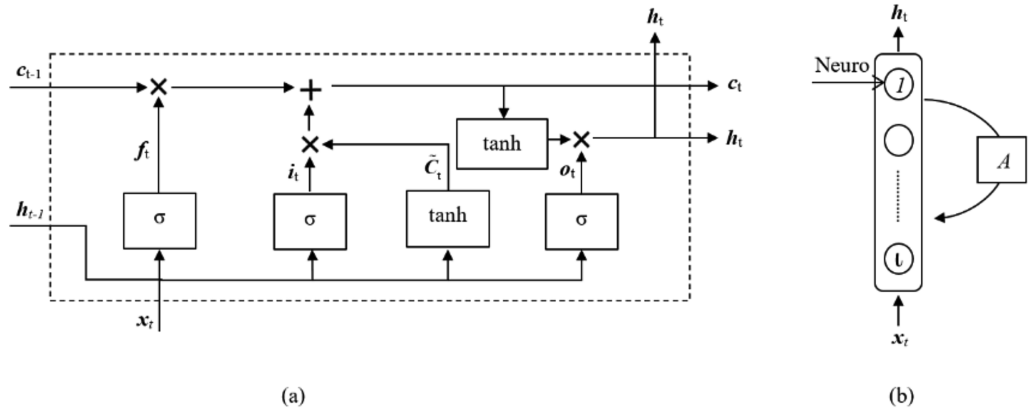


Fig. 3. LSTM network model structure: (a) LSTM cell structure; and (b) LSTM simplified visualization.

b, enabling robust temporal data handling through specialized gating mechanisms.

$$\left.\begin{aligned} \boldsymbol{f}_{\mathrm{t}} &= \sigma\left(\boldsymbol{W}_f\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_f\right)\\ \boldsymbol{i}_t &= \sigma(\boldsymbol{W}_i\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_i)\\ \boldsymbol{o}_t &= \sigma(\boldsymbol{W}_o\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_o) \end{aligned}\right\} \tag{1}$$

where $[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]$ denotes the concatenation of two vectors: the previous hidden state $\boldsymbol{h}_{t-1}$ and the current input $\boldsymbol{x}_t$; σ represents the activation function; and $\boldsymbol{W}$ and $\boldsymbol{b}$ denote the weight matrix and bias that govern the behavior of the gates, respectively. Once the output information is computed, Eq. (2) can derive both the short-term memory state ($\boldsymbol{h}_t$) and the long-term memory state ($\boldsymbol{C}_t$):

$$\left.\begin{aligned} \boldsymbol{C}_t &= \boldsymbol{f}_t\odot\boldsymbol{C}_{t-1}+\boldsymbol{i}_t\odot\tilde{\boldsymbol{C}}_t\\ \tilde{\boldsymbol{C}}_t &= \tanh(\boldsymbol{W}_{\mathrm{c}}\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\mathrm{c}})\\ \boldsymbol{h}_t &= \boldsymbol{O}_t\odot\tanh\boldsymbol{C}_t \end{aligned}\right\} \tag{2}$$

where $\odot$ denotes element-wise multiplication (the Hadamard product).

3. Research methodology

This section outlines the methodological framework for parameterizing the HS model. It begins by describing the traditional conceptual parameterization process. The study then introduces the automated workflow, which leverages data augmentation, finite element modeling, and DL. Finally, the boundary conditions and soil parameter ranges used for model training and validation are defined.

3.1. Traditional conceptual parameterization workflow

The traditional conceptual parameterization workflow, provided by GEOSLOPE International Ltd., (2021), which is an established methodology, consists of three main steps: (1) the experimental approach, (2) the numerical approach, and (3) the result verification. In the experimental approach, the soil sample is prepared, triaxial tests are performed, and stress–strain data are collected (Fig. 4a–c). Following that step, manual parameterization begins. The numerical approach (i.e. manual parametrization) involves parametrization of model constraints, including: (1) strength parameters (c', ϕ', and R_{f}) (Fig. 4d and e) and (2) stiffness parameters ($E_{\mathrm{oed}}^{\mathrm{ref}}, E_{50}^{\mathrm{ref}}, E_{\mathrm{ur}}^{\mathrm{ref}}, K_0^{\mathrm{NC}}$, and m) (Fig. 4f–j). Since this study emphasizes the parameterization process, a concise overview of how these parameters are derived manually is provided in the following sections. For a detailed explanation, readers are encouraged to refer to GEOSLOPE International Ltd., (2021) for comprehensive insights:

(1) Concerning the strength parameters, in Fig. 4d, the c' is derived from the shear-normal stress space as the intersection of the failure line with the vertical axis, and the ϕ' corresponds to the slope of the failure line. In Fig. 4e, the R_{f} is calculated based on the slope of parallel lines in the relationship between $\varepsilon_1 q_{\mathrm{f}}/q$ and axial strain ε_1.
(2) For stiffness parameters, the reference stress (σ^{ref}) is used to determine the failure deviatoric stress ($q_{\mathrm{f}}^{\mathrm{ref}}$) following the Mohr–Coulomb strength parameters (Fig. 4g). The E_{50}^{ref} is calculated as the slope of a line passing through the point corresponding to 50 % of the failure deviatoric stress derived from a specific triaxial test (Fig. 4f). The $E_{\mathrm{ur}}^{\mathrm{ref}}$ is determined using the least squares method, as depicted in Fig. 4f. The $E_{\mathrm{oed}}^{\mathrm{ref}}$ is defined as the tangent stiffness at the point where vertical stress equals the reference stress, as shown in Fig. 4h. $E_{\mathrm{oed}}^{\mathrm{ref}}$ requires results from an oedometer test but can be approximated as a ratio of the half-failure reference stiffness, typically ranging between 50 % and 80 %, depending on the soil type (GEOSLOPE International Ltd., 2020). Similarly, the coefficient of lateral earth pressure, defined as the ratio of lateral stress (σ_3) to vertical stress (σ_1), can be computed by the solver ($K_0^{\mathrm{NC}} = 1 - \sin\phi'$) (Fig. 4i). The m exponent is determined as the slope of a trend line derived from the relationship between $\ln\left(E_{50}/E_{50}^{\mathrm{ref}}\right)$. Typically, this exponent is assumed to remain consistent for both loading and unloading-reloading phases (Fig. 4j).

3.2. Automated conceptual parameterization workflow

Fig. 5 illustrates the automated conceptual parameterization workflow. The proposed tool streamlines the traditional parameterization process shown in Fig. 4d–j by automatically processing laboratory data (Fig. 4a–c) using the user-friendly automatic parametrization of the HS model (APHS) tool. This tool integrates a novel multi-LSTM model focusing exclusively on triaxial test data. In contrast to conventional methods, which depend on oedometer tests or empirical correlations for determining parameters like $E_{\mathrm{oed}}^{\mathrm{ref}}$, and require unloading-reloading stress–strain data to estimate $E_{\mathrm{ur}}^{\mathrm{ref}}$, this approach capitalizes on the proposed multi-LSTM's ability to detect intricate patterns solely from triaxial loading test data.

3.3. Boundary conditions and soil parameter ranges

This study aims to ensure the tool's practical applicability and accurate predictions by establishing an appropriate parameter range. This approach is essential for optimizing the dataset size and ensuring adequate search space coverage, which is critical given that even minor prediction errors can lead to significant simulation deviations. With this in mind, Table 2 summarizes the parameter ranges applied in this study, including strength parameters such as c', ϕ', R_{f}, and ψ, and stiffness parameters such as $E_{\mathrm{oed}}^{\mathrm{ref}}$, E_{50}^{ref}, $E_{\mathrm{ur}}^{\mathrm{ref}}$, m, and ν_{ur}, which are determined as follows.

Field data, laboratory tests, and relevant literature defined the boundaries for strength parameters (c', ϕ', and ψ). In May 2023, field sampling was conducted at multiple locations across Hong Kong to determine the basic soil properties of the shallow layers. A total of 33 samples were collected from 11 different locations. Laboratory tests conducted by ETS-Testconsult Ltd., (2023) indicate that the samples primarily consist of moist, brown, silty clayey gravelly sand. These samples were subsequently tested using small shear box tests under normal stresses of 20 kPa, 30 kPa, and 50 kPa to determine their cohesion and internal friction angle. The median values were 3.67 kPa for effective cohesion and 30.5° for the internal friction angle, as shown in Fig. 6. The range of these parameters is consistent with findings from the literature. Maunsell Geotechnical Services Ltd. (2007) reported cohesion values ranging from 0 kPa to 8 kPa and internal friction angles between 31° and 36° for shallow colluvium and completely decomposed tuff layers. Zhou (2008) noted that typical completely decomposed granite (CDG) soils in Hong Kong exhibit an effective cohesion of approximately 0 kPa and an internal friction angle of around 35°. Additionally, Wong (2020) recommended values of 5 kPa for cohesion and 36° for the internal friction angle for fill/colluvium soils based on extensive testing. Observations from the small shear box tests indicate that dilation can reasonably be

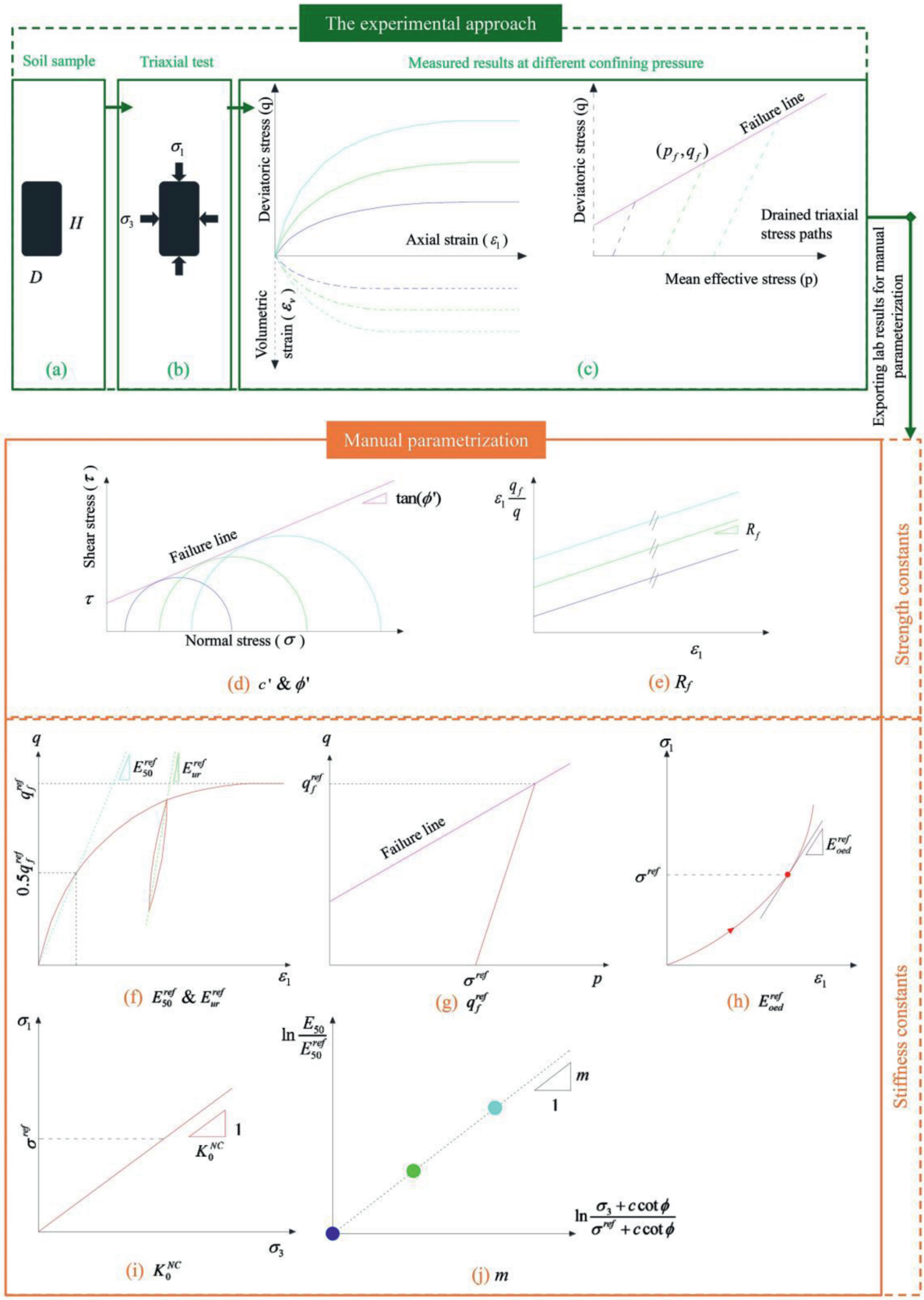


Fig. 4. Traditional parameterization workflow: (a–c) The experimental approach; (d, e) Parameterization of strength constants; and (f–j) Parameterization of stiffness constants (Section 3.1).

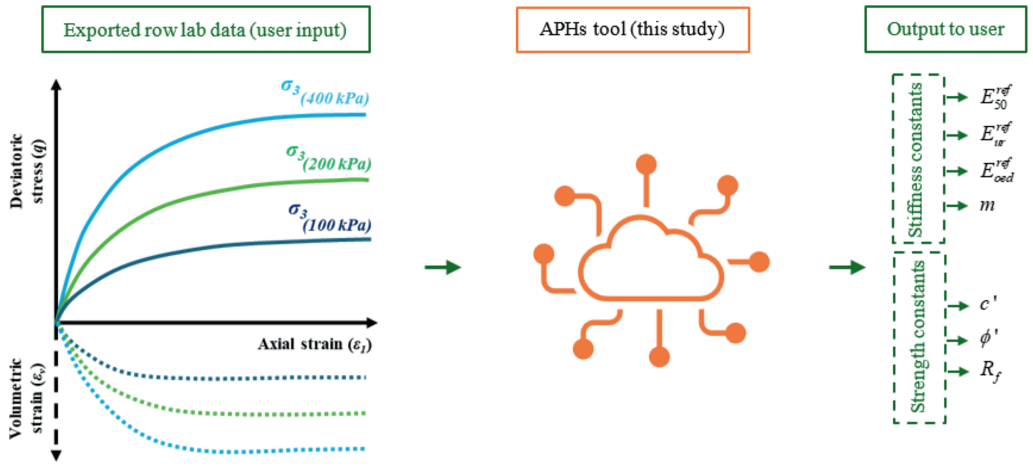


Fig. 5. The workflow of the proposed APHS tool (Section 3.2).

Table 2
HS model parameter range for model training and validation (this study).

Parameter	Minimum value	Maximum value
Over-consolidation ratio (OCR)	1	
K_o^{NC}	Calculated by the solver $(1-\sin\phi)$	
ν_{ur}	0.15	0.35
E_{ur}^{ref} (kPa)	10,000	50,000
E_{50}^{ref} (kPa)	2000	$0.5E_{ur}^{ref}$
E_{oed}^{ref} (kPa)	$0.5E_{50}^{ref}$	$0.8E_{50}^{ref}$
m	0.15	0.99
σ^{ref} (kPa)	100	
c' (kPa)	0	8
ϕ' (°)	29	36
ψ (°)	0	
R_f	0.5	1

assumed to be negligible, consistent with Zhou (2008)'s findings, which demonstrated that dilation in saturated CDG soils in Hong Kong is insignificant.

Concerning stiffness parameters (i.e. E_{oed}^{ref}, E_{50}^{ref}, E_{ur}^{ref}, and ν_{ur}), relevant literature guided the establishment of the parameter boundaries (refer to Table 2). Local studies, such as those by Lee et al. (2008), Lam (2018), and Lu et al. (2018), reported typical ranges for soil parameters suitable for modeling. Specifically, they found that $E_{oed}^{ref} = 0.7E_{50}^{ref}$ to E_{50}^{ref}, while $E_{ur}^{ref} = 2E_{50}^{ref}$ to $3E_{50}^{ref}$. For example, Lu et al. (2018) reported values of 20.78 MPa for E_{ur}^{ref}, 7.573 MPa for E_{oed}^{ref}, and 10.39 MPa for E_{50}^{ref} for a CDG layer, with ν_{ur} of 0.2 and $m = 0.517$. Lam (2018) considered values of 10.5 MPa, 45 MPa, and 15 MPa for E_{oed}^{ref}, E_{ur}^{ref}, E_{50}^{ref}, respectively, while Chan et al. (2021) used $E_{oed}^{ref} = E_{50}^{ref} = 5$ MPa and $E_{ur}^{ref} = 15$ MPa for a fill layer.

The minimum and maximum values listed in Table 2 were determined to ensure reasonable coverage of the parameter values discussed above. For c', the range of 0–8 kPa was adopted. Similarly, for ϕ', since all literature suggests values greater than 30°, which aligns with our laboratory tests, lower values (i.e. below 28°) were not considered. This provides a comprehensive range between 29° and 36°. E_{ur}^{ref} was chosen to cover a broader range (10–50 MPa) than typically reported ones. For parameters that lack a clearly identified range, such as R_f, m, and ν_{ur}, a wider range has been adopted to consider all possible scenarios. Most importantly, to ensure that the generated search space aligns with the theoretical assumptions, the maximum values of E_{50}^{ref} are capped to $0.5E_{ur}^{ref}$. Similarly, E_{oed}^{ref} is assumed to fall within the range of $0.5E_{50}^{ref}$ to $0.8E_{50}^{ref}$. Additionally, each parameter set is thoroughly reviewed for reliability, as described in Section 4.1. These soils are assumed to be fully saturated, normally consolidated, and non-dilative (Ng et al., 2001; Zhou, 2008). These conditions establish a solid framework, with future work expanding it to a broader range of scenarios.

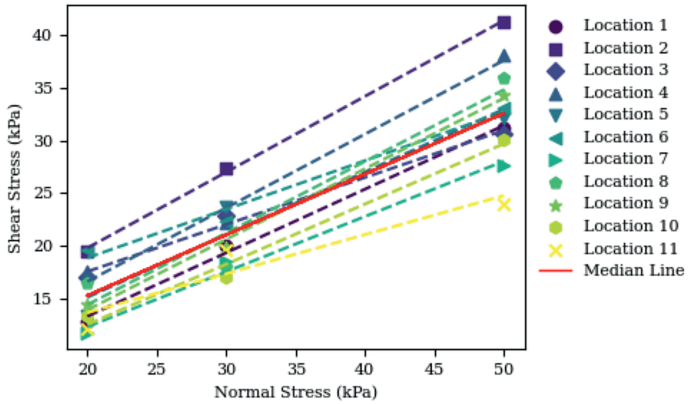


Fig. 6. Shear strength vs. normal stress relationship.

3.4. APHS model development

3.4.1. LHS

Modeling all scenarios within the parameter range specified in Section 3.3 (Table 2) requires an extensive number of samples, making manual sampling impractical. To address this, the study utilizes LHS (McKay et al., 1979), a statistical method that ensures efficient and comprehensive parameter space coverage. LHS is a stratified sampling technique that ensures that each parameter's range is divided into (n) intervals of equal probability, with one sample taken from each interval. For a given parameter (P_i) within the range of $[P_{i,\min}, P_{i,\max}]$, the sampling values are defined, as illustrated in Eq. (3). This process is repeated for all parameters, ensuring an evenly distributed and non-overlapping sampling across the multi-dimensional parameter space. LHS achieves efficient coverage, reducing the number of simulations needed compared to simple random and Monte Carlo (MC) sampling methods (McKay et al., 1979; Wang, 2022).

$$P_i^j = P_{i,\min} + (P_{i,\max} - P_{i,\min})\frac{j - r_i}{n} \quad (j \in \{1, 2, \ldots, n\}) \tag{3}$$

where j is the sample index; r_i is a random value in [0,1), ensuring uniform randomness within each interval; and n is the number of samples.

3.4.2. Numerical model for triaxial tests

The samples generated using LHS within the parameter ranges specified in Table 2 were simulated using an axisymmetric finite element model in SIGMA/W (GeoStudio, 2023). Key considerations include: (1) modeling strain up to 40 % to cover a wide range of behavior; (2) ensuring that all parameter sets reached apparent failure before 40 % strain, and the sets not meeting this condition were adjusted manually; (3) conducting triaxial simulations under confining pressures (σ_3) of 100 kPa, 200 kPa, and 400 kPa; and (4) extracting q and volumetric strain (ε_v) at ε_1 increments of 0.4 %. Extracting ε_1 in small increments helps to represent each curve with enough points, minimizing the impact of any potentially noisy data. Fig. 7 illustrates the model's boundary conditions and dimensions to simulate a conventional drained triaxial test. The model is built as an axisymmetric two-dimensional (2D) simulation (in an r-z coordinate) that leverages the assumption of rotational symmetry about the axis. This allows it to capture the essential three-dimensional (3D) behavior without the full computational cost of a complete 3D simulation.

3.4.3. LSTM model development

The LSTM model was developed and trained using an augmented dataset generated within the parameter ranges outlined in Table 2. Each dataset sample, representing a simulated triaxial test, comprises six curves: q vs. ε_1 and ε_v vs. ε_1, each under σ_3 = 100 kPa, 200 kPa, and 400 kPa. The temporal dependencies within these curves make the LSTM architecture a suitable choice.

The model aims to predict strength and stiffness parameters for the HS model, but several challenges necessitate tailored pre-processing and design considerations:

(1) The varying scales of parameters (e.g. E_{50}^{ref} vs. c') require normalization to ensure consistent training.
(2) The distinct nature of predicted parameters necessitates a multi-model approach for optimal accuracy.
(3) Some parameters, like c', may include zero values, requiring metrics that account for sparsity.
(4) Triaxial test measurements are used as model inputs (no additional feature selection process is necessary), with no inter-sample temporal dependencies (i.e. no overlap between samples).
(5) Each sample has inherent temporal dependencies within its features.

The following subsections detail critical steps in developing the LSTM model.

3.4.3.1. Data preparation. The data preparation process is structured into key steps to ensure compatibility with the LSTM model, optimizing its performance through proper pre-processing. These steps are described as follows:

(i) Normalization. Input features $\boldsymbol{X}$ and target variables $\boldsymbol{y}$ are standardized to ensure consistency and mitigate sensitivity to varying scales of the features. Standardization transforms values into a standard normal distribution with zero mean and unit variance, as described by Eq. (4) (LeCun et al., 2015).
(ii) Reshaping and temporal ordering. Input features are reshaped into a 3D format (Eq. (5)) to capture the sequential and temporal dependencies, represented by axial strain increments, required by the LSTM model.
(iii) Multi-model approach: Seven distinct LSTM models are trained, each focusing on predicting a single HS model parameter per case. This approach emphasizes the unique characteristics of each target and applies weighted losses to

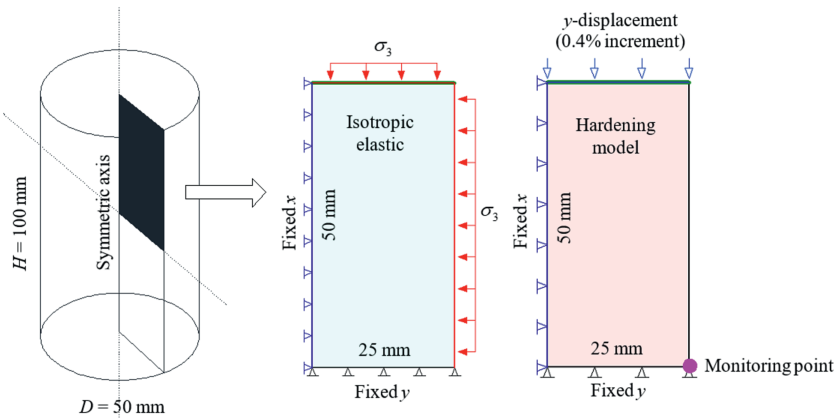


Fig. 7. Axisymmetric finite element model (SIGMA/W).

account for their differing scales and importance (Bengio, 2012).

(iv) Data splitting and randomization. Case indices are shuffled to ensure a random distribution, minimizing biases from ordered data. The dataset is then split, with the validation set comprising 20 % of the training set, maintaining representativeness.

$$\boldsymbol{X}_{\text{input}} = (\boldsymbol{X} - \mu_{\boldsymbol{X}}) \Big/ \sigma_{\boldsymbol{X}} \tag{4}$$

$$X_{\text{input}} = \left(n_{\text{cases}}, timesteps, n_{\text{features}}\right) \tag{5}$$

where $\mu_{\boldsymbol{X}}$ is the mean, $\sigma_{\boldsymbol{X}}$ is the standard deviation of the feature set, *timesteps* represents the axial strain increments, n_{features} is the number of input features, and n_{cases} is the number of samples.

3.4.3.2. Feature selection. This study utilizes triaxial test results as input features, as these measurements directly capture the physical behavior of soil under varying confining pressures. Given these features' inherent physical meaning and well-established relationships with the target parameters, no additional feature selection process is necessary (refer to Section 3.1 and Fig. 4). The input dataset includes q vs. ε_1 and ε_v vs. ε_1 curves under $\sigma_3 = 100$ kPa, 200 kPa, and 400 kPa, and ν_{ur}. ν_{ur}, which affects soil behavior, is determined through laboratory tests by analyzing axial and lateral strain data (Atkinson, 2007). This value is essential for understanding the soil's elastic properties and its stress response. Each triaxial test is considered an independent sample, where temporal dependencies are present only within the sample itself. There are no temporal dependencies between different samples.

To quantify the correlation between each feature and the target parameters, feature importance was assessed using the permutation method. This approach evaluates the effect of each feature on model performance by permuting its values and observing the resulting change in error (Fisher et al., 2019). By comparing the baseline error with the error after permuting each feature, we quantified the importance of each feature. The error was evaluated using mean absolute error (MAE). This approach is ideal for time-series models like LSTM, as it preserves the temporal structure of the data, making it more suitable than model-agnostic methods like SHAP, which are typically used for static datasets.

3.4.3.3. Model structure, training, and optimization. Fig. 8 presents a schematic view of the model architecture designed for predicting a single output parameter. The input, representing a triaxial test, comprises numerical variables (6 six temporal triaxial curves and one constant feature ν_{ur}). Temporal variables are sampled at regular axial strain intervals to maintain a consistent window size across all samples, as detailed in Section 3.4.2. The pre-processed data were reshaped into a 3D tensor, as presented in Eq. (5). This study employs seven parallel LSTM networks, each following the architecture depicted in Fig. 8, to accurately predict the seven HS model parameters. The design, training, and optimization processes of the model are detailed in the subsequent sections.

To enhance the performance of the LSTM model, each hidden layer employed the Leaky rectified linear unit (Leaky-ReLU) activation function to improve the model's nonlinear characteristics. The Leaky-ReLU addresses the “dying ReLU” problem by allowing a slight, non-zero gradient for negative input values (Qi et al., 2024). This study incorporated dropout and L2 regularization techniques to mitigate overfitting in the LSTM model. Applying both methods can improve the generalization performance on unseen data. Dropout minimizes the risk of neurons becoming overly reliant on one another, while L2 regularization ensures that the weights remain small (Srivastava et al., 2014). The mean squared error (MSE) was utilized as the loss function, while the Adam optimizer was employed for model training as more adaptable and efficient. An Early stopping criterion was applied to select the best-performing model weights and prevent overfitting or underfitting issues. Learning curves were analyzed to monitor the model's performance during training and validation.

The hyperparameters of the LSTM model, such as the number of neurons, dropout rate, the L2 parameter, the batch normalization technique, and the Leaky-ReLU parameter, were determined through the Bayesian optimization technique (Snoek et al., 2012) with a search space outlined in Table 3.

The implementation was carried out using TensorFlow, a Python-based DL framework that enables the development of high-level neural networks (Abadi et al., 2016). Keras Tuner (O'Malley et al., 2019), a library for hyperparameter optimization, was employed to explore model configurations efficiently through Bayesian optimization. The model architecture was built using the sequential API, incorporating layers with customizable regularization options. In addition, the implementation utilized several Python libraries to support data analysis and visualization, including Pandas for data manipulation (Reback et al., 2020), NumPy for numerical computations (Harris et al., 2020), Matplotlib for data visualization (Hunter, 2007), Seaborn for statistical plotting (Waskom, 2021), and scikit-learn for data pre-processing (Pedregosa et al., 2011). Model training was performed on the high-performance computing (HPC) platform, leveraging its high-performance Jupyter Notebook environment. The hardware configuration included an NVIDIA RTX 3080 Ti GPU, 64 GB of RAM, six CPU cores, and one GPU card, running on Anaconda (2023-03-01) with Python 3.10.9 (The Hong Kong Polytechnic University HPC, 2025).

3.4.3.4. Model evaluation. For this objective, three primary metrics were employed to evaluate model performance: MAE, mean absolute percentage error (MAPE), and normalized mean absolute error (NMAE). These metrics assessed the model's accuracy in predicting each parameter. MAE quantifies the average absolute error across all predictions, as shown in Eq. (6). MAPE measures the error percentage relative to the actual values, as described in Eq. (7). However, MAPE can be sensitive to small or large values of (y_i), leading to disproportionate errors. To mitigate this issue, the NMAE is proposed. NMAE normalizes the MAE over the parameter range, providing a more balanced metric regardless of the parameter's magnitude (Eq. (8)). Lower values of MAE, MAPE, and NMAE indicate higher accuracy, with values approaching zero being ideal. Additionally, the coefficient of determination (R^2) was used to quantify prediction accuracy further. R^2 is a statistical measure of how well the predicted values align with the actual values (Eq. (9)). A higher R^2 value (closer to 1) indicates better predictive performance, while a lower value signifies worse accuracy (Hanane, 2023).

The MC dropout is integrated to quantify predictive uncertainty (Gal and Ghahramani, 2016; Zhang et al., 2021b). In this technique, the dropout layers are maintained. For each input $\boldsymbol{x}$, we perform $T = 1000$ stochastic forward passes, which yield a set of predictions $\left\{f^{(t)}(x)\right\}_{t=1}^{T}$ where the final prediction is then computed as the mean of these outputs. Predictive uncertainty is approximated by the sample standard deviation (Eq. (10)). It is important to note that while all these metrics offer valuable insights into model

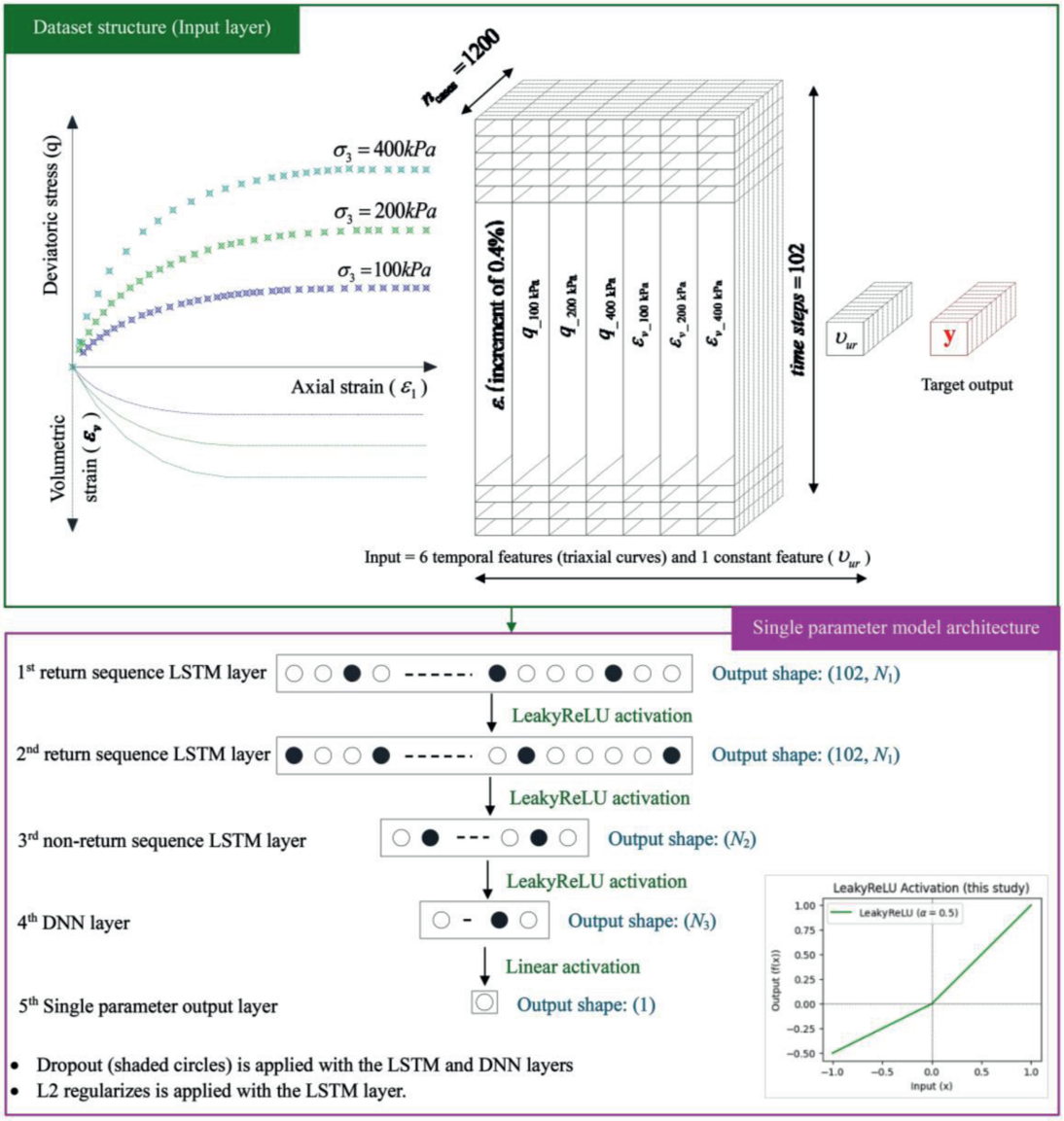


Fig. 8. LSTM model architecture for a single parameter.

Table 3
Search space of the LSTM model design parameters.

Parameter	Search space/value
Features	Six temporal features (triaxial curves) and one constant feature (ν_{ur})
Window size (ε_1_range)	102
Hidden layers (hidden units)	Two layers, LSTM return sequence (N_1 = 256, 512, 1024 units) One layer, LSTM non-return sequence (N_2 = 128, 256, 512 units) One layer, deep neural network (DNN) (N_3 = 64, 128, 256 units)
Loss function	$MSE = \sum_{i=1}^{n}(y_i - \tilde{y}_i)/n$, where y_i represents the actual value, $\tilde{y}_i$ is the predicted value, and n is the number of data points
Optimizer	Adam
Dropout	D = 0.1, 0.3, 0.5, 0.7, 0.9
Leaky-ReLU	LR = 0.1, 0.3, 0.5, 0.7, 0.9
L2_regularizers	$L2$ = 0.0001, 0.001, 0.01, 0.1
Batch normalization	B = True, False

reliability, MAE is beneficial because it quantifies errors in the same units as the parameter, thereby facilitating direct deterministic geotechnical interpretation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \tilde{y}_i| \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| \times 100\% \tag{7}$$

$$NMAE = \frac{MAE}{y_{\max} - y_{\min}} \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{9}$$

where y_i represents the actual value; $\tilde{y}_i$ is the predicted value; n is the total number of predictions; $y_{\max}$ and $y_{\min}$ represent the maximum and minimum values of the parameter range, respectively; and $\bar{y}$ is the mean of the observed data.

$$\sigma(x) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(f^{(t)}(x) - f'(x)\right)^2} \tag{10}$$

where $f^{(t)}(x)$ represents the model's prediction for the input x during the t-th stochastic forward pass, and $f'(x)$ is the mean of all stochastic predictions obtained from T forward passes.

4. Results and discussion

4.1. Augmented dataset

Fig. 9 shows the frequency distribution of the HS parameters, with histograms depicting how frequently each parameter takes specific values across the dataset. Fig. 10 presents a scatter pair plot, displaying the relationships between pairs of HS parameters, with each point representing a case and its corresponding parameter values. Most parameters show a relatively uniform frequency distribution. During the development of the numerical simulation, we ensured that each parameter set, which corresponded to laboratory tests, had an apparent failure occurring within the strain range. Consequently, parameters such as the R_f and m exhibit higher frequencies for specific values (Fig. 9) as they were manually adjusted, ensuring that failure is reached. To illustrate, m was manually adjusted for the cases that correspond to a lower value of the E_{50}^{ref}, and the E_{oed}^{ref} was adjusted by increasing the value of m. This can be visualized in the pair plot between these parameters in Fig. 10. Similarly, the upper limits of the R_f were modified (reduced for the cases with lower E_{50}^{ref} and E_{oed}^{ref}, as visualized in Fig. 10). This explains why specific values of these parameters occur more frequently than others.

The distribution of E_{50}^{ref} is skewed, as it is constrained by the maximum value of E_{ur}^{ref}, where the maximum value of $E_{50}^{ref} = 0.5E_{ur}^{ref}$. A similar pattern is observed with E_{oed}^{ref}, which is assumed to be within the range of $0.5E_{50}^{ref}$ to $0.8E_{50}^{ref}$. Fig. 10 confirms these relationships, demonstrating that they align with the theoretical assumption and conform to the expected range. Parameters such as c', ϕ', and ν_{ur} are uniformly distributed. ν_{ur} is occasionally simulated outside the predefined bounds to enhance the model's learning capability, as this parameter has minimal influence on the simulation results.

Fig. 11 illustrates the relationships between ε_1, q, and ε_v as simulated using the SIGMA/W axisymmetric finite element model (GeoStudio, 2023). These simulations represent the augmented parameter space defined in Section 3.3 and detailed in Figs. 9 and 10. Fig. 11 comprises six subplots, each corresponding to a specific σ_3, and visualizes 1200 curves representing the response of 1200 unique cases, each defined by a labeled parameter set. Additionally, these plots include a heatmap representation of ν_{ur}. The relationships are integral to both model training and validation as input features.

4.2. DL tuning

DL is fundamentally an empirical process that requires careful tuning to achieve optimal performance. In this study, the Bayesian optimization technique (Snoek et al., 2012) was employed to optimize the parameters for each DL model. A maximum of 100 trials were conducted, with 250 training epochs per trial, to initially explore the parameter space effectively. Fig. 12 presents the density map of each variable against the loss. Some variables exhibit a consistently lower loss for specific models, while others do not show a clear pattern. It is also observed that multiple combinations of hyperparameters can result in similar loss values. Consequently, the optimal parameter set identified through Bayesian optimization was subsequently used to train the models, following a coarse-to-fine tuning approach (refer to Figs. 16–20 in Ebrahim et al. (2024) for further details). Early stopping criteria were applied, and the training and validation processes were

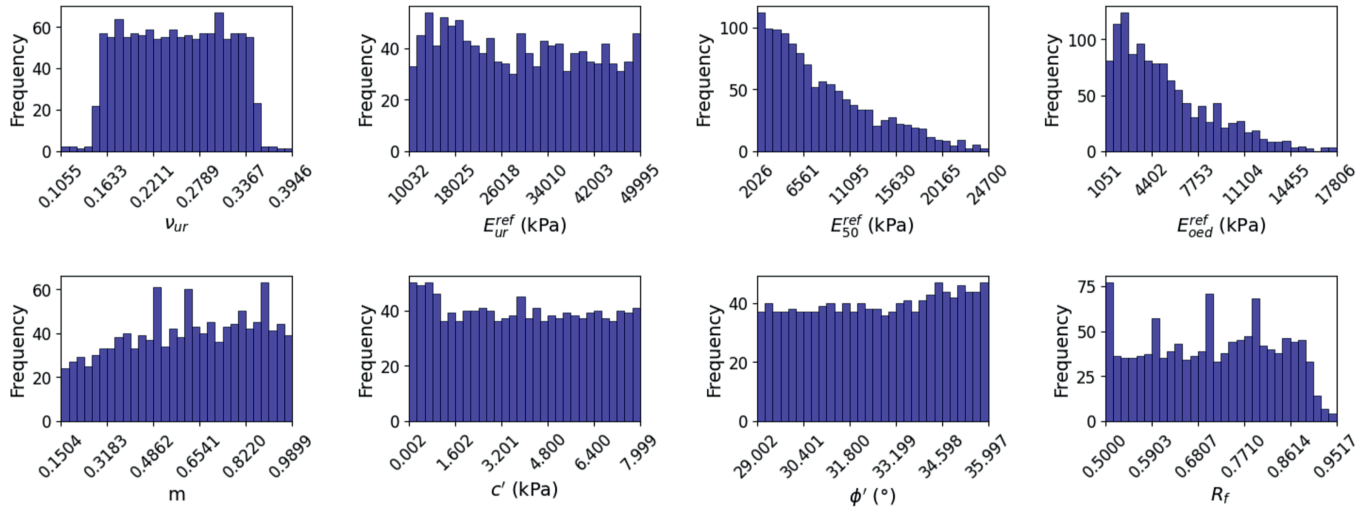


Fig. 9. Frequency distribution of HS parameters (augmented dataset).

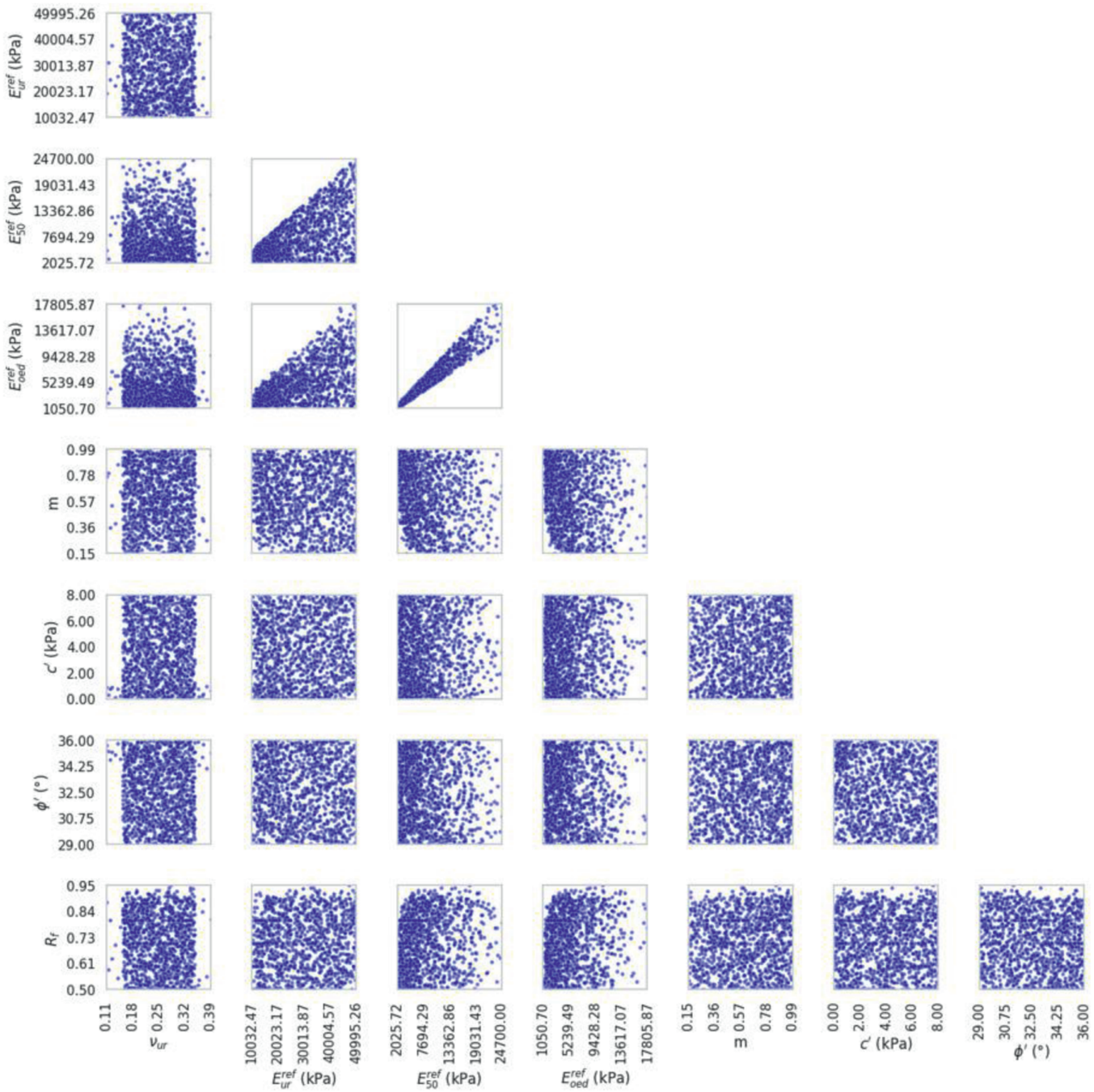


Fig. 10. Pair plot relationships between HS parameters (augmented dataset).

closely monitored by visualizing the loss curves to refine parameters iteratively. For instance, this process helped to identify and address overfitting or underfitting by adjusting regularization values (i.e. *L2*) accordingly (Ebrahim et al., 2024). The error benchmark for model evaluation was set to an NMAE of less than 1 %, with the best-performing model either selected from the tuning process or manually adjusted based on learning curves, being the one that met this criterion. This methodology helps to mitigate the risk of being trapped in local minima, and it also helps to balance overfitting and underfitting. As a result, computational efficiency was improved, and higher accuracy was achieved. The final optimized parameter sets are presented in Table 4.

4.3. Model training and evaluation

A hold-out validation strategy was employed to ensure the robustness and reliability of the LSTM models. The dataset, comprising 1200 unique samples with no overlapping responses, was randomized and split into 1000 training and 200 validation samples. The model was trained using early stopping to prevent overfitting, and the optimized weights were saved. This process was repeated with different randomized splits, using the previously saved weights as initialization for further optimization. Given the uniqueness of each sample in the dataset, this strategy allowed the model to generalize effectively by exposing it to

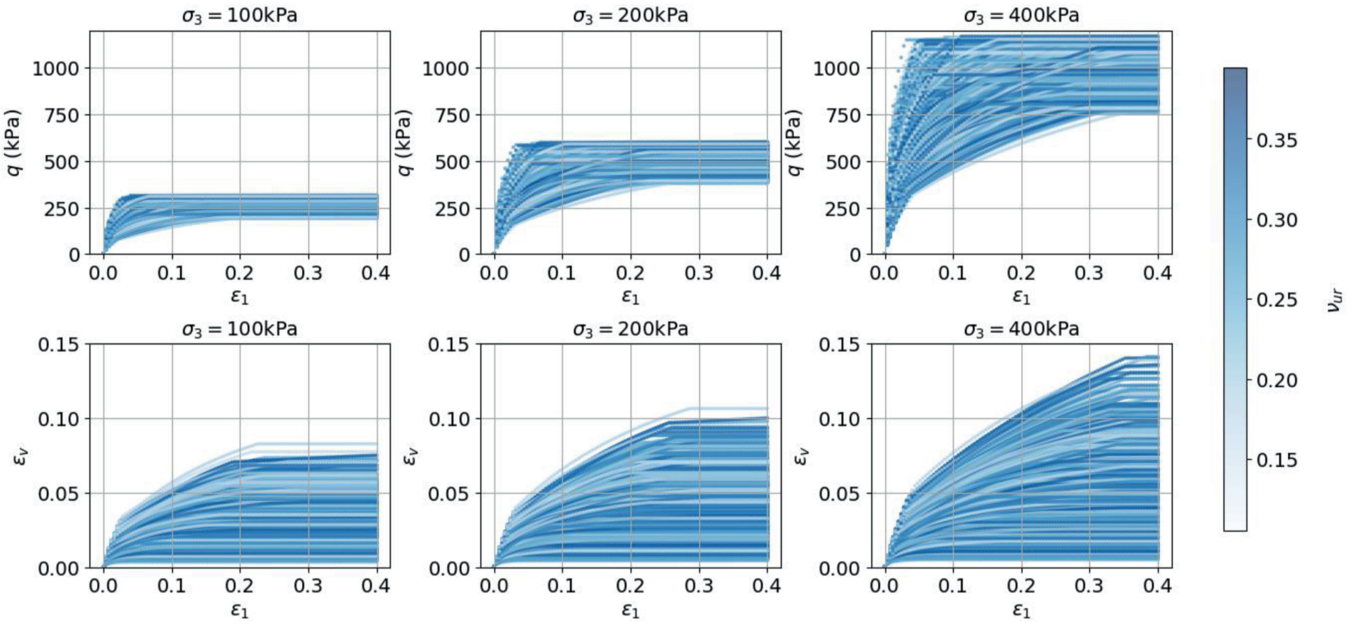


Fig. 11. The augmented stress–strain curves (q vs. ε_1 and ε_v vs. ε_1) under $\sigma_3 = 100$ kPa, 200 kPa, and 400 kPa, and ν_{ur} as a heatmap.

diverse training-validation splits while incorporating dropout variability (Ebrahim et al., 2025). To ensure model consistency, L2 regularization stabilizes weight updates and reduces sensitivity to data variations (Srivastava et al., 2014). For instance, Fig. 13 illustrates the training process for some parameters, such as E_{50}^{ref} and R_f, where initial difficulties in predicting specific samples were resolved through resampling.

Fig. 14 presents the learning curves for the optimized LSTM model (taking ϕ' and m as an example). The curves highlight the model's strong performance, as evidenced by the declining loss values, particularly during the initial epochs, followed by stabilization at a consistently low loss level. The close alignment between the training and validation loss curves indicates minimal risk of overfitting. However, minor noise and oscillations are observed during the training process, which can be attributed to the Adam optimization algorithm and random dropout techniques.

Figs. 15 and 16 present scatter plots of predicted vs. actual values for the strength and stiffness parameters, respectively, demonstrating the robust performance of the LSTM models in predicting the HS parameters. Each subplot evaluates the model's accuracy for a specific parameter, with both training and validation sets clearly delineated (training samples as blue circles and validation samples as red circles). The model achieved exceptional accuracy with an NMAE of 0.66 % and 0.74 % for c' (Fig. 15a). For ϕ' (Fig. 15b), the NMAE was 0.24 % and 0.22 %. The model also demonstrated high accuracy in predicting R_f (Fig. 15c) with an NMAE of 0.83 % and 0.84 %. E_{ur}^{ref} (Fig. 16a) achieved NMAEs of 0.9 % and 1.23 %, while E_{50}^{ref} (Fig. 16b) reached 0.7 % and 0.83 %. E_{oed}^{ref} (Fig. 16c) showed NMAEs of 0.39 % and 0.42 %, and the m (Fig. 16d) demonstrated NMAEs of 0.25 % and 0.24 % for the training and validation sets, respectively.

Notably, the proximity of training and validation errors across all parameters highlights the absence of overfitting or underfitting, emphasizing the model's robustness. Figs. 15 and 16 also demonstrate that the validation samples are well-distributed across the parameter space, ensuring comprehensive evaluation and consistent performance for all HS parameters. Table 5 presents the evaluation metrics for all models, including the MAE, MAPE, NMAE, and R^2. The models demonstrated exceptional performance, with average NMAE, MAPE, and R^2 values of approximately 0.606 %, 0.975 %, and 0.9987, respectively, underscoring their accuracy and reliability, indicating an accuracy of roughly 99.4 %.

In scenarios characterized by complex or noisy data, we integrated MC dropout to quantify predictive uncertainty using an unseen test set of 100 samples. Figs. 15 and 16 illustrate the actual vs. predicted values (magenta), the scatter plots of MAE values (gray), and the corresponding histograms of residuals for each parameter (blue and magenta). The analysis shows that residuals exhibit minimal error across parameters, with approximately bell-shaped distributions, except for a slight skew in the stiffness parameters. Notably, the estimated uncertainties (σ) are minor relative to the range of each parameter. For instance, $\sigma_{E_{ur}^{ref}}$ recorded around 0.76 (ranging from 10 MPa to 50 MPa). Moreover, the model performance is robust, with R^2 values exceeding 0.99 and MAE values comparable to those from the deterministic predictions (see Table 5). The proximity of the performance could be illustrated as follows: (1) extracting axial strain in small increments helps to represent each curve with enough points, minimizing the impact of any potentially noisy data; (2) adopting a multi-approach that assigns distinct weights for each parameter; and (3) applying a structured tuning process, from coarse to fine, with integrated dropout and regularization. Given the similar performance of deterministic and stochastic predictions, using the standard deviations of residuals from training and evaluation offers a time-efficient way to estimate uncertainty without requiring additional sampling or computationally expensive techniques at prediction time.

4.4. Model testing, validation, and deployment

After the training phase, the model was tested using unseen data. The first test was a numerical sample parameterized by GEOSLOPE International Ltd., (2021), represented by Ottawa sand (Krachtoudi, 2011). This test served two purposes: to evaluate the model's ability to parameterize the HS model for unseen data

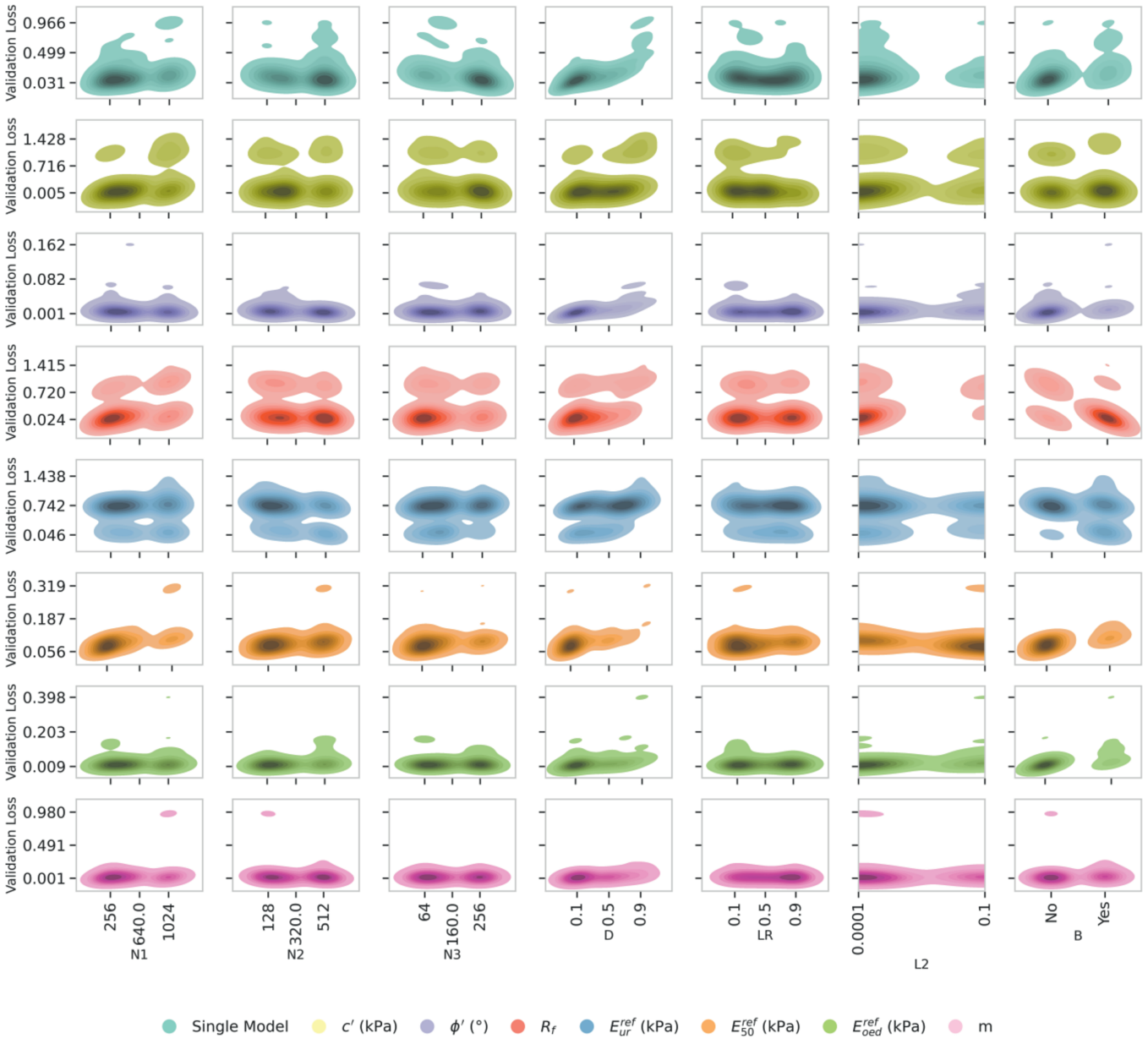
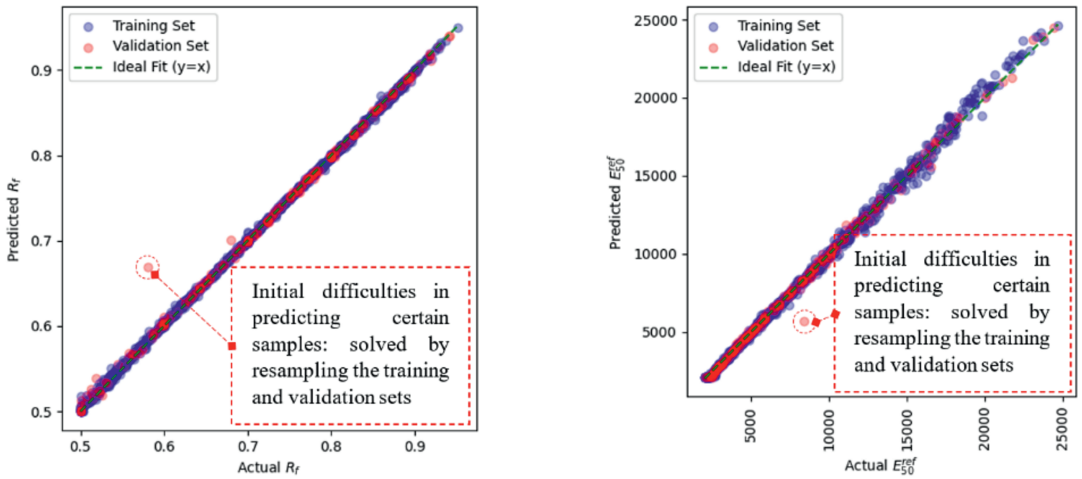


Fig. 12. Density maps of hyperparameters vs. validation loss.

Table 4
The final multi-LSTM model parameters.

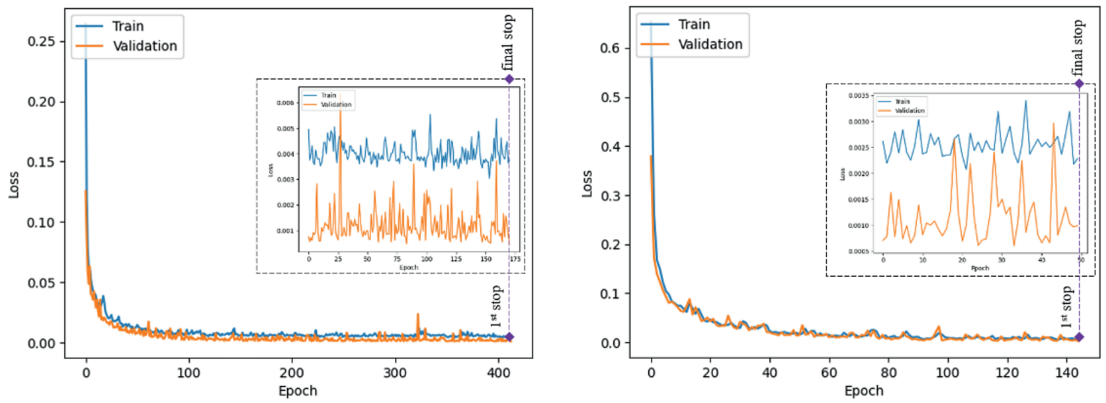
Parameter	Value (target)							
	Single LSTM[b]	Multi-LSTM						
		c'	ϕ'	R_f	E_{ur}^{ref}	E_{50}^{ref}	E_{oed}^{ref}	m
N_1	512	256	256	512	512	512	256	512
N_2	512	256	128	256	256	256	128	256
N_3	256	128	64	128	128	128	64	128
D	0.1	0.1	0.1	0.3	0.3	0.3	0.1	0.1
LR	0.7	0.9	0.9	0.5	0.5	0.5	0.9	0.9
$L2$	0.001	0.0001	0.0001	0.001	0.0001	0.001	0.0001	0.0001
B	False	False	False	True	False	True	False	False
Initial tuning durations (s)[a]	21,881	21,542	26,068	28,708	29,342	22,815	23,510	21,932

[a] These durations are based on the high computational power provided by the HPC server.
[b] This refers to a single model that predicts all seven parameters simultaneously.



(a) R_f (b) E_{50}^{ref}

Fig. 13. Model evaluation process using hold-out validation strategy (1st sampling dataset).



(a) Internal friction angle (ϕ') (b) Power for the stress level dependency (m)

Fig. 14. The training learning curves.

accurately and to validate the numerical model used to generate the training dataset. Since this soil sample falls within the parameter range used during model training, it provided a suitable benchmark for comparison. Table 6 presents the HS model parameters calculated by GEOSLOPE International Ltd., (2021) alongside the parameters retrieved using the APHS tool developed in this study. The comparison demonstrates that the APHS tool achieved an accuracy of approximately 99.6 %, with minimal absolute error for each parameter. Furthermore, Fig. 17 visualizes the simulated results obtained using parameters from both the APHS tool and the conventional parameterization method. The results confirm the precision of the APHS tool, validating its reliability and the accuracy achieved during training and validation when applied to unseen data.

The APHS tool was further applied to another real-world laboratory case study to validate its reliability. Zhou (2008) conducted conventional consolidated drained triaxial tests on recompacted saturated CDG soil. These tests aimed to determine its shear strength under σ_3 of 100 kPa, 200 kPa, and 400 kPa as part of a soil nailing research project. The laboratory data, along with ν_{ur} of 0.3 (Zhou, 2008), were used as inputs for the APHS tool. The HS parameters extracted by the tool were then simulated using an axisymmetric finite element model (SIGMA/W) (GeoStudio, 2023), and the simulation results were compared against the original laboratory data for further validation, as shown in Fig. 18.

From the laboratory tests at 15 % and 20 % strains, the calculated shear strength parameters were $c' = 0$ kPa and $\phi' = 35°$ and 35.1°, respectively. The APHS tool predicted $c' = 0.49$ kPa and an ϕ' of 34.96°, showing minimal differences. These minor differences did not significantly influence the simulation results, as demonstrated in Fig. 18. The differences in ϕ' align with the uncertainties shown in Figs. 15 and 16 and the confidence intervals (CIs) in Table 7. The discrepancies in c' may stem from inherent model uncertainties or from measurement noise. Measurement noise could also have been introduced during the digitization of laboratory data using the WebPlotDigitizer tool (Rohatgi, 2022). To illustrate, the original data from Zhou (2008) were converted from portable document format (PDF) to numerical format for analysis, which might have contributed to these discrepancies. The extracted parameters and their corresponding 99 % CI are presented in Table 7, providing users with a prediction range to account for any possible deviations. The CIs are calculated using the standard deviations of residuals during the training and evaluation process. The results in Table 7 align with the MC dropout uncertainty estimates, further validating the rapid interface predictions.

The APHS tool, developed using the Streamlit framework in

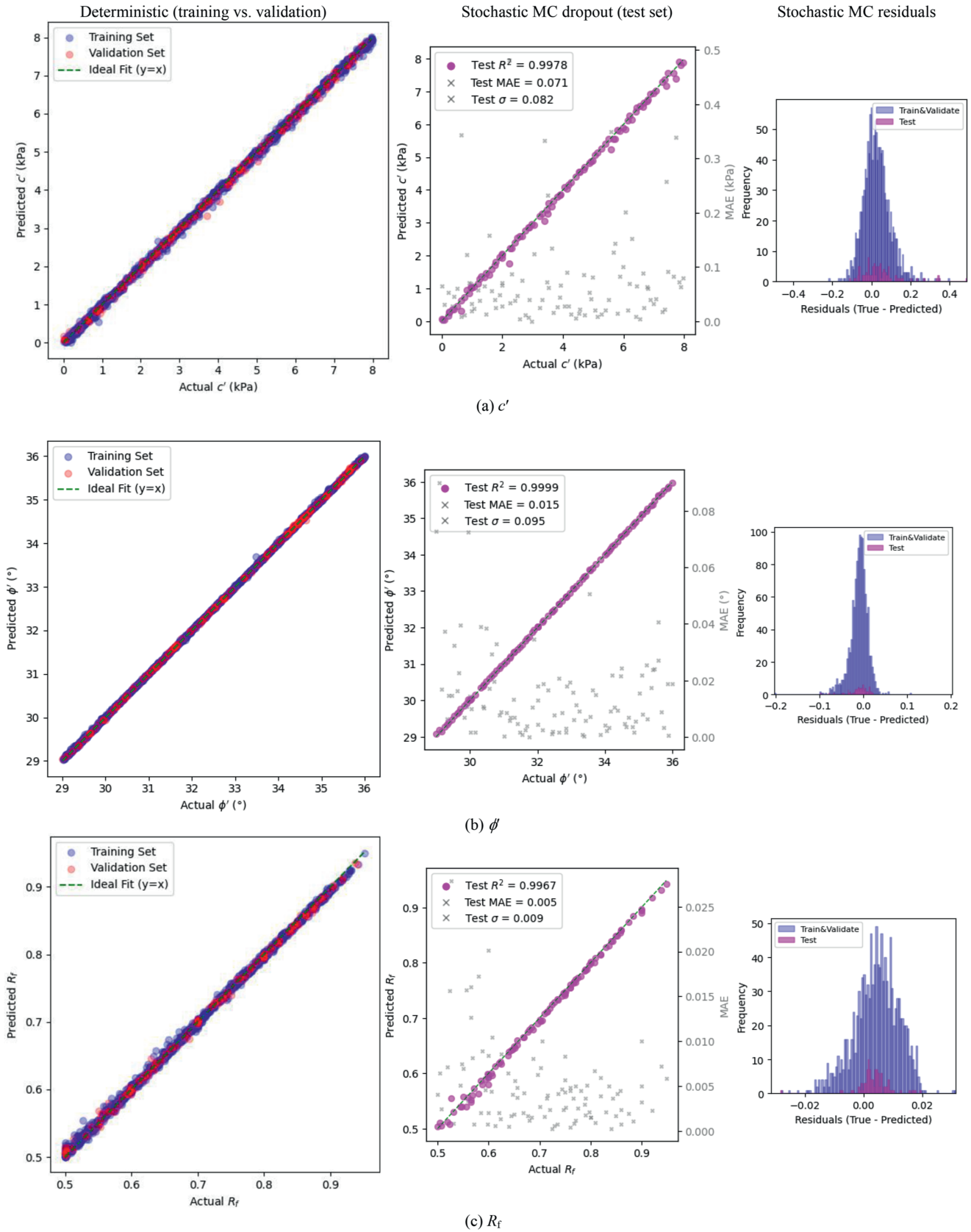


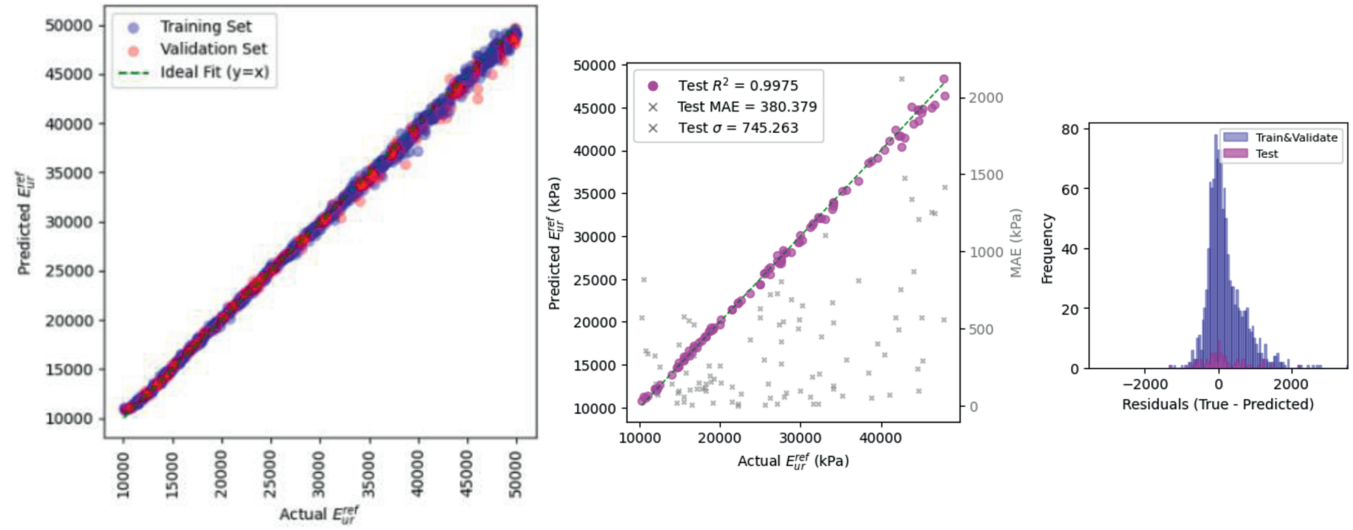
Fig. 15. Deterministic (training and validation sets) and stochastic (test set) predicted strength parameters vs. actual HS strength parameters.

Python (Khorasani et al., 2022), demonstrates remarkable efficiency in automating the HS model parameterization process, as illustrated in Fig. 19. The tool's web-based interface eliminates the traditional manual curve-fitting approach, significantly reducing the time and effort required for parameter identification while maintaining high accuracy.

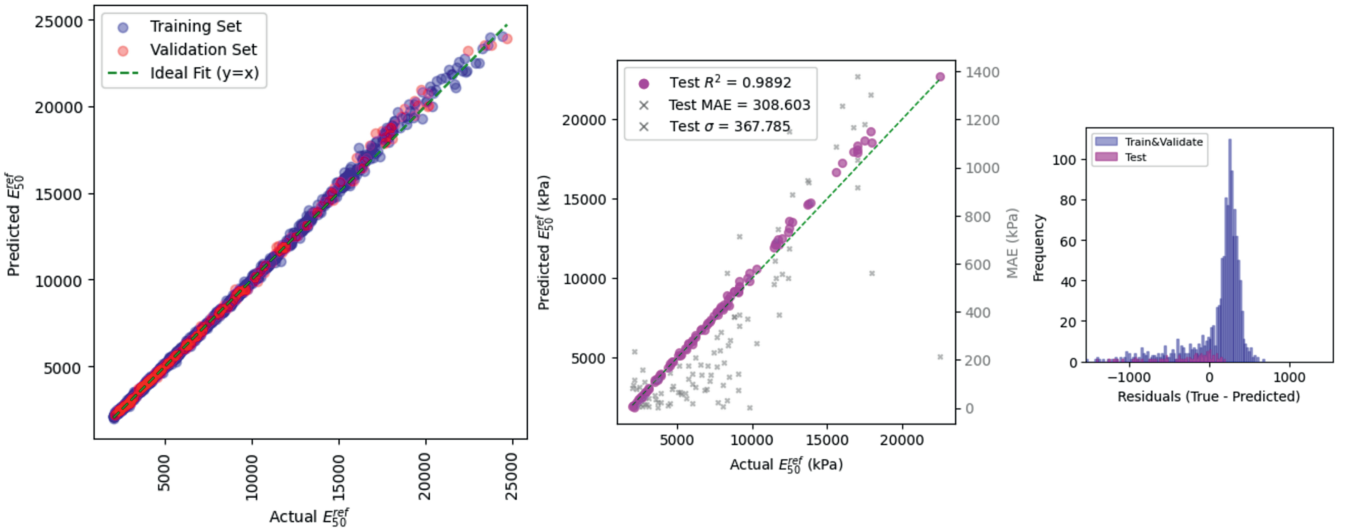
4.5. Discussion

This study introduces the APHS tool, a significant advancement in automating and streamlining the parameterization process of the HS model. Conventional methods typically rely on oedometer tests, unloading/reloading stress–strain data, empirical correlations, domain-specific assumptions, or expert judgment, all of which can be time-consuming, costly, and error-prone. In contrast, the APHS tool exclusively uses data from conventional triaxial tests, making it a versatile solution for geotechnical applications. The APHS tool employs DL, specifically LSTM networks, to capture the relationship between stress and strain under loading conditions. By modeling this relationship as a time-series problem, the tool effectively leverages the sequential evolution of stress–strain responses, offering a more efficient approach.

The development of the APHS tool systematically addressed key challenges related to data generation and model optimization. Given the limitation of insufficient triaxial laboratory data, the study focused on a specific case study in Hong Kong to establish a parameter space suitable for numerical simulations. This parameter space was derived from field samples collected in Hong Kong and supported by existing literature, ensuring the tool's practical applicability to real-world conditions. To efficiently explore this parameter space, LHS was employed, facilitating the generation of diverse synthetic datasets that comprehensively cover the key parameters of the HS model. The numerical model SIGMA/W was then integrated to simulate soil responses under these defined

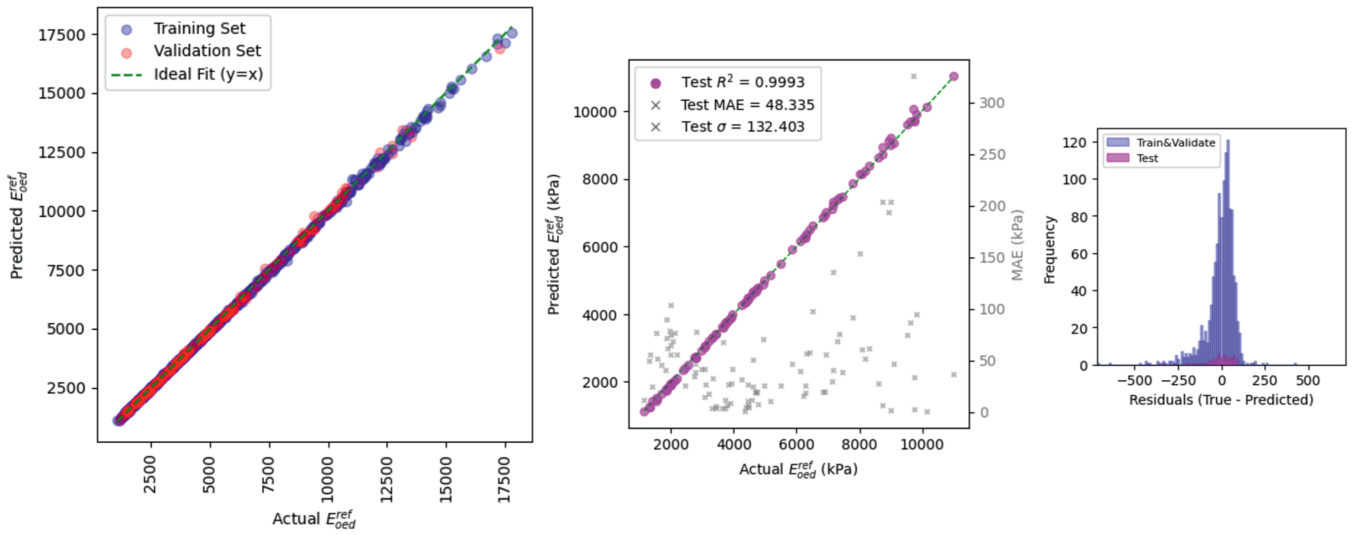


(a) E_{ur}^{ref}

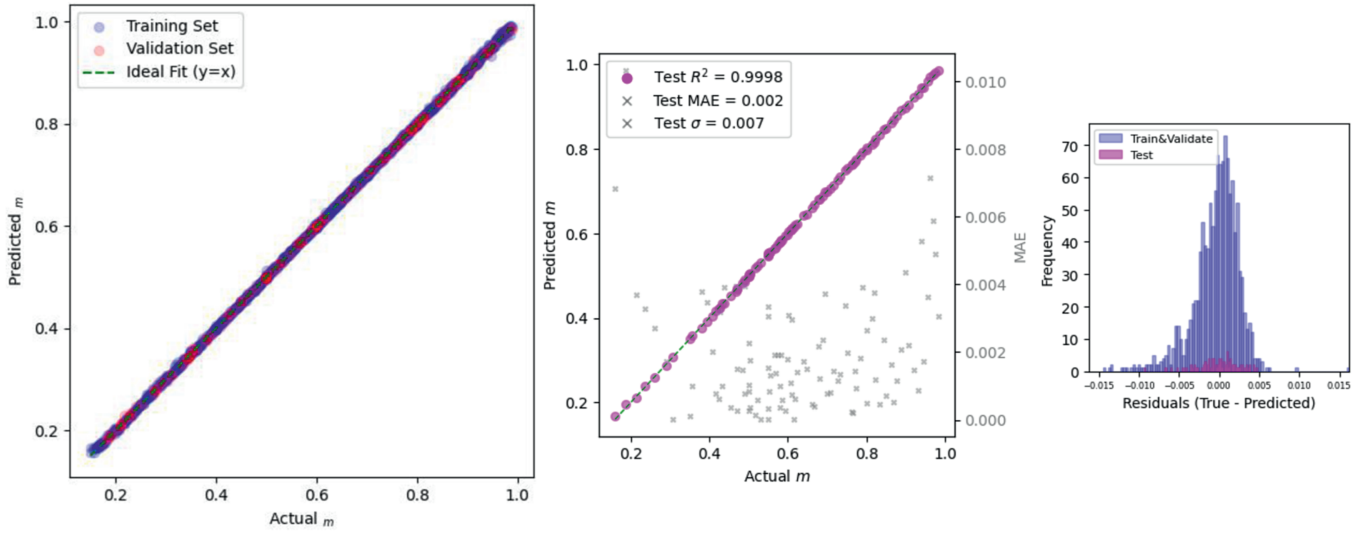


(b) E_{50}^{ref}

Fig. 16. Deterministic (training and validation sets) and stochastic (test set) predicted stiffness parameters vs. actual HS stiffness parameters with stochastic test MAE and residuals.



(c) E_{oed}^{ref}



(d) m

Fig. 16. (*continued*).

conditions, producing stress–strain curves for DL training and validation.

For model optimization, a critical consideration was the variation in scale among input parameters, such as the stress–strain values, and the target parameters, such as c' vs. E_{ur}^{ref}. Data preparation played a significant role in this process, particularly through normalization techniques that ensured uniformity across all input features. An extensive coarse-to-fine-tuning strategy was employed. The process included early stopping, close monitoring of the loss curves, and reshuffling the training and validation sets to mitigate overfitting. Given the uniqueness of each sample, this approach enabled the model to generalize effectively.

Seven novel parallel LSTM models were implemented to enhance accuracy, each dedicated to predicting a specific parameter. This approach independently optimizes each parameter, reflecting the physical processes observed in conventional parameterization. For example, as shown in Fig. 20, the multi-model assigns distinct weights to each parameter. It gives higher weight to features such as deviatoric stress for specific targets (i.e., c'). Meanwhile, volumetric strain has less impact, which is expected. This approach ensures that the model successfully mimics the underlying physical processes (Fig. 4 and Section 3.1). Similarly, the E_{oed}^{ref} is found to be controlled with the volumetric strain features. Most significantly, as expected, E_{ur}^{ref} is influenced mainly by deviatoric stress due to the constraints applied (i.e. $E_{50}^{ref} = 0.5E_{ur}^{ref}$) during dataset generation (Fig. 10 and Table 2). Additionally, since E_{50}^{ref} is correlated with E_{oed}^{ref}, E_{ur}^{ref} also exhibits a relative correlation with volumetric strain, which governs E_{oed}^{ref}. Given the absence of unloading/reloading data, the LSTM model derived correlations from a robust dataset, informed by literature-defined boundaries and validated LHS-generated parameter sets (Section 4.1).

In contrast, the single-model approach applies uniform loss weights based on the average predicted error of all seven parameters, which may increase the risk of overfitting and reduce generalization. This is evident in Table 8, where the multi-model achieves lower MAE values. For instance, E_{ur}^{ref} exhibits a reduction from 1016.57 kPa (single model) to 380.38 kPa (multi-model) in the testing set. This trend is even more pronounced in the tested case study, where the multi-model shows an improvement, with an MAE of 104.41 kPa compared to 1477.93 kPa for the single

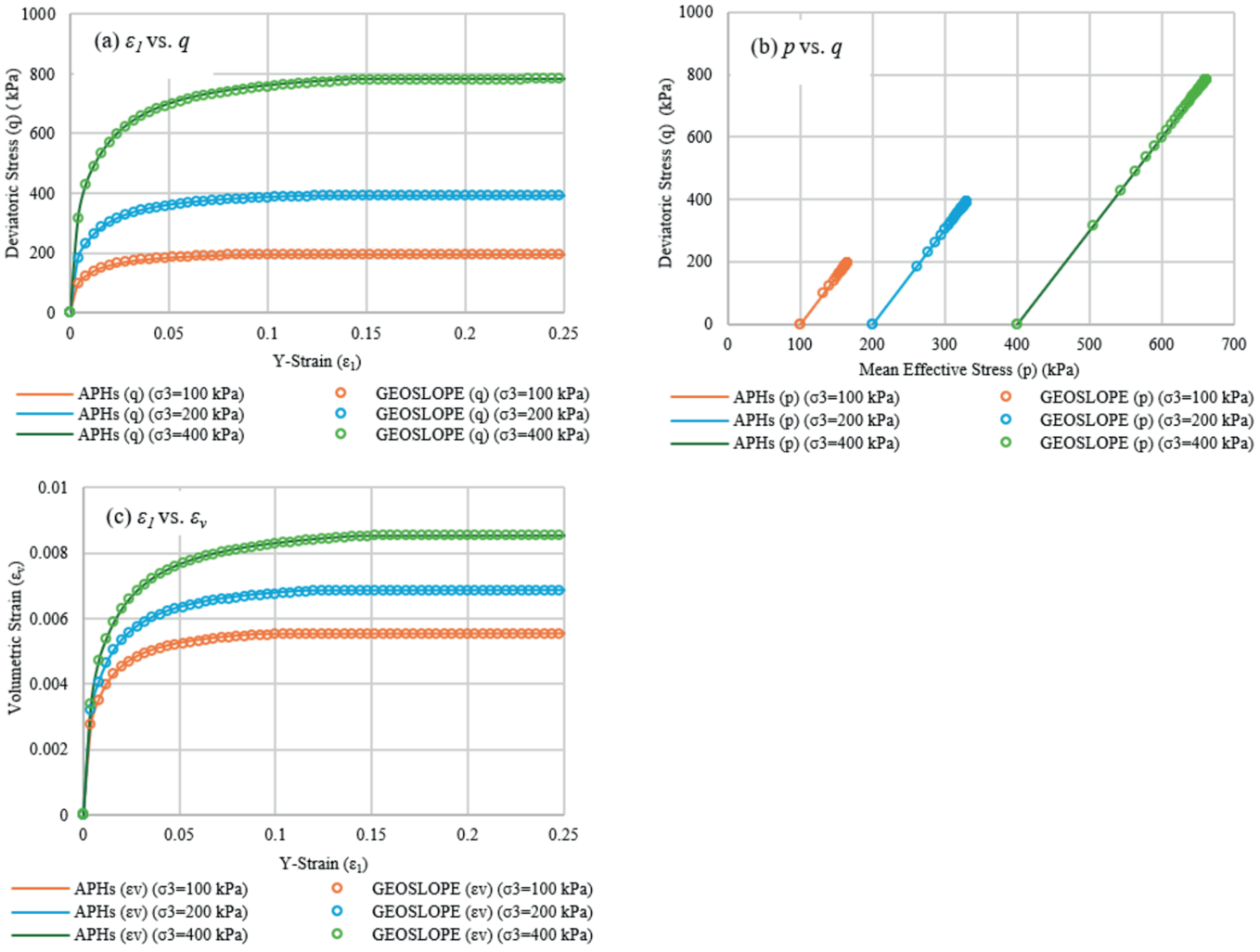


Fig. 17. Comparison of simulated results: APHS vs. conventional parameterization (GEOSLOPE International Ltd., 2021).

model for E_{50}^{ref}. The multi-model approach delivers more accurate results on unseen data and shows robustness through strong alignment between its stochastic and deterministic predictions, as confirmed by the MC dropout simulation (Table 8 and Figs. 15 and 16).

In contrast to studies that use optimization with physically informed models, this study adopts a DL approach. Rather than relying on established equations, our model learns underlying physical patterns directly from raw data, enhancing its flexibility and establishing it as a standalone tool with potentially high predictive accuracy. To illustrate its practical utility, we compared our approach with that of Jin et al. (2016), who employed a real-coded genetic algorithm to optimize parameters for Hostun sand. WebPlotDigitizer (Rohatgi, 2022) converted triaxial test plots into numerical data along useable strain sequences. Taking ϕ' as an example, Jin et al. (2016) employed a search range (10°–50°) and obtained values between 27° and 29° for different constitutive models. Remarkably, our study predicts 28.24° for the Hostun sand despite using data outside the training scope (minimum training value of 29°). As shown in Fig. 21, APHS closely match laboratory results with only minor errors.

To further enhance the simulation, fine-tuning is enriched by integrating a feature importance analysis (Fig. 20) to guide parameter adjustments. For example, deviations in volumetric strain indicate that E_{oed}^{ref} needs adjusting. Remarkably, the accurate value of $E_{\text{oed}}^{\text{ref}}$ (adjusted to 1000 kPa) falls between 911 and 1337 kPa (Table 9) despite the training minimum being 1050 kPa. Similarly, deviations in the magnitude of the deviatoric stress indicate that ϕ' needs adjusting (adjusted to 27.37°). The slight deviation in ϕ' may partly reflect laboratory bias. This is evident in the envelope plot (Fig. 21), where the APHS envelope is tangent to the first two Mohr circles but deviates slightly from the third. Overall, APHS delivers accurate parameterizations even beyond the training range and effectively complements physically informed models. They refine the optimization search space from 10° to 50° down to approximately 28°, and the feature importance in Fig. 20 further guides model calibration.

Analyzing soil samples beyond the trained parameter ranges while accounting for laboratory and numerical biases allows us to evaluate how these errors propagate through simulation results. For Hostun sand, the absolute errors were approximately 0.87° and 0.126 MPa for ϕ' and $E_{\text{oed}}^{\text{ref}}$, respectively, resulting in only minor deviations in both deviatoric stress and volumetric strain. This suggests that when soil properties fall within the training range, the MAE values in Table 5 (0.015° and 0.068 MPa) indicate minimal expected deviations, as shown in Figs. 17 and 18. However, experts should interpret simulated results, considering the specific

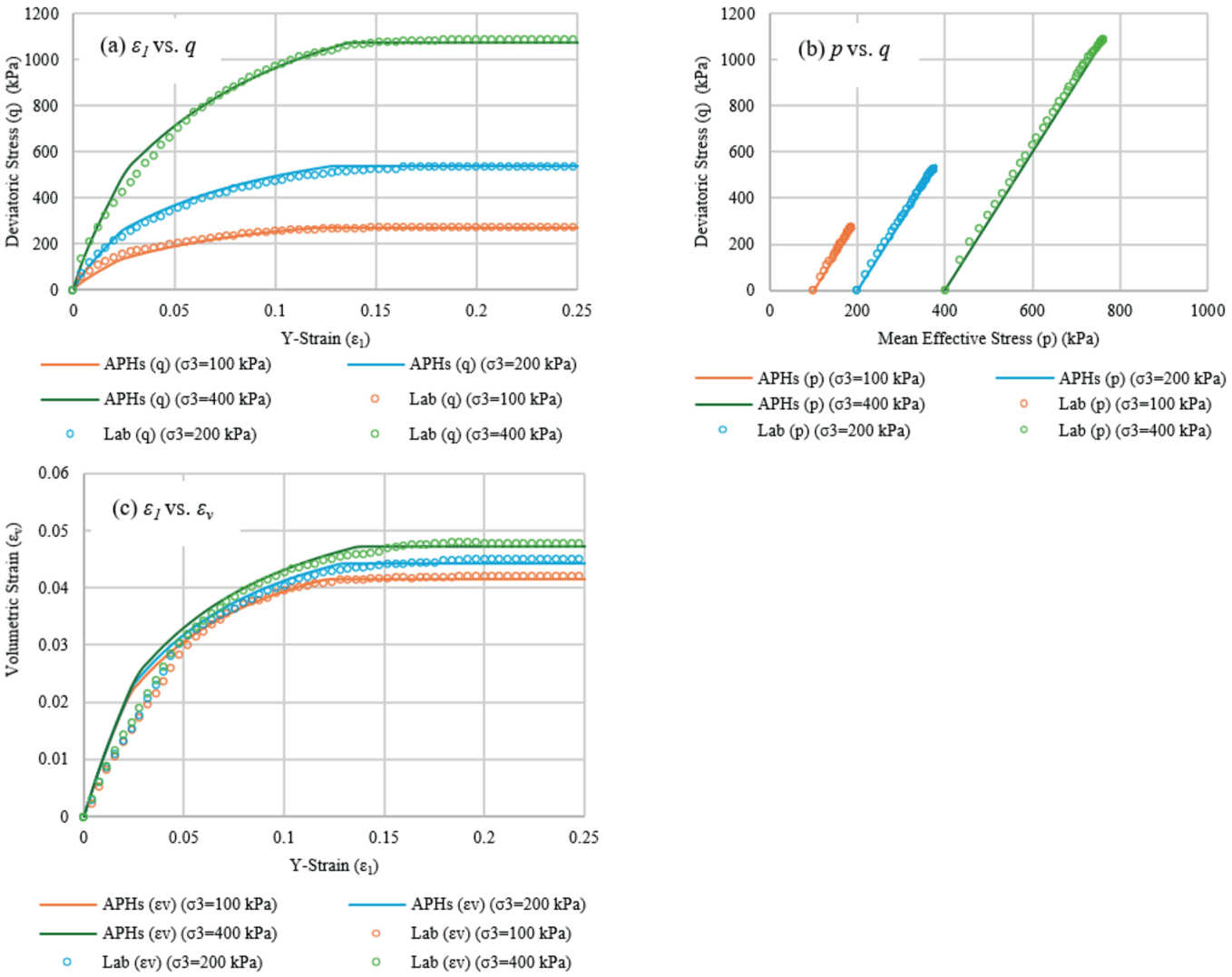


Fig. 18. Comparison of simulated results: APHS vs. laboratory triaxial test results (Zhou, 2008).

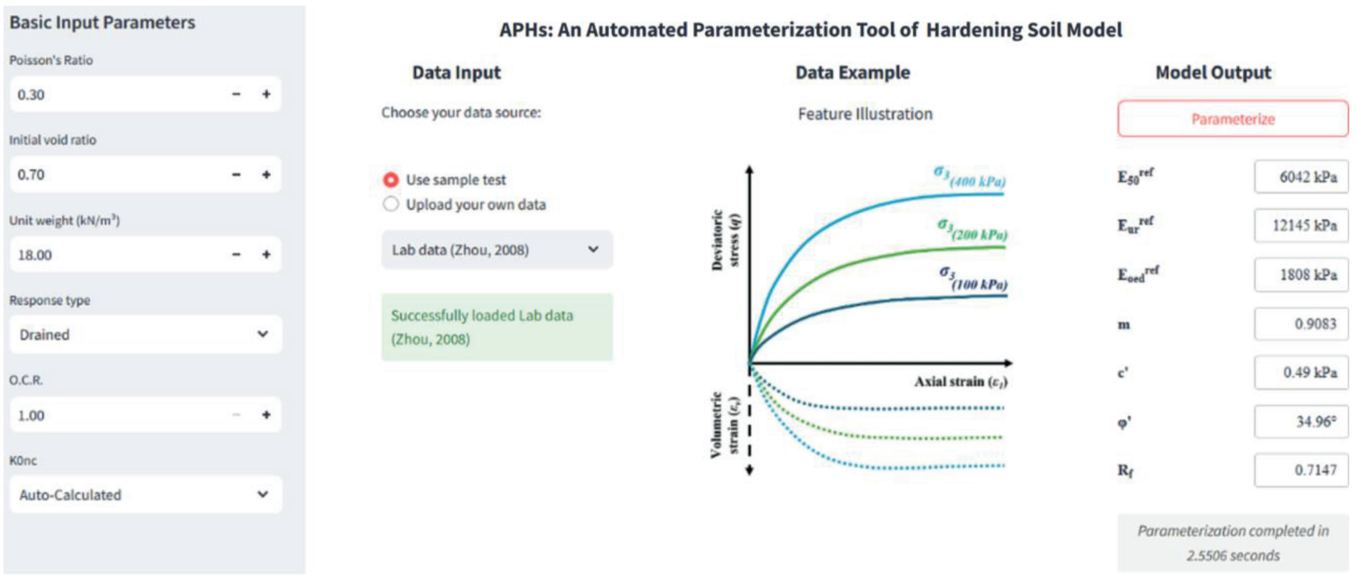


Fig. 19. The APHS tool developed in this study.

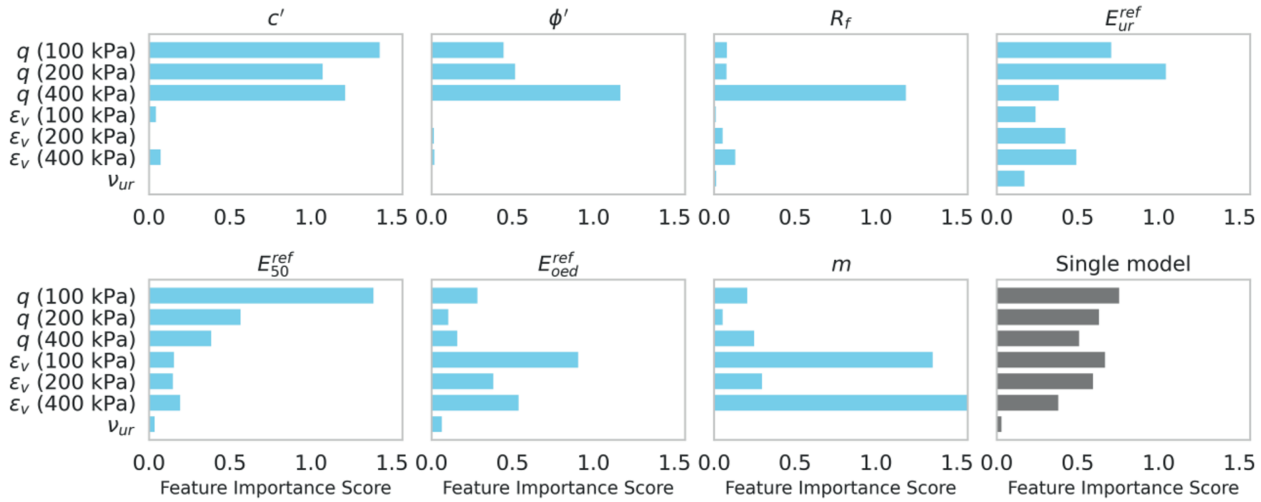


Fig. 20. Permutation feature importance: Multi-model (blue) vs. single model (gray).

Table 5
The model evaluation metrics (NMAE, MAPE, R^2, and MAE).

Parameter	NMAE		MAPE		R^2		MAE	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
c'	0.66	0.74	–	–	0.9991	0.9987	0.052 kPa	0.085 kPa
ϕ'	0.24	0.22	0.05	0.05	0.9999	0.9999	0.017°	0.015°
R_f	0.83	0.84	0.56	0.55	0.9984	0.9985	0.004	0.004
E_{50}^{ref}	0.9	1.23	1.28	1.59	0.9982	0.9958	361.512 kPa	490.832 kPa
	0.7	0.83	2.09	2.18	0.9976	0.9969	156.208 kPa	188.001 kPa
E_{oed}^{ref}	0.39	0.42	1.25	1.27	0.9994	0.9991	65.006 kPa	67.919 kPa
m	0.25	0.24	0.43	0.4	0.9998	0.9999	0.002	0.002
Average	0.567	0.646	0.943	1.007	0.9989	0.9984	–	–

Table 6
Comparison of the APHS tool and the conventional method for parametrizing HS model.

Method	c' (kPa)	ϕ' (°)	R_f	E_{ur}^{ref} (kPa)	E_{50}^{ref} (kPa)	E_{oed}^{ref} (kPa)	m
GEOSLOPE	0	29.6	0.941	45000	17745	11500	0.68
APHS	0.041	29.589	0.9359	45361.88	17849.41	11528.095	0.682
MAE	0.041	0.011	0.0051	361.88	104.41	28.095	0.002
MAPE (%)	–	0.037	0.54	0.80	0.588	0.24	0.294

Table 7
HS model parameters with 99 % CI lower and upper bounds using the APHS tool.

Method	c' (kPa)	ϕ' (°)	R_f	E_{ur}^{ref} (kPa)	E_{50}^{ref} (kPa)	E_{oed}^{ref} (kPa)	m
APHS	0.493	34.960	0.715	12,145.493	6041.435	1807.876	0.908
99 % CI lower	0.306	34.907	0.702	10,815.303	5326.814	1515.862	0.902
99 % CI upper	0.682	35.014	0.727	13,475.684	6756.055	2099.89	0.915

Table 8
The single model vs. multi-model evaluation metrics: MAE.

Case	MAE values						
	c' (kPa)	ϕ' (°)	R_f	E_{ur}^{ref} (kPa)	E_{50}^{ref} (kPa)	E_{oed}^{ref} (kPa)	m
Training set	0.052 (0.151)	0.017 (0.073)	0.004 (0.005)	361.512 (557.42)	156.208 (204.77)	65.006 (88.22)	0.002 (0.009)
Validation set	0.085 (0.152)	0.015 (0.071)	0.004 (0.005)	490.832 (575.87)	188.001 (206.83)	67.919 (87.16)	0.002 (0.009)
Testing set (stochastic dropout)	0.071 (0.174)	0.015 (0.075)	0.005 (0.011)	380.38 (1016.57)	308.6 (393.42)	48.34 (96.73)	0.002 (0.009)
Test case (GEOSLOPE International Ltd., 2021)	0.041 (0.324)	0.011 (0.026)	0.005 (0.019)	361.88 (768.02)	104.41 (1477.93)	28.095 (357.83)	0.002 (0.017)

Note: Values in parentheses correspond to the single LSTM model.

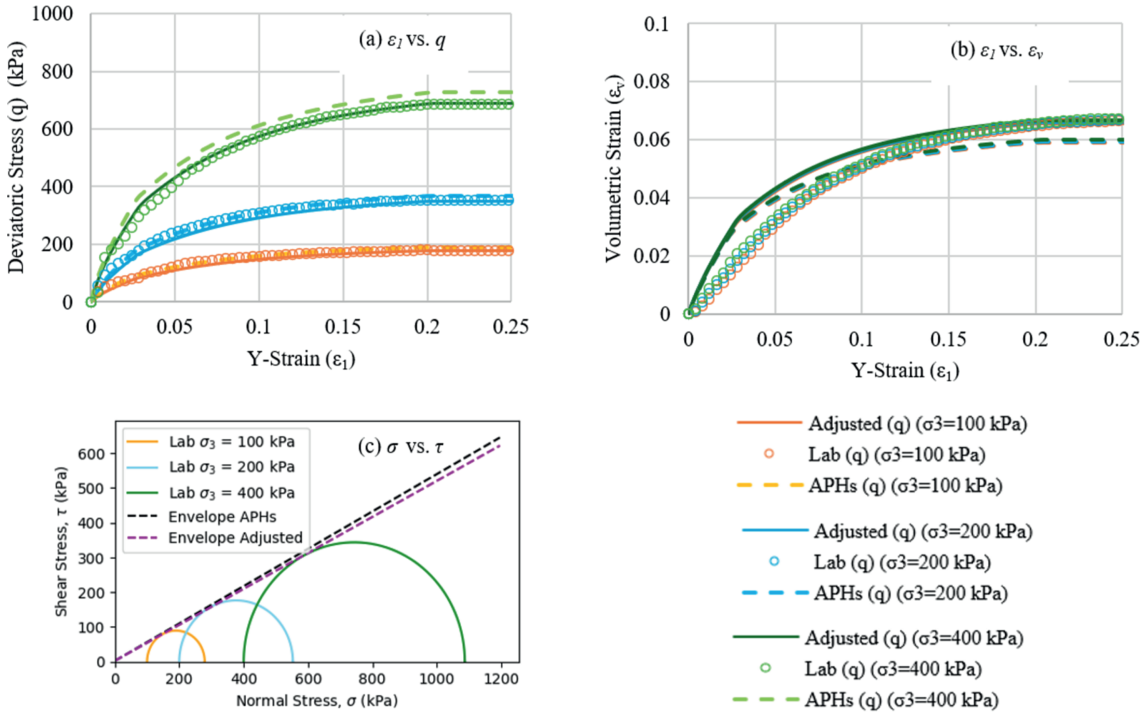


Fig. 21. Comparison of simulated results: APHS vs. laboratory test results (Jin et al., 2016).

Table 9
HS model parameters for Hostun sand with 99 % CI using the APHS tool.

Method	c' (kPa)	ϕ' (°)	R_f	E_{ur}^{ref} (kPa)	E_{50}^{ref} (kPa)	E_{oed}^{ref} (kPa)	m
APHS	2.5026	28.2405 (27.3671)*	0.8551	10,692.4502	4591.1729	1124.23 (1000)*	0.9834
99 % CI lower	2.3297	28.191	0.8422	9483.7782	4057.8961	911.133	0.9769
99 % CI upper	2.6755	28.2901	0.8679	11,901.1222	5124.4496	1337.3333	0.9899

Note: Values in parentheses are adjustments based on controlling features (Fig. 20) and visualized deviations (Fig. 21). The symbol "*" indicates that the data fall outside the trained boundaries of the APHS tool.

characteristics and sensitivity of the problem under investigation.

One of the most notable achievements of this approach is its ability to accurately estimate stiffness parameters without requiring oedometer tests, making it an efficient tool. Notably, the parameters retrieved by the APHS tool for CDG-recompacted saturated soil were found to be $E_{oed}^{ref} = 0.3E_{50}^{ref}$, which is found to fall outside the augmented samples for which the tool was developed and even beyond the ratios proposed in the literature. For example, Lu et al. (2018), Lee et al. (2008), and Lam (2018) suggested a ratio between 0.7 and 1, while GEOSLOPE International Ltd., (2021) recommended a range of 0.5–0.8 in cases where oedometer tests are unavailable. Comparable results were observed for Hostun sand. This demonstrates the robust performance of the APHS tool in accurately determining the HS parameters.

The results demonstrate the performance of the APHS tool, achieving an accuracy of 99.4 %. Scatter plots of predicted vs. actual values highlight the tool's ability to generalize effectively, with the predicted values aligning closely with the 1:1 line for both training and validation samples. The performance of the APHS tool was tested, validated, and deployed using unseen datasets to confirm its reliability and precision. The first validation reference is a soil sample previously parameterized by GEOSLOPE International Ltd., (2021). The comparison demonstrated that the APHS tool achieved an accuracy of approximately 99.6 %, with minimal absolute error across all predicted parameters. The second validation case focused on a CDG-recompacted saturated soil. The simulated results using parameters retrieved from the APHS tool were favorably comparable to the original laboratory tests (Zhou, 2008). This study validates the initial hypotheses proposed at the outset. One of the key implications of this research is the development of a smart laboratory system. This work paves the way for transforming traditional into innovative laboratories, leveraging the power of DL to enhance efficiency and accuracy. APHS can be locally installed on any device with the required libraries to deliver parameterizations within seconds.

4.6. Model computational burden and future directions

While our results demonstrate that APHS yields accurate predictions, the reader needs to note the considerable computational power required. Essentially, prediction performance relies on two factors: (1) the data's volume and quality and (2) the DL algorithm's capability, both of which impact computational requirements during training and for end users.

First, comprehensive laboratory datasets with accurately measured parameters are scarce; for instance, according to the authors' knowledge, one of the minor parametrized cases available is the one provided by GEOSLOPE International Ltd., (2021). Conducting laboratory tests usually requires extensive manual parameterization, which is time-consuming, highlighting the advantage of our proposed method in handling the sheer volume

Table 10
Summary of current achievements and future direction.

Aspect	Current setup	Current achievement	Future directions	Target improvement
Dataset volume	A comprehensive dataset of 1300 samples	High accuracy	Exploring alternative sampling techniques (e.g. Sobol sequences, MC)	Reduction of dataset size while maintaining high accuracy for further efficiency
Dataset quality	102 time steps represent each curve	(1) High accuracy; (2) Mitigate laboratory data bias	Explore transformers/CNNs	Achievement of faster training with fewer trainable parameters while capturing long-term behavior
Dataset generation	Based on 2D-axisymmetric simulation assumptions	Testing on real-world samples confirms the validity of the 2D axisymmetric simulation assumptions	Considering 3D simulations	Enhancement of realism and robustness with further training as needed
Tool versatility	Tolerated for Hong Kong soil, as outlined in Section 3	(1) High accuracy; (2) User interface; (3) Fast parametrizations; (4) Integrable with any finite element models or physically informed tools	Incorporating anisotropy, dynamic loading conditions, other constitutive models, and wider parameter ranges	Broadening applicability and versatility to different soils and conditions

of data needed. However, there remains potential to further reduce the required sample size by experimenting with different sample sizes and alternative sampling strategies such as Sobol sequences or MC methods.

Secondly, we acknowledge that individual laboratory tests may introduce bias regarding the dataset's quality. To mitigate this, we ensured that 102 time steps represent each curve. However, this level of detail means that each sample (comprising six curves, 102 time steps per curve, and seven parameters) demands significant computational resources during training. While the LSTM has proven effective in delivering accurate results, its computational requirements are substantial. Therefore, exploring alternative DL architectures, such as transformer networks and CNNs, could potentially reduce the computational burden without compromising performance.

Thirdly, regarding the computational burden, our experiments (using a single model as an example) illustrate this challenge well. For instance, the first LSTM layer, which processes seven input features, incorporates roughly 1,064,960 trainable parameters (calculated as $4 \times (7 + 512 + 1) \times 512$). Both the second and third LSTM layers require approximately 2,099,200 parameters each. In addition, a 256-unit dense layer adds about 131,328 parameters, and the final 7-unit output layer contributes another 1799 parameters, resulting in a total of roughly 5.4 million trainable parameters. The sequential processing across all 102 time steps further amplifies memory and time requirements. To clarify, model tuning on the HPC server takes about 6 h. In contrast, tuning on a regular PC (8 CPU cores, 128 GB memory, with no GPU) could extend the duration to several days. For example, 20 iterations with 250 epochs required approximately 2 days, reflecting the substantial computational resources required.

Our findings are intended to serve as a foundation, opening avenues for further future efforts toward extensive dataset generation and further optimization of the computational burden. Table 10 summarizes the study's current achievements and future directions.

5. Conclusions

This study introduces a transformative approach to soil parameterization by developing the APHS tool, leveraging the power of LSTM networks to achieve unprecedented accuracy and efficiency in estimating critical strength and stiffness parameters (c', ϕ', R_f, E_{oed}^{ref}, E_{50}^{ref}, E_{ur}^{ref}, and m). Key findings include:

(1) With a model accuracy of 99.4 %, the APHS tool demonstrates precision in predicting parameters solely from triaxial test data, making the process more accessible and cost-effective. Validation cases affirm the tool's reliability, with predicted parameters aligning with laboratory results. Notably, for CDG-recompacted soil, the tool accurately retrieved parameters even when they fell outside the range of ratios traditionally proposed in the literature, showcasing that this tool not only accelerates the HS model parameterization process but also simplifies it while maintaining high precision.
(2) The aforementioned outstanding performance stems from two key advancements. Firstly, the study employed advanced data generation techniques using a combination of the LHS and SIGMA/W simulations to address the limitations of insufficient experimental data. This approach created a robust and comprehensive parameter space tailored to Hong Kong's unique geological conditions, ensuring the tool's practical applicability. Secondly, the tool's innovative multi-model architecture enabled the assignment of distinct weights to each parameter, effectively mimicking the physical processes involved in parameter derivation and ensuring robustness across varying scales. By framing stress–strain relationships as a time-series problem, the tool leveraged the sequential nature of triaxial test results to enable precise and consistent parameter estimation.
(3) This research establishes a paradigm shift in geotechnical engineering, paving the way for smart laboratory systems that integrate DL to optimize and automate traditionally manual processes. While current findings validate the tool's performance under defined conditions, future work will focus on expanding its parameter space to accommodate a broader spectrum of soil types and boundary conditions.

CRediT authorship contribution statement

Kyrillos Ebrahim: Methodology, Data curation, Writing – original draft, Investigation, Conceptualization, Validation, Formal analysis. **Tarek Zayed:** Supervision, Writing – review & editing, Resources, Conceptualization, Validation, Methodology, Investigation. **Ridwan Taiwo:** Data curation, Writing – review & editing, Visualization. **Ashraf El-Hamalawi:** Conceptualization, Writing – review & editing, Formal analysis.

Availability of data and materials

The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used GPT 4o, Copilot, DeepSeek, and Grammarly for code debugging, rephrasing, and checking grammar and spelling. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the support from the Innovation and Technology Support Programme (ITSP) under grant number ITS/033/20FP, and the Water Supplies Department of the Hong Kong Special Administrative Region, China. The authors would also like to acknowledge the open-source files provided by GEOSLOPE International Ltd. (2021) and the valuable insights shared in the webinar series on the Hardening Soil Model, available at https://www.seequent.com/sigma-w-material-model-series-hardening-soil-model/. The authors also extend their gratitude to the Departments of Building and Real Estate and Civil and Environmental Engineering at The Hong Kong Polytechnic University for providing valuable resources to support this study.

Abbreviations and notations

HS	Hardening
APHS	Automatic parametrization of HS
CDG	Completely decomposed granite
CI	Confidence interval
HPC	High-performance computing
LHS	Latin hypercube sampling
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MC	Monte Carlo
MSE	Mean squared error
NMAE	Normalized mean absolute error
c'	Effective cohesion
ϕ'	Effective internal friction angle
R_f	Failure ratio
E_{50}^{ref}	Secant stiffness in drained triaxial tests
E_{ur}^{ref}	Unloading/reloading stiffness
m	Power for the stress level dependency of stiffness
q	Deviatoric stress
ε_1	Axial strain
ε_v	Volumetric strain
σ_3	Lateral stress
R^2	Coefficient of determination

References

Abadi, M., Agarwal, A., Barham, P., et al., 2016. TensorFlow: large-scale machine learning on heterogeneous systems. https://doi.org/10.48550/arXiv.1603.04467.

Atkinson, J., 2007. The Mechanics of Soils and Foundations, second ed. CRC Press, London, UK.

Bangun, R., Krisnamurti, Indra, N., 2020. Analysis of slope stability in soft soil using hardening soil modeling and strengthening of bamboo mattress. Int. J. GEOMATE 19 (73), 226–234.

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures. In: Montavon, G., Orr, G.B., Müller, K.R. (Eds.), Neural Networks: Tricks of the Trade. Springer, Berlin, Heidelberg, Germany.

Bhadiyadra, K., Ong, D.E.L., 2024. Mechanics of rainfall-induced landslides after a prolonged dry period based on laboratory tests and numerical models incorporating soil–water characteristic curves. Geosci. J. 14 (7), 174.

Brinkgreve, R.B.J., 2019. Automated model and parameter selection: incorporating expert input into geotechnical analyses. GeoStrata Mag. Arch. 23 (1), 38–45.

Brosz, F., Machaček, J., Zachert, H., 2024. Automatic parameter calibration of advanced constitutive soil models and application to a boundary value problem. In: Geotechnical Engineering Challenges to Meet Current and Emerging Needs of Society. CRC Press, London, UK, pp. 1–6.

Bui, L.M., Wu, L., Cheng, Y., Dong, D.J., 2024. Evaluation of using hardening soil model for predicting wall deflections caused by deep excavation: a case study at the Ho Chi Minh metro line 1, Vietnam. Arch. Civ. Eng. 70 (1), 357–373.

Chan, C., Chiu, D., Lo, F., Kwan, J., Lee, S., Leung, A., 2021. Back analyses of two deep excavations in Hong Kong using the mohr-coulomb model with linear elasticity and the hardening soil model. Int. J. Geoeng. Case Hist. 7 (1), 137–163.

Di Fiore, F., Nardelli, M., Mainini, L., 2024. Active learning and Bayesian optimization: a unified perspective to learn with a goal. Arch. Comput. Methods Eng. 31 (5), 2985–3013.

Dong, Y., Burd, H., Houlsby, G., 2016. Finite-element analysis of a deep excavation case history. Geotechnique 66 (1), 1–15.

Ebrahim, K.M.P., Fares, A., Faris, N., Zayed, T., 2024. Exploring time series models for landslide prediction: a literature review. Geoenviron. Disasters 11 (1), 25.

Ebrahim, K., Taiwo, R., Zayed, T., 2025. OWHK: operational volumetric water content forecasting model for shallow rainfall-induced landslides in Hong Kong. Eng. Geol., 108228, 355.

ETS-Testconsult Ltd., 2023. The Direct Shear test (Small Shear Box Apparatus). ETS-Testconsult Ltd., Hong Kong, China.

Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res. 20 (177), 1–81.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning, 48. PMLR), New York, USA, pp. 1050–1059.

GEOSLOPE International Ltd., 2020. Stress–Strain Modeling with Geostudio 2021. Technical Report. GEOSLOPE International Ltd., Calgary, Canada.

GEOSLOPE International Ltd., 2021. Parameterizing the Hardening Soil Model. Technical Report. GEOSLOPE International Ltd., Calgary, Canada.

GeoStudio, 2023. User's Guide for SIGMA/W (Version 2023). GeoStudio, Calgary, Canada.

GCO (Geotechnical Control Office), 1982. Mid-levels study: report on geology, hydrology and soil properties. GCO, Public Works Department of Hong Kong, Hong Kong, China.

Hanane, D., 2023. Evaluation metrics for regression models. https://machinelearning-basics.com.

Harris, C.R., Millman, K.J., van der Walt, S.J., et al., 2020. Array programming with NumPy. Nature 585 (7825), 357–362.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95.

Jin, Y.F., Yin, Z.Y., Shen, S.L., Hicher, P.Y., 2016. Selection of sand models and identification of parameters using an enhanced genetic algorithm. Int. J. Numer. Anal. Methods GeoMech. 40 (8), 1219–1240.

Khorasani, M., Abdou, M., Hernández Fernández, J., 2022. Streamlit at work. In: Web Application Development with Streamlit. Apress, Berkeley, USA, pp. 389–401.

Krachtoudi, C.Z., 2011. Numerical Simulation of the Response of Ottawa Sand in the Triaxial and the Torsional Shear Apparatus. National Technical University of Athens, Athens, Greece. MSc Thesis.

Lam, A.K.M., 2018. An engineering solution for a hillside project in Hong Kong. Geotech. Res. 5 (3), 170–181.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, S.W., Pickles, A.R., Henderson, T.O., Li, E.S.F., 2008. Three-dimensional modelling of deep excavation in decomposed granite: influence of small strain stiffness and presence of individual piles. In: Proceedings of the HKIE Geotechnical Division Annual Seminar 2008, Hong Kong, China, pp. 171–180.

Li, K.Q., Yin, Z.Y., 2025. State of the art of coupled thermo–hydromechanical–chemical modelling for frozen soils. Arch. Comput. Methods Eng. 32 (2), 1039–1096.

Li, K.Q., Yin, Z.Y., Zhang, N., Li, J., 2024. A PINN-based modelling approach for hydromechanical behaviour of unsaturated expansive soils. Comput. Geotech. 169, 106174.

Liu, H.C., Zhang, N., Yin, Z.Y., Wang, Y., 2025. Interpretation of subsurface stratigraphic variations from limited boreholes using Dual Bi-LSTM. Can. Geotech. J. 62, 1–15.

Liu, M., Zhuang, P., Lai, F., 2024. A Bayesian optimization-genetic algorithm-based approach for automatic parameter calibration of soil models: application to

clay and sand model. Comput. Geotech. 176, 106717.
Liu, Z.Q., Guo, D., Lacasse, S., Li, J.H., Yang, B.B., Choi, J.C., 2020. Algorithms for intelligent prediction of landslide displacements. J. Zhejiang Univ. - Sci. 21 (6), 412–429.
Lu, K.K., Yin, J.H., Lo, S.C., 2018. Modeling small-strain behavior of Hong Kong CDG and its application to finite-element study of basement-raft footing. Int. J. GeoMech. 18 (9), 04018104.
Machaček, J., Staubach, P., Tavera, C.E.G., Wichtmann, T., Zachert, H., 2022. On the automatic parameter calibration of a hypoplastic soil model. Acta Geotech. 17 (11), 5253–5273.
Marzouk, I., Brinkgreve, R., Lengkeek, A., Tschuchnigg, F., 2024. APD: an automated parameter determination system based on in-situ tests. Comput. Geotech. 176, 106799.
Maunsell Geotechnical Services Ltd., 2007. Detailed study of the 21 August 2005 debris flow on the natural hillside near Fei Ngo Shan service reservoir. GEO Rep. 233. Geotechnical Engineering Office, Hong Kong, China.
McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.
Murray, C., Chaurasia, P., Hollywood, L., Coyle, D., 2022. A comparative analysis of state-of-the-art time series forecasting algorithms. In: International Conference on Computational Science and Computational Intelligence (CSCI). Las Vegas, USA, pp. 89–95.
Murray, C., Du Bois, N., Hollywood, L., Coyle, D., 2023. State-of-the-art deep learning models are superior for time series forecasting and are applied optimally with iterative prediction methods. SSRN Electron. J. https://doi.org/10.2139/ssrn.4361707.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7 (4), 308–313.
Ng, C.W.W., Shi, Q., 1998. A numerical investigation of the stability of unsaturated soil slopes subjected to transient seepage. Comput. Geotech. 22 (1), 1–28.
Ng, C.W.W., Sun, H.S., Lei, G.H., Shi, J.W., Mašín, D., 2015. Ability of three different soil constitutive models to predict a tunnel's response to basement excavation. Can. Geotech. J. 52 (11), 1685–1698.
Ng, C.W., Wang, B., Tung, Y.K., 2001. Three-dimensional numerical investigations of groundwater responses in an unsaturated slope subjected to various rainfall patterns. Can. Geotech. J. 38 (5), 1049–1062.
Nguyen, T., Truong, T.T., Nguyen-Thoi, T., Van Hong Bui, L., Nguyen, T.H., 2022. Evaluation of residual flexural strength of corroded reinforced concrete beams using convolutional long short-term memory neural networks. Structures 46, 899–912.
Nguyen-Minh, T., Bui-Ngoc, T., Shiau, J., Nguyen, T., Nguyen-Thoi, T., 2023. Coupling isogeometric analysis with deep learning for stability evaluation of rectangular tunnels. Tunn. Undergr. Space Technol. 140, 105330.
Nguyen-Minh, T., Bui-Ngoc, T., Shiau, J., Nguyen, T., Nguyen-Thoi, T., 2025. Hybrid deep learning and isogeometric analysis for bearing capacity assessment of sand over clay. J. Rock Mech. Geotech. Eng. 17 (8), 5240–5265.
O'Malley, T., Bursztein, E., Long, J., et al., 2019. KerasTuner. GitHub. San Francisco, USA. https://github.com/keras-team/keras-tuner.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Qi, X., Wei, Y., Mei, X., Chellali, R., Yang, S., 2024. Comparative analysis of the linear regions in ReLU and LeakyReLU networks. In: Luo, B., Cheng, L., Wu, Z.G., Li, H., Li, C. (Eds.), Neural Information Processing: 30Th International Conference on Communications in Computer and Information Science (ICONIP 2023), 1962. Springer, Singapore, pp. 528–539.
Reback, J., McKinney, W., Brockmendel, J., et al., 2020. Pandas: powerful Python data analysis toolkit. https://pandas.pydata.org/.
Rohatgi, A., 2022. WebPlotDigitizer, Version 4.6. https://automeris.io/WebPlotDigitizer.
Schanz, T., Vermeer, P.A., Bonnier, P.G., 1999. The hardening soil model: formulation and verification. In: Beyond 2000 in Computational Geotechnics: Ten Years of PLAXIS International. Routledge, London, UK, pp. 281–296.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117.
Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. In: 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360). IEEE, New York, USA, pp. 69–73.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 60.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In: Proceedings of the 26th International Conference on Neural Information Processing Systems, 2. Curran Associates Inc., Red Hook, USA, pp. 2951–2959.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Storn, R., Price, K., 1997. Differential evolution – a Simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11 (4), 341–359.
Surarak, C., 2010. Geotechnical Aspects of the Bangkok MRT Blue Line Project. Griffith University, Nathan, Australia. PhD Thesis.
Surarak, C., Likitlersuang, S., Wanatowski, D., Balasubramaniam, A., Oh, E., Guan, H., 2012. Stiffness and strength parameters for hardening soil model of soft and stiff Bangkok clays. Soils Found. 52 (4), 682–697.
The Hong Kong Polytechnic University HPC, 2025. Student HPC. https://studenthpc.polyu.edu.hk.
Truty, A., Obrzud, R., 2015. Improved formulation of the hardening soil model in the context of modeling the undrained behavior of cohesive soils. Studia Geotechnica Mech. 37 (2), 61–68.
Varangaonkar, P., Rode, S.V., 2023. Lightweight deep learning model for automatic landslide prediction and localization. Multimed. Tool. Appl. 82 (21), 33245–33266.
Wang, L., Xiao, T., Liu, S., Zhang, W., Yang, B., Chen, L., 2023. Quantification of model uncertainty and variability for landslide displacement prediction based on Monte Carlo simulation. Gondwana Res. 123, 27–40.
Wang, Z., 2022. Comparative study of Latin hypercube sampling and Monte Carlo method in structural reliability analysis. Highl. Sci. Eng. Technol. 28, 61–69.
Waskom, M.L., 2021. Seaborn: statistical data visualization. J. Open Source Softw. 6 (60), 3021.
Wolpert, D.H., 1996. The lack of a priori distinctions between learning algorithms. Neural Comput. 8 (7), 1341–1390.
Wong, H.Y., 2020. Shear strength properties of Hong Kong soils for slope stability. HKIE Trans. 27 (1), 48–54.
Wong, J., Gupta, A., Ooi, C., Chiu, P.H., Liu, J., Ong, Y.S., 2025. Physics-informed neuro-evolution (PINE): a survey and prospects. https://arxiv.org/html/2501.06572v1.
Xing, Y., Yue, J., Chen, C., 2019. Interval estimation of landslide displacement prediction based on time series decomposition and long short-term memory network. IEEE Access 8, 3187–3196.
Xu, H.R., Yin, J.N., Zhang, N., 2025a. Transformer-based deformation measurement of underground structures from a single-camera video. Autom. Constr. 172, 106070.
Xu, K., Zhang, N., Yin, Z.Y., Li, K., 2025b. Finite element-integrated neural network for inverse analysis of elastic and elastoplastic boundary value problems. Comput. Methods Appl. Mech. Eng. 436, 117695.
Zhang, N., Shen, S.L., Zhou, A., Jin, Y.F., 2021a. Application of LSTM approach for modelling stress–strain behaviour of soil. Appl. Soft Comput. 100, 106959.
Zhang, N., Xu, K., Yu Yin, Z., Li, K.Q., Jin, Y.F., 2025. Finite element-integrated neural network framework for elastic and elastoplastic solids. Comput. Methods Appl. Mech. Eng. 433, 117474.
Zhang, P., Jin, Y.F., Yin, Z.Y., 2021b. Machine learning-based uncertainty modelling of mechanical properties of soft clays relating to time-dependent behavior and its application. Int. J. Numer. Anal. Methods GeoMech. 45 (11), 1588–1602.
Zhang, P., Yin, Z.Y., Jin, Y.F., Ye, G.L., 2020. An AI-based model for describing cyclic characteristics of granular materials. Int. J. Numer. Anal. Methods GeoMech. 44 (9), 1315–1335.
Zhou, W.G., 2008. Experimental and Theoretical Study on Pullout Resistance of Grouted Soil Nails. The Hong Kong Polytechnic University, Hong Kong, China. PhD Thesis.

Kyrillos Ebrahim earned a BSc degree in Civil Engineering (2017) with first-class honors and an MSc degree in Structural Engineering (2022) from Mansoura University, Egypt. From 2018 to 2023, he worked at Mansoura University's Structural Engineering Department, serving as a Research Assistant, Quality Manager, and Senior Geotechnical Engineer, and contributed to the lab's ISO/IEC 17025:2017 accreditation process. He is currently a PhD candidate at the Department of Building and Real Estate (BRE) at the Hong Kong Polytechnic University and is a visiting research exchange student at Loughborough University in the UK. His research encompasses deep learning applications in geotechnical engineering, landslide prediction, soil–structure interactions, and numerical analysis. He is an alumnus of the LARAM International School on Landslide Risk Management (2025).