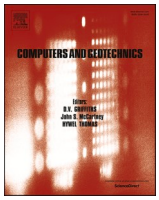
Computers and Geotechnics 190 (2026) 107761

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo



Research Paper

A FEM-PINN approach to modelling elastoplastic soil behaviour in boundary value problems

Mingpeng Liu [a,*], Qinghua Zhang [b], Raul Fuentes [a]

[a] Institute of Geomechanics and Underground Technology, RWTH Aachen University, Germany
[b] School of Qilu Transportation, Shandong University, China

ARTICLE INFO

Keywords:
Physics-informed neural network
Elastoplastic soil
Constitutive modelling
Finite element-neural network coupling
Boundary value problems

ABSTRACT

This study presents a physics-informed neural network (PINN) framework for modelling the elastoplastic behaviour of soils and its integration into the finite element method (FEM). The network jointly predicts stress, void ratio, and plastic strain, while the incremental strain decomposition is imposed as a physics-informed constraint in the loss function. This avoids explicit plastic yield functions or hardening rules and ensures compatibility with measurable laboratory quantities. The void ratio further defines a state-dependent elastic stiffness tensor, enabling a fully implicit FEM–PINN coupling in which element stiffness is updated at every Newton iteration. Compared with a multilayer perceptron (MLP) baseline, the PINN achieves superior accuracy, stability, and interpretability. The framework is validated on three unseen boundary value problems—biaxial compression, cavity expansion–contraction, and foundation loading—where it reproduces benchmark solutions with excellent agreement in stress, strain, and void ratio. It also captures stress concentration and unloading–reloading responses. Overall, the proposed FEM–PINN framework provides a robust and interpretable alternative to conventional soil constitutive models, combining the flexibility of data-driven learning with physics-based constraints to enable reliable analysis of complex geotechnical problems.

1. Introduction

Soil presents complex mechanical behaviour, such as stress or strain path dependency, dilatancy, state variables-dependency, and critical state characteristics, to name but a few (Wu et al., 2023). In an attempt to predict soil behaviour, constitutive models are used. e.g., the Modified Cam-clay model (Roscoe and Burland, 1968) and the Clay And Sand Model (Yu, 1998). Among those, elastoplastic models using the critical state concept have achieved great popularity since their inception (Liu et al., 2024b; Zhang et al., 2021). They follow a well-understood philosophy and methodology that can be explained in relatively simple terms mathematically and this enables seamless implementation in numerical frameworks. The finite element method (FEM) combined with constitutive models remains the most widespread tool to predict soil behaviour in engineering boundary value problems (BVPs) (Lai et al., 2025a; 2025b).

Over the past few years, the successful application of neural networks has been an alternative to soil constitutive modelling (Guan and Yang, 2023; Wang et al., 2024). Some data-driven models have been developed to represent the mechanical behaviour of soil by directly learning the stress–strain relationships from simulations or measurements (Eghbalian et al., 2023; Mozaffar et al., 2019). Neural networks can replace conventional constitutive modelling, including the potential to learn complex path-dependent behaviour when appropriate state variables are used (Mozaffar et al., 2019). From these developments, neural networks have been recently proposed as surrogate models and used in FEM. Zhang et al. (2020; 2022a; 2022b) have investigated the application of the long short-term memory (LSTM) neural network in modelling complex history-dependent soil constitutive behaviours as well as incorporating it into finite elements for different BVPs, such as biaxial tests and foundations with internal erosion. Guan et al. (2023) successfully applied several neural networks to learn the stress–strain relationship and to model cavity expansion and a strip footing. However, such pure data-driven models also come with some drawbacks. They usually require a large amount of training data and have poor extrapolation ability (Eghbalian et al., 2023). Many argue, this is rooted in the lack of proper physical constraining of the output (Eghbalian et al., 2023; Haghighat et al., 2023).

These drawbacks have led some scholars to develop physical-based neural networks that enforce the outputs to comply with a framework

* Corresponding author.
E-mail address: liu@gut.rwth-aachen.de (M. Liu).

https://doi.org/10.1016/j.compgeo.2025.107761
Received 14 May 2025; Received in revised form 6 October 2025; Accepted 7 November 2025
Available online 12 November 2025

of physical equations (Haghighat et al., 2021; Zhang et al., 2024). Physics-Informed Neural Networks (PINNs) can theoretically be applied to any given physical process described by partial differential equations (PDEs) (Raissi et al., 2019). The PINN endows the outputs of neural networks with a clear physical meaning and constrains them into a framework of physical relationships, thereby improving the prediction and, critically, the extrapolability accuracy of models (Meng and Karniadakis, 2020; Vahab et al., 2022).

Recently, some scholars extended the restriction of PDEs in PINN to more general respecting physical-based laws, including physic-based postulates in constitutive modelling that are not necessarily expressed in terms of PDEs (Haghighat et al., 2023; Vahab et al., 2022). In fact, PINN has been applied to solid mechanics (Haghighat et al., 2021), fluid mechanics (Cai et al., 2021), thermodynamics (Masi and Stefanou, 2022), and elastoplastic mechanics (Haghighat et al., 2023), to name but a few. Linka and Kuhl (2023) designed the constitutive artificial neural networks that inherently satisfy common kinematical, thermodynamical, and physical constraints, in which the network weights have a clear physical interpretation. Masi and Einav (2024) developed the Neural Integration for Constitutive Equations (NICE) by replacing the evolution equations and internal energy function with two neural operators, in order to discover constitutive models at the material point level from scarce and incomplete observations. Su et al. (2024) used the plastic strain as the output and constrained it into elastoplastic constitutive modelling, providing the true value. However, such models usually consider solely the stress and strain variables to improve prediction accuracy, but discard some important information in soils, e.g., plastic strain or void ratio (Zhang and Mohr, 2020). Even in those incorporating soil state variables (Eghbalian et al., 2023; Su et al., 2024), they use the true value for plastic strain as part of the learning process: however, this is difficult to measure, thereby lowering the practicality and explainability of their models.

For soils, physical variables, such as void ratio and plastic strain, are significant in the loading process, as they 1) provide physical information of soils, 2) are necessary for calculating the stress–strain relationship and plastic dissipation, and 3) participate in the formation of material stiffness matrix when used in FEM. However, many of data-driven or PINN models (Haghighat et al., 2023) neglect this role. Subsequently, neural networks usually ignore some physical or plasticity information in replacing constitutive model, lowering their applicability. Additionally, in the integration with FEM, the material stiffness matrix is either a simple constant form or just omitted using explicit algorithms (Zhang et al., 2022a). The use of an implicit algorithm where the material stiffness matrix is assembled using PINN has not been yet explored.

Hence, compared to neural-network constitutive models integrated into finite element solvers (e.g., Hashash et al. (2004); Kirchdoerfer and Ortiz (2016)), the present framework differs in that the network jointly predicts stress, void ratio and plastic strain. The predicted void ratio is then used to compute the state-dependent elastic stiffness tensor, which enter both the physics-informed loss (through elastic–plastic strain decomposition) and the finite element stiffness assembly in the Newton iterations. Likewise, the predicted plastic strain closes the strain decomposition. Compared to stress-only surrogates commonly reported in the literature, this design provides the following contributions:

(i) provides state consistency and interpretability, as e and σ are measurable quantities that can also be easily obtained in the laboratory.
(ii) a novel approach in PINNs to calculate ε^p as output combining a simplified elastic stiffness tensor and incremental strain decomposition. This avoids assumptions on plastic yield or hardening.
(iii) implements a fully robust implicit FEM–PINN coupling where the global stiffness matrix is updated from PINN-predicted state each loading step, in contrast to most prior work that uses constant stiffness or explicit updates.

We position the contributions and paper overall as a state-aware and soil-specific synthesis that enhances interpretability, physical consistency, and solver stability relative to existing stress-only PINN approaches. Section 2 introduces the PINN-based constitutive modelling framework. Section 3 illustrates the development of the PINN, including the data generation, network architecture, and loss function. The PINN is then used to predict the soil constitutive behaviour and compared with deep neural networks. Section 4 incorporates the PINN into the FEM framework to enhance its practical application whilst comparing it to the benchmark in BVPs.

2. Methodology

2.1. Neural network-based constitutive modelling

For a strain-controlled constitutive model, the current stress state is dependent on the previous stress and historical strain (Zhang et al., 2023), and the physical state variables of soils (e.g., void ratio) (Zhang et al., 2022a). Following this logic, a neural network-based modelling strategy shown in Fig. 1 can be described as:

$$\left(\boldsymbol{\sigma}^{i+1}, \boldsymbol{J}^{i+1}\right) = \mathcal{NN}\left(\boldsymbol{\sigma}^{i}, \boldsymbol{\varepsilon}^{i}, \Delta\boldsymbol{\varepsilon}^{i}, \boldsymbol{J}^{i}\right) \tag{1}$$

where $\boldsymbol{\sigma}$, $\boldsymbol{\varepsilon}$, $\Delta\boldsymbol{\varepsilon}$, and $\boldsymbol{J}$ represent the stress tensor, strain tensor, strain increment tensor, and physical variables. Superscript "i" or "$i+1$" denotes the loading step. $\mathcal{NN}$ represents the neural network. Various deep neural network architectures, renowned for their capability to model complex nonlinear relationships between inputs and outputs, have been extensively developed for constitutive modelling. Notable examples include the multilayer perceptron (MLP) (Linka and Kuhl, 2023; Masi and Einav, 2024), the long short-term memory (LSTM) neural network (He et al., 2023; Zhang et al., 2022a), or some hybrid networks (Guan and Yang, 2023).

2.2. Classical elastoplastic model

In conventional elastoplastic models, the total strain rate is decomposed into elastic and plastic components, as expressed in Eq. (2):

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^{e} + \dot{\boldsymbol{\varepsilon}}^{p} \tag{2}$$

where superscript " ˙ " represents the rate of a variable; $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}^e$, and $\boldsymbol{\varepsilon}^p$ denote the total, elastic, and plastic strain tensor, respectively. The elastic strain can be solved using the general elasticity theory. However, the specification of plastic strains requires further consideration. Eq. (3) describes a conventional elastoplastic constitutive model:

$$\dot{\boldsymbol{\sigma}} = \boldsymbol{D}_e : \dot{\boldsymbol{\varepsilon}}^{e} = \boldsymbol{D}_e : \left(\dot{\boldsymbol{\varepsilon}} - \dot{\boldsymbol{\varepsilon}}^{p}\right) \tag{3.1}$$

$$f(\boldsymbol{\sigma}, \mathscr{H}) \leq 0 \tag{3.2}$$

$$\dot{\boldsymbol{\varepsilon}}^{p} = \dot{\Lambda}\frac{\partial g(\boldsymbol{\sigma}, \mathscr{H})}{\partial \boldsymbol{\sigma}} \tag{3.3}$$

$$\dot{\mathscr{H}} = H(\mathscr{H}, \boldsymbol{\sigma}, \boldsymbol{\varepsilon}, \dot{\boldsymbol{\sigma}}, \dot{\boldsymbol{\varepsilon}}, \boldsymbol{J}) \tag{3.4}$$

where Eq. (3.1) describes the general stress–strain relationship, in which $\boldsymbol{D}_e$ is the elastic stiffness tensor, $f(\boldsymbol{\sigma}, \mathscr{H})$ is the yield function or loading surface in which $\mathscr{H}$ denotes the hardening variables. Eq. (3.3) is the flow rule, where g denotes the plastic potential function and Λ is the plastic multiplier, and Eq. (3.4) is the hardening law H.

In the numerical implementation, these rate equations are transformed into the algebraic equations by time integral. For this reason, Eq. (3.1) can be written into incremental form (Zhou et al., 2022), as shown in Eq. (4), with a schema illustration of incremental strain decomposition in Fig. 2.

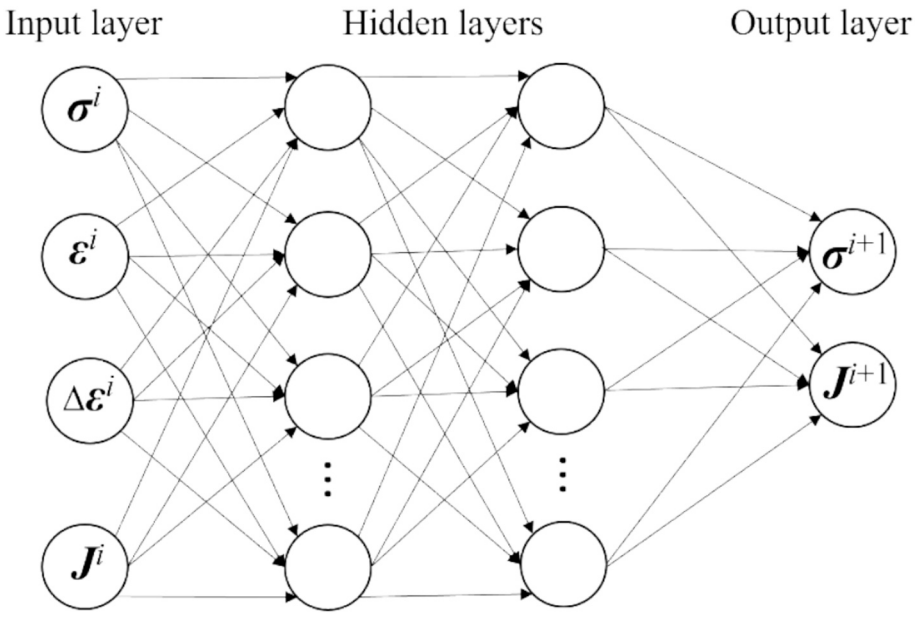


Fig. 1. Schema of neural network-based constitutive modelling.

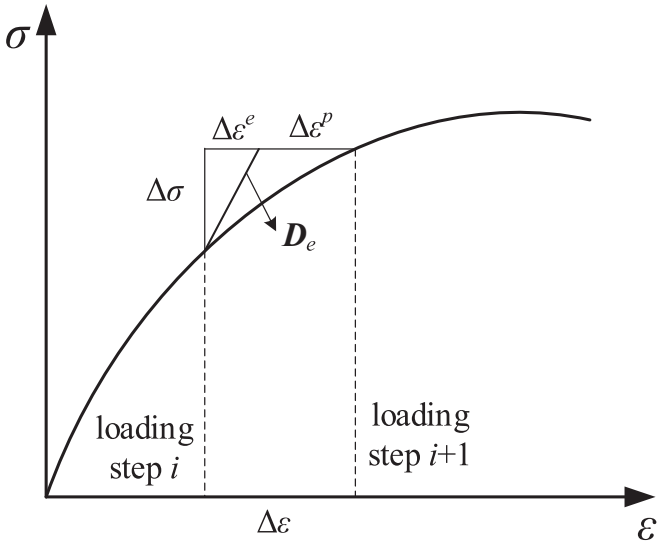


Fig. 2. Incremental stress–strain relationship.

$$\Delta\boldsymbol{\sigma} = \boldsymbol{D}_e : \Delta\boldsymbol{\varepsilon}^e = \boldsymbol{D}_e : (\Delta\boldsymbol{\varepsilon} - \Delta\boldsymbol{\varepsilon}^p) \tag{4}$$

Importantly, $\boldsymbol{D}_e$ is defined here as the state-dependent elastic stiffness tensor, i.e., the proportionality tensor between the elastic strain increment $\Delta\boldsymbol{\varepsilon}^e$ and the resulting stress increment $\Delta\sigma$ during a load step. In the general elasticity theory, $\boldsymbol{D}_e$ can be calculated using Eq. (5):

$$\boldsymbol{D}_e = \left(K - \frac{2}{3}G\right)\mathbf{I}\otimes\mathbf{I} + 2G\Pi \tag{5}$$

where K and G are the elastic bulk and shear moduli, respectively; $\mathbf{I}\otimes\mathbf{I}$ is the fourth order volumetric projector; $\mathbf{I}$ is the second order identity tensor; and Π is the fourth order deviatoric projection tensor. The moduli K and G are state dependant through e and calculated as (Yu, 1998):

$$K = \frac{1+e}{\kappa}p \tag{6.1}$$

$$G = \frac{3(1-2\mu)}{2(1+\mu)}K \tag{6.2}$$

where μ is the Poisson's ratio; κ is the slope of the swelling line; e is the void ratio; p is the mean effective stress defined as $p = \mathrm{tr}(\boldsymbol{\sigma})/3$. It should be noted that Eq. (6) is a concise but effective, representative, and widely-used formula to compute the elasticity. However, this formula is also a simplified expression.

2.3. Physical-informed neural network for elastoplastic soil

We designed the PINN framework in Fig. 3. The void ratio e and plastic strain tensor $\boldsymbol{\varepsilon}^p$ are chosen as the physical variables. Together with the stress tensor $\boldsymbol{\sigma}$, they are used as the output. The rationality of using e and $\boldsymbol{\varepsilon}^p$ as outputs are: (i) they are physically informative for soils; (ii) the outputs can be directly used to form the stiffness matrix; (iii) they close the stress–strain relationship used in the loss function, avoiding explicit yield/flow rules. In Fig. 3, the main body of the neural network is chosen as MLP as recommended by Roy and Guha (2023). Details of the MLP approach can be found in Liu et al. (2024a). With the current step i of stress, strain, strain increment, and physical variables ($\boldsymbol{\sigma}^i$, $\boldsymbol{\varepsilon}^i$, $\Delta\boldsymbol{\varepsilon}^i$, $\boldsymbol{J}^i$) at hand, the MLP is employed to predict the next step $i+1$ of stress and physical variables ($\boldsymbol{\sigma}^{i+1}$, $\boldsymbol{J}^{i+1}$). For this purpose, only the initial values of input are required to start the framework, while they can be updated in further predictions. Strain increments specifying the loading path should be preset as they can be provided by the FEM framework.

The predictions of $\boldsymbol{\sigma}$ and e can be trained using true values. The loss functions of $\boldsymbol{\sigma}$ and e are defined as the error between the predicted and true values, and they are quantities easily available from experiments. For the assessment of plastic strain that hard be directly measured in experiments, we use the elastoplastic stress–strain relationship to infer it:

$$\Delta\boldsymbol{\varepsilon}^i - \left(\boldsymbol{D}_e{}^i\right)^{-1} : \Delta\boldsymbol{\sigma}^i - \Delta\boldsymbol{\varepsilon}^{p,i} = 0 \tag{7}$$

Eq. (7) is obtained by rewriting Eq. (4), where $\Delta\boldsymbol{\sigma}^i = \boldsymbol{\sigma}^{i+1} - \boldsymbol{\sigma}^i$ and $\Delta\boldsymbol{\varepsilon}^{p,i} = \boldsymbol{\varepsilon}^{p,i+1} - \boldsymbol{\varepsilon}^{p,i}$. Hence, the complete loss function $\mathscr{L}$ consists of three parts – see Eq. (8).

$$\mathscr{L} = w_e\mathscr{L}_e + w_\sigma\mathscr{L}_\sigma + w_{\varepsilon^p}\mathscr{L}_{\varepsilon^p} \tag{8.1}$$

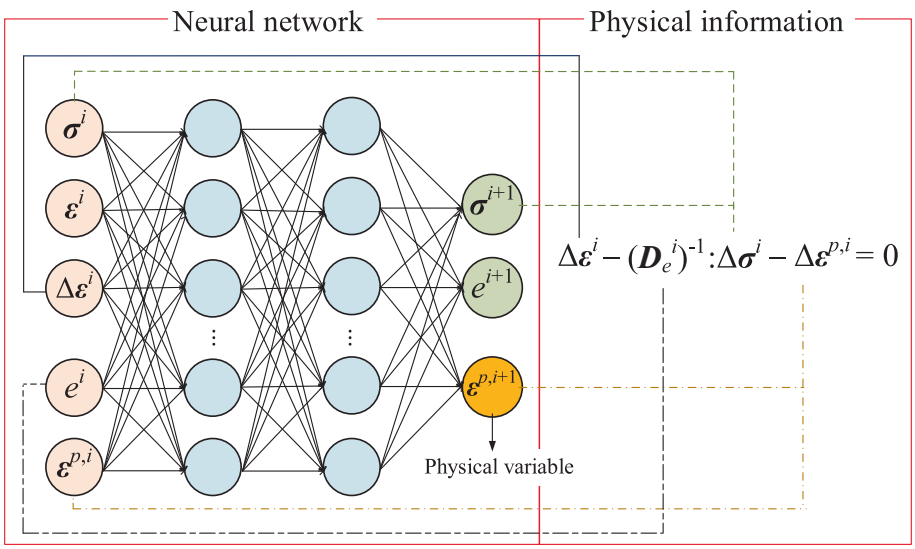


Fig. 3. The PINN for elastoplastic soil.

$$\mathscr{L}_e = \frac{1}{N_T}\sum_{i=1}^{N_T}\left(e_i - e_i^*\right)^2 \tag{8.2}$$

$$\mathscr{L}_{\boldsymbol{\sigma}} = \frac{1}{N_T}\sum_{i=1}^{N_T}\left(\boldsymbol{\sigma}_i - \boldsymbol{\sigma}_i^*\right)^2 \tag{8.3}$$

$$\mathscr{L}_{\boldsymbol{\varepsilon}^p} = \frac{1}{N_T}\sum_{i=1}^{N_T}\|\Delta\boldsymbol{\varepsilon}^i - \left(\boldsymbol{D}_e{}^i\right)^{-1} : \Delta\boldsymbol{\sigma}^i - \Delta\boldsymbol{\varepsilon}^{p,i}\|_2 \tag{8.4}$$

where $\mathscr{L}_e$, $\mathscr{L}_{\boldsymbol{\sigma}}$, and $\mathscr{L}_{\boldsymbol{\varepsilon}^p}$ represent the loss functions of void ratio, stress tensor, and plastic strain tensor, respectively. w_e, $w_{\boldsymbol{\sigma}}$, and $w_{\boldsymbol{\varepsilon}_p}$ denote the adaptive weights for scaling the values of each term by quantifying its relative importance (Zhang et al., 2025). $\mathscr{L}_e$ and $\mathscr{L}_{\boldsymbol{\sigma}}$ compute the error between the predicted and true values, while $\mathscr{L}_{\boldsymbol{\varepsilon}^p}$ denotes the error of the physical equation. N_T is the number of sample values. Superscript "*" refer to the true values of variable. The $\mathscr{L}_{\boldsymbol{\varepsilon}^p}$ loss enforces the physical equation by making the L_2 norm ($\|^{\circ}\|_2$) zero, see Eq. (8.4).

The loss function explores the removal of any explicit expression of plasticity. It is important to mention here that $\boldsymbol{\varepsilon}^p$ is inferred directly from Eq. (7) and no data is necessary. Only an initial value to start iterations is needed – typically 0. This design is intentional, since plastic strain is generally not observable in laboratory testing, and thus the approach remains fully compatible with experimental datasets. The only adapted assumption lies on the strain decomposition. In our case, the adaptive weights are all set to 1, indicating the equal contribution of each loss item for corresponding state variables.

3. Development of PINN

3.1. Data generation and processing

3.1.1. Benchmark – Bounding surface model

To investigate the capability of our approach, we selected the critical state bounding surface model (BSM), which is capable of capturing the complex behaviour of soils, as a benchmark. Seven material parameters are used in the model as λ, κ, μ, M, N, Γ, and n, details of which are given in Appendix. The bounding surface model was selected as the data generator because it can describe most of the complex behaviour of both sand and clay, provides plastic strains at every increment, and allows generation of large training datasets. We claim that this choice does not restrict the framework: any constitutive model or experimental dataset could serve as the data source.

The adopted constitutive parameters of the BSM are summarized in Table 1. The initialization parameters used are pre-consolidation pressure p_c, initial void ratio e_0, and initial stress state p_0. We define the value of p_c as 100 kPa and e_0 as 0.85, respectively. The p_0 is set to a confining pressure of 100 kPa to model an initially normally consolidated soil.

3.1.2. Random loading path

To consider complex strain-loading paths that the soil may experience, we use the material element point to generate strain paths. The random strain-loading path proposed by Mozaffar et al. (2019) is employed to generate training data in this study, which is also used by Zhang et al. (2022a) and He and Semnani (2023). For convenience and efficiency, we simplify the 3D problems to plane strain conditions, in which case the stress and strain components are (σ_{11}, σ_{22}, σ_{33}, σ_{12}) and (ε_{11}, ε_{22}, $\varepsilon_{33} = 0$, ε_{12}). Next, we explain how to generate the complex strain paths. Considering in all cases the soil starts from the initial state, the strain path should be initialized as $\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{12} = 0$, as shown in Fig. 4(a). Then, the strain path will evolve in the next 150 loading steps. For one strain path, three control points are set at 50, 100, and 150 loading steps and their strain values are randomly taken from the range of [-0.15, 0.15]. The strain-loading path is subsequently constructed by interpolating between control points using polynomial regression, and the same method to strain components ε_{11}, ε_{22}, and ε_{12}, as shown in Fig. 4(a). BSM produces the stress path (σ_{11}, σ_{22}, σ_{33}, σ_{12}) shown in Fig. 4(b). The stress–strain path are very complex cases, including strain direction reversals, in order to challenge the robustness and generalization of the PINN.

3.1.3. Data processing

Using BSM and this approach, 20 strain-loading paths are generated for training and testing, samples of which are shown in Fig. 5, where the stress and strain are shown by the isotropic and deviatoric tensors. It shows that adopting a random strain path guarantees a wide range of volumetric and shear strains, as well as their increments. Most of the generated volumetric strain as well as its increment are mainly distributed around zero, as shown in Fig. 5(b), (c), (d), and (e). Similarly, the generated mean effective stress p is mainly distributed around the initial

Table 1
Parameters of BSM.

λ	κ	μ	M	N	Γ	n
0.15	0.05	0.3	1.2	2.69	2.62	1.6

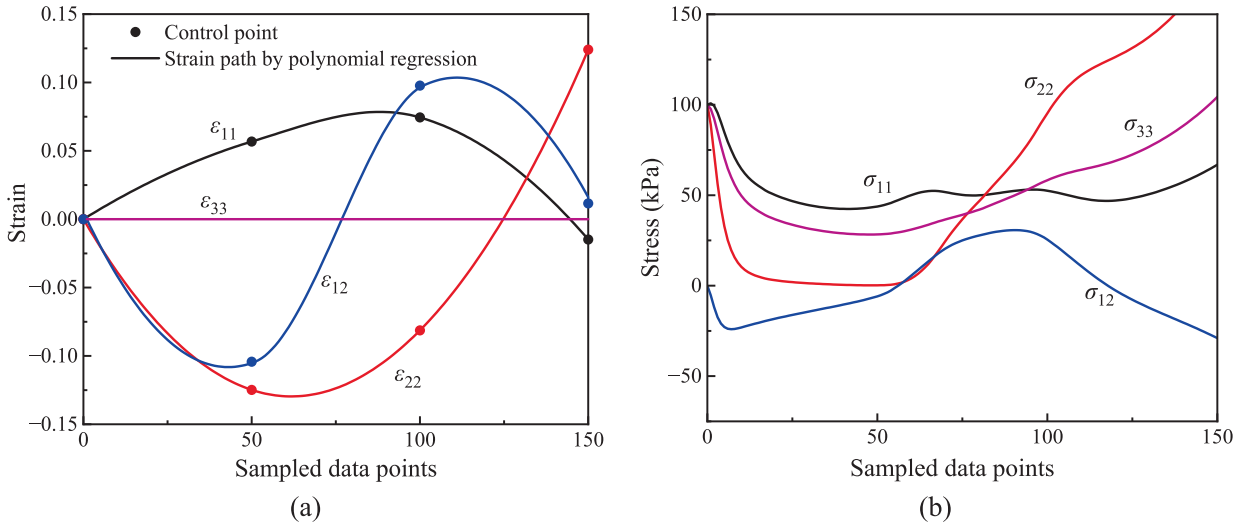


Fig. 4. Data sampling strategy for plane strain condition: (a) random strain-loading path, and (b) corresponding stress path.

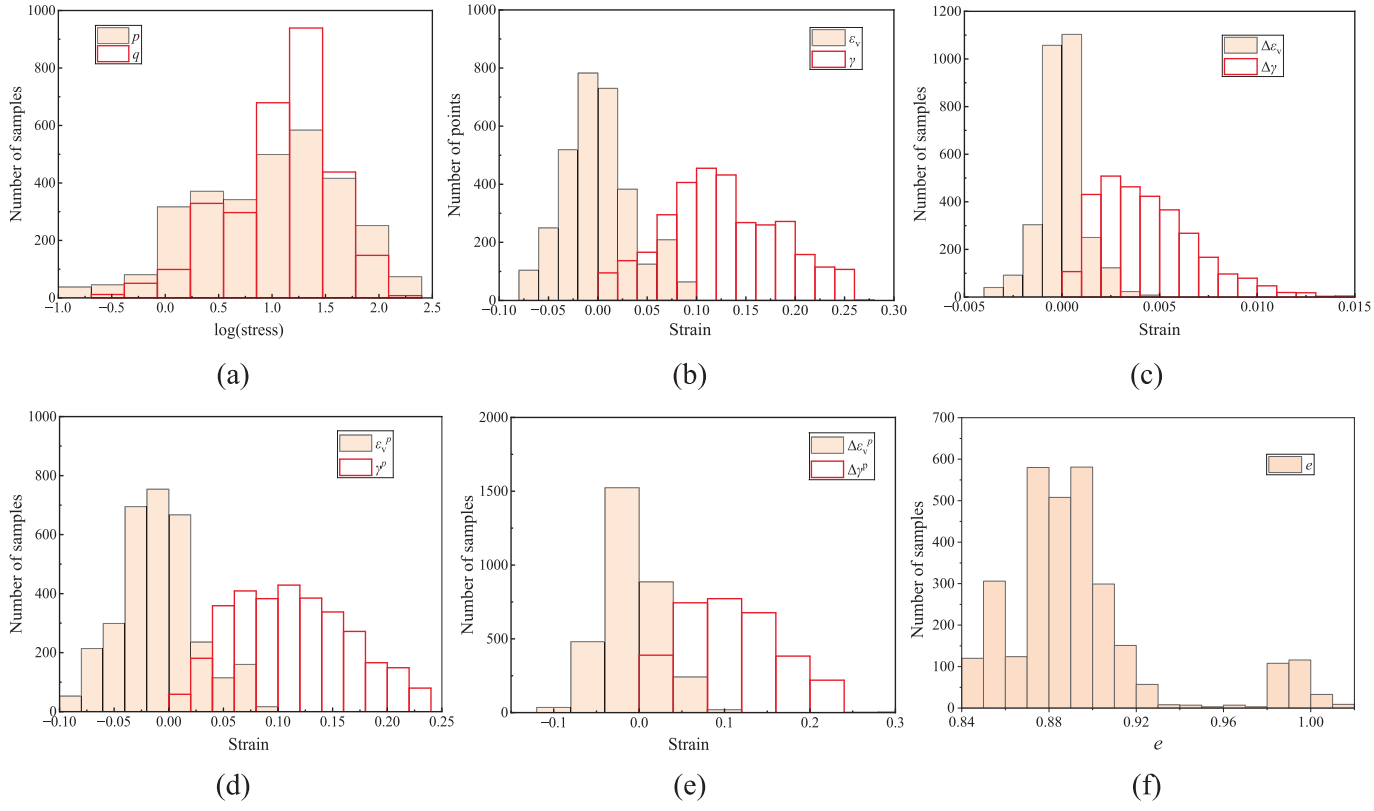


Fig. 5. Distribution of generated samples. (a) mean effective stress p and deviatoric stress q; (b) volumetric strain ε_v and shear strain γ; (c) volumetric strain increment $\Delta\varepsilon_v$ and shear strain increment $\Delta\gamma$; (d) plastic volumetric strain ε_v^p and plastic shear strain γ^p; (e) plastic volumetric strain increment $\Delta\varepsilon_v^p$ and plastic shear strain increment $\Delta\gamma^p$; (f) void ratio e.

confining pressure. The shear strain, its increment, and its plastic components have a wide range because of the continued change of strain paths.

The following equation is used to normalize the input data to the common range of 0 ~ 1:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{9}$$

where x is the raw input variables before normalization, x_{norm} is the input variables after normalization, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the input variables, respectively.

3.2. Network architectures

For the neural network in the PINN, the input layer has 15 neurons corresponding to 1 for void ratio, 4 for stress components, 3 for strain

components, 3 for strain increment components, and 4 for plastic strain components, which inputs the variables in current step. The output layer is designed with 9 neurons, of which 1 for void ratio, 4 for stress, and 4 for plastic strain components, outputting the variables in next step.

Three optional numbers of hidden layers (2, 3, 4) as well as four optional numbers of neurons in each hidden layer (40, 60, 80, 100) are investigated for the sake of finding the optimal network architecture. Three activate functions (Relu, Sigmoid, and Tanh) are also studied. A widely used learning rate is prescribed as 0.001 (Zhang et al., 2022a). The adaptive moment estimation (Adam) optimizer is utilized since it makes the neural networks converge faster and better handle sparse gradients (Pan and Zhang, 2022). The batch size is set to a default value of 128. Additionally, the traditional MLP (in the form of Fig. 1) with the same network architectures with the PINN is set for the comparison.

3.3. Training of PINN

The generated 20 loading paths are divided into training and testing datasets using the split ratio of 80 %: 20 %, that is 16 paths for training and 4 paths for testing. Besides, the training data is shuffled in each epoch to enhance the robustness of neural networks. Fig. 6 shows the evolution of the loss function of a typical network architecture for MLP and PINN, taking 3 hidden layers and 80 neurons in each layer with the activation function of Relu as an example. The training of both MLP and PINN are shown to converge to a small loss value around 10^{-4}. For PINN model, the final loss value is slightly higher than MLP, which illustrates that adding additional term in loss function makes the training a little more difficult.

3.4. Number of hidden layers, neurons, and activation function

The coefficient of determination (R^2) and mean absolute percentage error (MAPE) are defined below for investigating optimal network architecture:

$$R^2 = 1 - \frac{\sum_{i=1}^{N_T}(y_i - y_i^*)^2}{\sum_{i=1}^{N_T}(y_i^* - \overline{y^*})^2} \tag{10}$$

$$\text{MAPE} = \frac{1}{N_T}\sum_{i=1}^{N_T}\left|\frac{y_i - y_i^*}{y_i^*}\right| \times 100\% \tag{11}$$

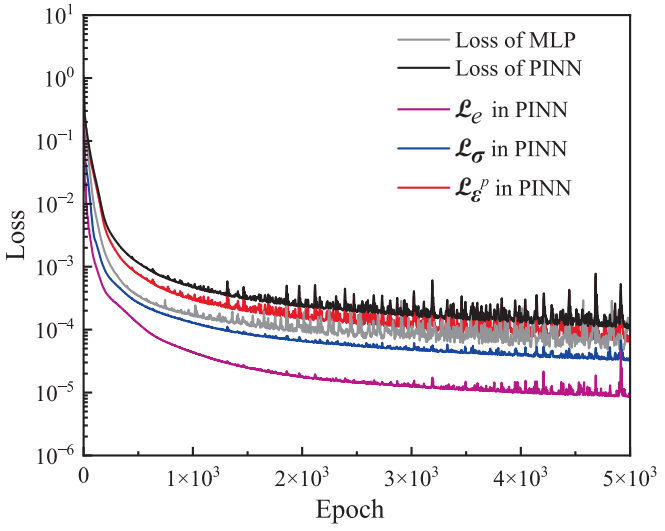


Fig. 6. Evolution of loss function during the training process.

where y^* and y denote the true and predicted values, respectively. $\overline{y^*}$ is the mean real value.

Fig. 7 displays the performance of MLP and PINN for different hidden layers and neurons in testing set (using activate function of Relu). The R^2 for the PINN are higher than those from MLP in most cases, while the MAPE from PINN is lower. Additionally, 3 hidden layers with 80 neurons reach the highest R^2 and lowest MAPE. Generally, more neurons in layers could improve the nonlinear predictive ability but have a higher risk of overfitting. In this study, 100 neurons reduced the R^2 but increased MAPE values compared to 80 neurons, illustrating a possible overfitting phenomenon. To the end, the network architecture is determined to be 3 hidden layers with 80 neurons in each layer, as it provides the best accuracy with the lowest overfitting possibility.

Fig. 8 compares the performance of different activation functions in the testing set (using 3 hidden layers and 80 neurons in each layer). Relu performs best with the highest R^2 and lowest MAPE values, followed by Tanh, while Sigmoid underperforms than the other two activate functions. Meanwhile, the PINN still performs better than MLP at all activation functions. Thereby, the optimal network architecture is determined to be 3 hidden layers with 80 neurons in each layer, using Relu.

After choosing the optimal network, the R^2 and MAPE for output variables are shown in Fig. 9 with respect to void ratio, stress tensor, and plastic strain tensor. The PINN for each group variable still performs better than the MLP with higher R^2 and lower MAPE in both training and testing sets. The void ratio performs best, followed by stress tensor, and finally plastic strain tensor. The error of void and stress compound on the prediction of plastic strain, resulting the lowest performance of the plastic strain tensor. Besides, stress tensors have four components while the void ratio has only one, which explains the higher R^2 of void ratio.

3.5. Prediction on complete loading paths

The recall mode is employed to predict a complete loading path, as illustrated in Algorithm 1. Herein, the trained network is called multiple times. The outputs of the current step are used as the inputs of the next step, in such a way to update the constitutive relationships, including stress, strain, plastic strain, and void ratio. This is much easier to implement than the implicit integration required in classical elasto-plastic theory (Potts and Gens, 1985).

Algorithm 1 Recall mode to predict a strain-loading path

Require: Pre-trained network, number of loading steps (nstep)
Initialization: $\boldsymbol{\sigma}^0$, $\boldsymbol{\varepsilon}^0$, e^0, $\boldsymbol{\varepsilon}^{p,0}$
for $i = 0$, nstep **do**
Prescribe the strain increment tensor $\Delta\boldsymbol{\varepsilon}^i$
Calculate the output of the neural network $\left(\boldsymbol{\sigma}^{i+1}, \boldsymbol{\varepsilon}^{i+1}, \boldsymbol{\varepsilon}^{p,i+1}\right) = \mathcal{NN}\left(\boldsymbol{\sigma}^i, \boldsymbol{\varepsilon}^i, \Delta\boldsymbol{\varepsilon}^i, \boldsymbol{e}^i, \boldsymbol{\varepsilon}^{p,i}\right)$
Update the state variables $\boldsymbol{\sigma}^i = \boldsymbol{\sigma}^{i+1}$, $\boldsymbol{\varepsilon}^i = \boldsymbol{\varepsilon}^i + \Delta\boldsymbol{\varepsilon}^i$, $\boldsymbol{e}^i = \boldsymbol{e}^{i+1}$, $\boldsymbol{\varepsilon}^{p,i} = \boldsymbol{\varepsilon}^{p,i+1}$
end for

By way of example, Fig. 10 exhibits a better predictive performance towards the constitutive response for a complete strain-loading path shown in Fig. 10(a) in testing set. Fig. 10(b) shows the evolution of stress, where the PINN demonstrates excellent agreement with the results of BSM, outperforming the MLP which tends to underestimate the stress components. In addition, Fig. 10(c) shows that PINN also achieves great prediction of the plastic strain and its evolution. Fig. 10(d) shows that the predicted void ratio from the PINN aligns closely with BSM results, whereas the MLP significantly underestimates the void ratio during the evolution process.

Fig. 11 provides a worst case of PINN predictions against the testing set where both the stress and strain components deviate slightly from the BSM. Whilst not perfect, the performance of PINN is substantially better than MLP, and the PINN can also successfully capture the evolution of

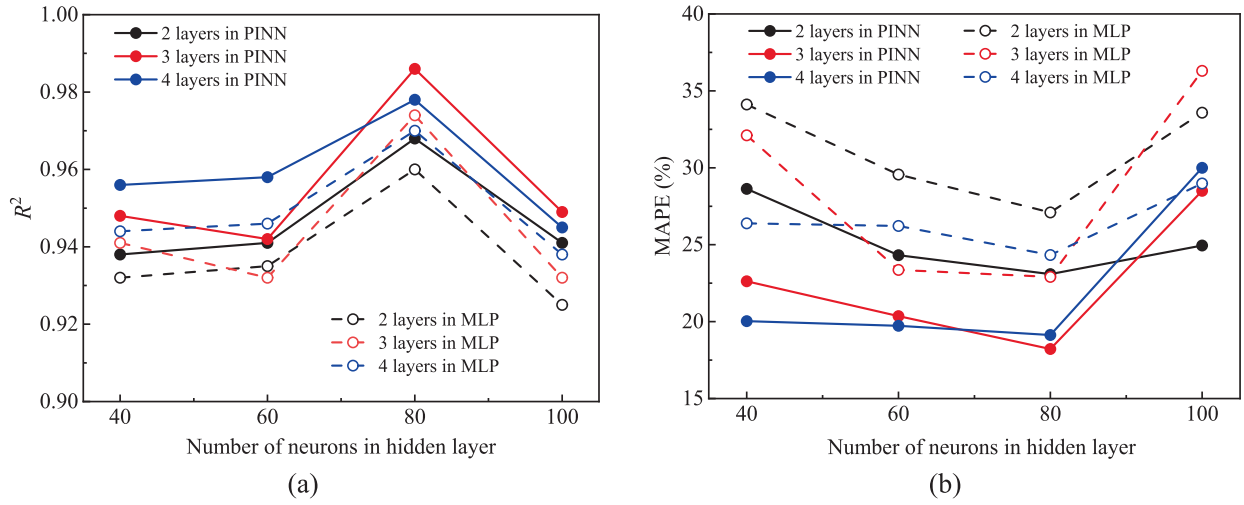


Fig. 7. Comparison of (a) R^2 and (b) MAPE for different network architectures in testing set.

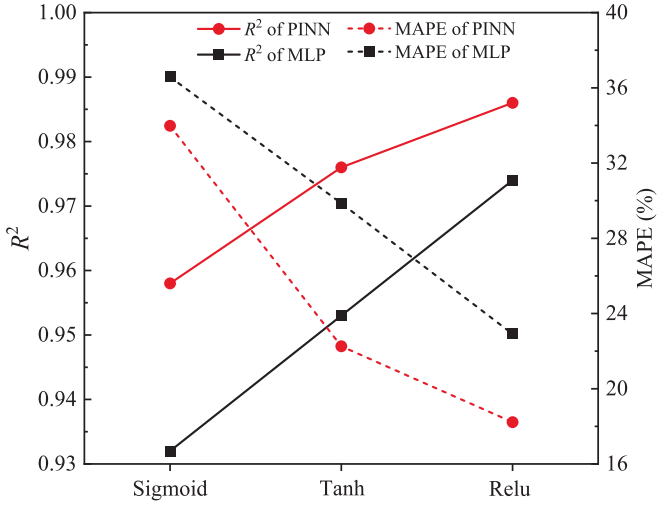


Fig. 8. Comparison of various activation functions.

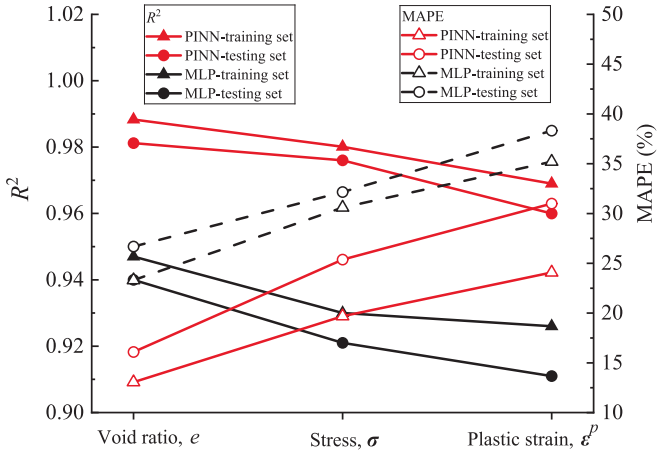


Fig. 9. Comparison of R^2 and MAPE for different group variables.

stress and strain. Besides, although the R^2 in training and testing sets have reached a very high value, there are still some discrepancies in predicting the whole loading path. This is because the recall mode uses the output in the last step instead of real value as the input in current step, thereby causing the error accumulation.

4. Integration with FEM: Results and discussion

4.1. Implicit algorithm for FEM-PINN coupling

To extend the application of the model to real cases, we incorporate the PINN into FEM for solving BVPs. Compared to constitutive models, the PINN does not need elastoplastic iterations. Fig. 12 and Algorithm 2 show the schema of the implementation of PINN into the FEM framework using an implicit algorithm. The Gauss points in the finite element domain are used to predict the displacement vector $\boldsymbol{u}$. Accordingly, the strain can be calculated via the shape function and its gradient. The strain in the current step is then transmitted to the neural network to predict the stress and physical variables in the next time step. The predicted stress and void ratio are then used to calculate the material stiffness matrix, which is returned to the Gauss points to assemble the global stiffness matrix. As for the form of material stiffness matrix, we select the state-dependent elastic stiffness tensor $\boldsymbol{D}_e$ in Eq. (5) as it: 1) has a good convergence, 2) easily relates the FEM to neural network-based constitutive modelling, and 3) can avoid any expression on plasticity. Subsequently, the global stiffness is assembled from the material stiffness matrix, and expressed as:

$$\boldsymbol{K}^{FE} = \int_{\Omega} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{D}_e \boldsymbol{B} \, \mathrm{d}\Omega \tag{12}$$

where $\boldsymbol{B}$ is the strain–displacement matrix, calculated from the gradient of shape function; Ω denotes the geometric domain. Finally, the nodal displacement and material stress are updated using Newton-Raphson iterations, thereby completing the FEM-PINN calculation. As the current void ratio and stress are predicted from the PINN, the state-dependent elastic stiffness tensor as well as the material stiffness matrix will evolve at each loading step. Besides, we emphasise that the resulting global matrix $\boldsymbol{K}^{FE}$ is an elastic tangent rather than the consistent elasto-plastic tangent of a return-mapping algorithm. While

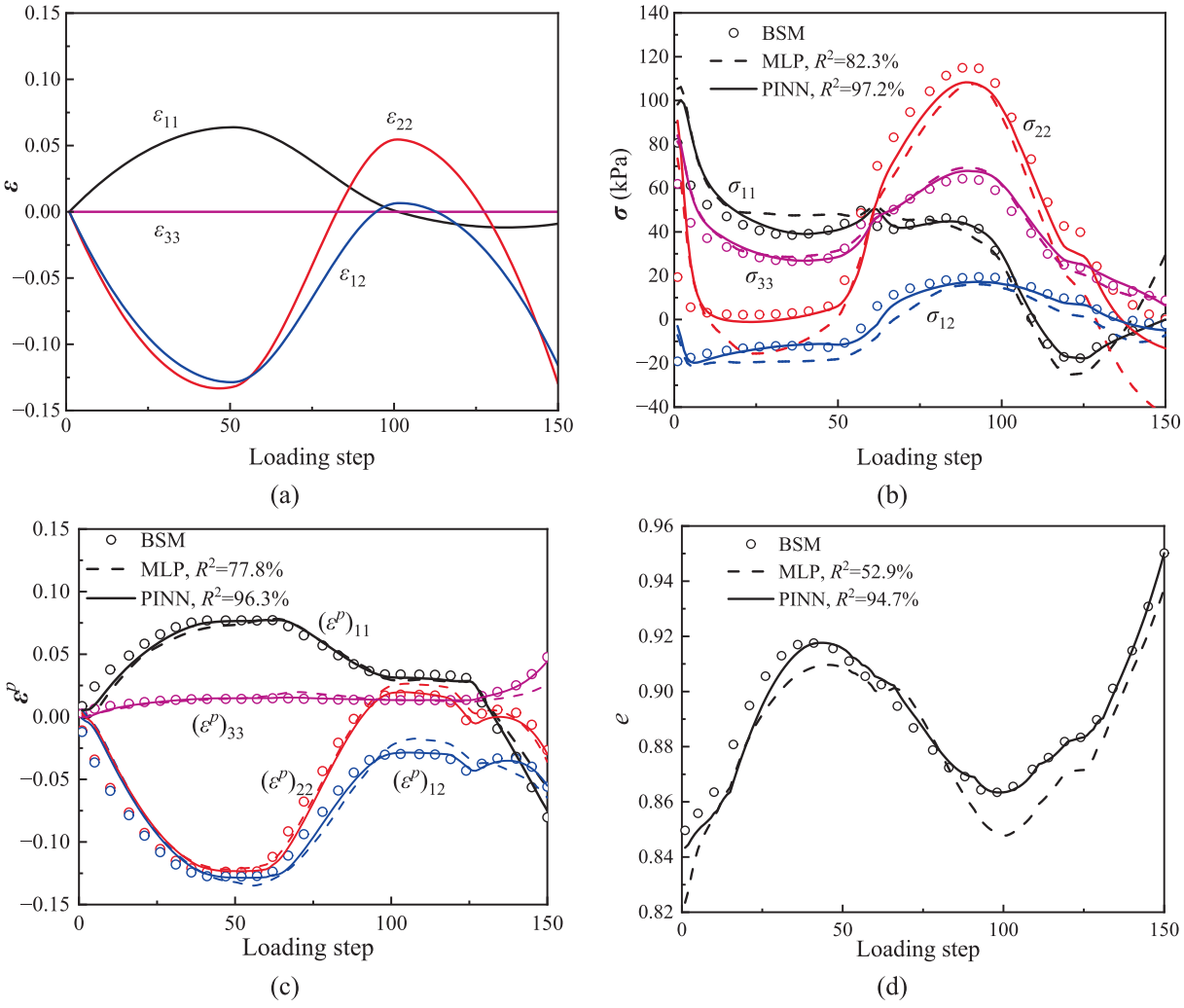


Fig. 10. Better prediction of constitutive relationship by PINN and comparison with BSM: (a) strain path; (b) stress path; (c) plastic strain; and (d) void ratio.

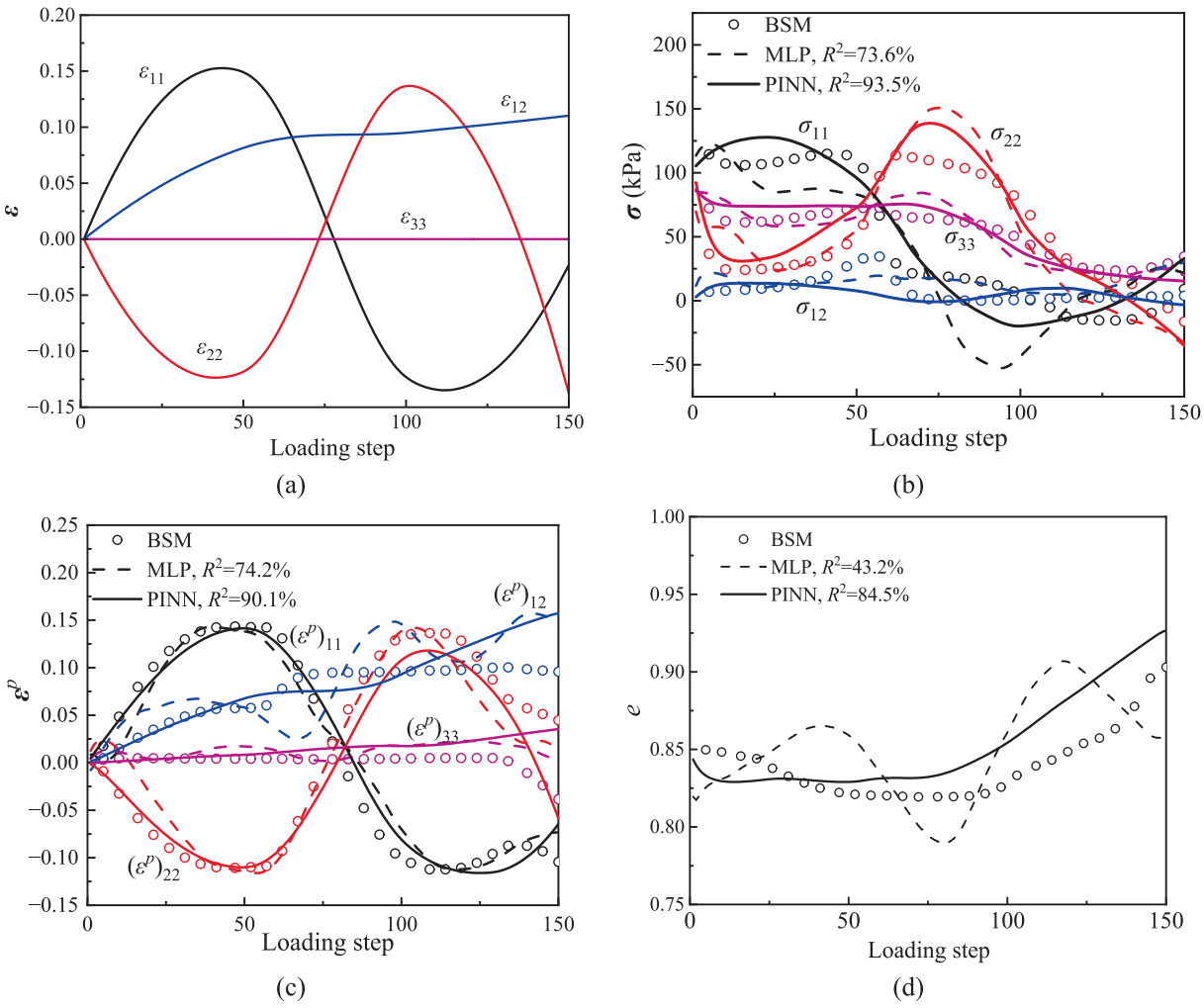


Fig. 11. Worse prediction of constitutive relationship by PINN and comparison with BSM: (a) strain path; (b) stress path; (c) plastic strain; and (d) void ratio.

approximate, this choice simplifies implementation and in practice, yielded robust convergence, requiring 2–5 Newton iterations per load step in the examples considered.

Algorithm 2 Implicit algorithm for FEM-PINN coupling

Require: Geometry, meshing, loading, boundary conditions, and pre-trained neural network
Initialization: $\boldsymbol{\sigma}^0$, $\boldsymbol{\varepsilon}^0 = 0$, e^0, $\boldsymbol{\varepsilon}^{p,0} = 0$; total loading step N_s; maximal New-Raphson iterations $N_{\max}$
for $i = 1, N_s$ **do**
Apply global loading step i
for $j = 1, N_{\max}$ **do**
Calculate the strain increment tensor $\Delta\boldsymbol{\varepsilon}^{i-1,j}$ via shape function and node displacement $\boldsymbol{U}$
Calculate stress tensor and physical variables via the neural network $\left(\boldsymbol{\sigma}^i, e^i, \boldsymbol{\varepsilon}^{p,i}\right) = \mathcal{NN}\left(\boldsymbol{\sigma}^{i-1}, \boldsymbol{\varepsilon}^{i-1}, \Delta\boldsymbol{\varepsilon}^{i-1}, e^{i-1}, \boldsymbol{\varepsilon}^{p,i-1}\right)$
Compute element elastic stiffness tensor ($\boldsymbol{D}_e$) via Eqs. (5) and (6)
Assemble global stiffness matrix $\boldsymbol{K}^{FE}$ via Eq. (12)
Calculate the internal force $\boldsymbol{F}_{\text{int}} = \boldsymbol{K}^{FE}\,\boldsymbol{U}$ and Residual force $\boldsymbol{R} = \boldsymbol{F}_{\text{int}} - \boldsymbol{F}_{\text{ext}} - \boldsymbol{F}_{\text{b}}$
if $\boldsymbol{R}$ < tolerance, **break** current loop
elseif $\boldsymbol{R}$ > tolerance
Calculate the displacement increment $\Delta\boldsymbol{U} = -\left(\boldsymbol{K}^{FE}\right)^{-1}\boldsymbol{R}$
Update node displacement vector $\boldsymbol{U} = \boldsymbol{U} + \Delta\boldsymbol{U}$
end if
end for
Update the state variables at Gauss points $\boldsymbol{\sigma}^{i-1} = \boldsymbol{\sigma}^i$, $\boldsymbol{\varepsilon}^{i-1} = \boldsymbol{\varepsilon}^i$, $e^{i-1} = e^i$, $\boldsymbol{\varepsilon}^{p,i-1} = \boldsymbol{\varepsilon}^{p,i}$
end for
Note: $\boldsymbol{F}_{\text{ext}}$ and $\boldsymbol{F}_{\text{b}}$ denote the external force and body force vectors, respectively.

In the next sections, we test the validity of the FEM-PINN framework using three different BVPs: biaxial compression, cavity expansion–contraction and foundation loading. The network was trained only on material-point data, and no BVP data were used during training. The BSM is also used as a benchmark. All cases running the BSM use the base parameters described in Table 1. We would like to highlight that the proposed framework does not violate the thermodynamic principles. Consistency was verified at all Gauss points by evaluating the plastic dissipation $\Delta w_p = \sigma : \Delta\varepsilon^p$ to be non-negative.

Another point is the computational costs of the PINN-based constitutive model. The computation time of the PINN is definitely related to the architecture of the neural network. More layers and neurons will cause a higher computational complexity. In this work, the chosen 3 hidden layers with 80 neurons presents a similar computation time with the benchmark BSM. However, if more available data is used in training, a more complex network architecture may be required. In that case, the computation time of the PINN in FEM framework also increases. A balance between the fitting ability and computational costs of the PINN-based model should be considered in this framework.

4.2. Case 1: Biaxial compression

First example aims to verify the ability of PINN in biaxial compression problems. The geometry is shown in Fig. 13 with a homogeneous soil column of size 0.08 × 0.02 m. The constant confining pressure of p_0 = 100 kPa is applied. A top-boundary displacement of S = 0.006 m is applied, corresponding to 7.5 % axial strain, while the bottom boundary

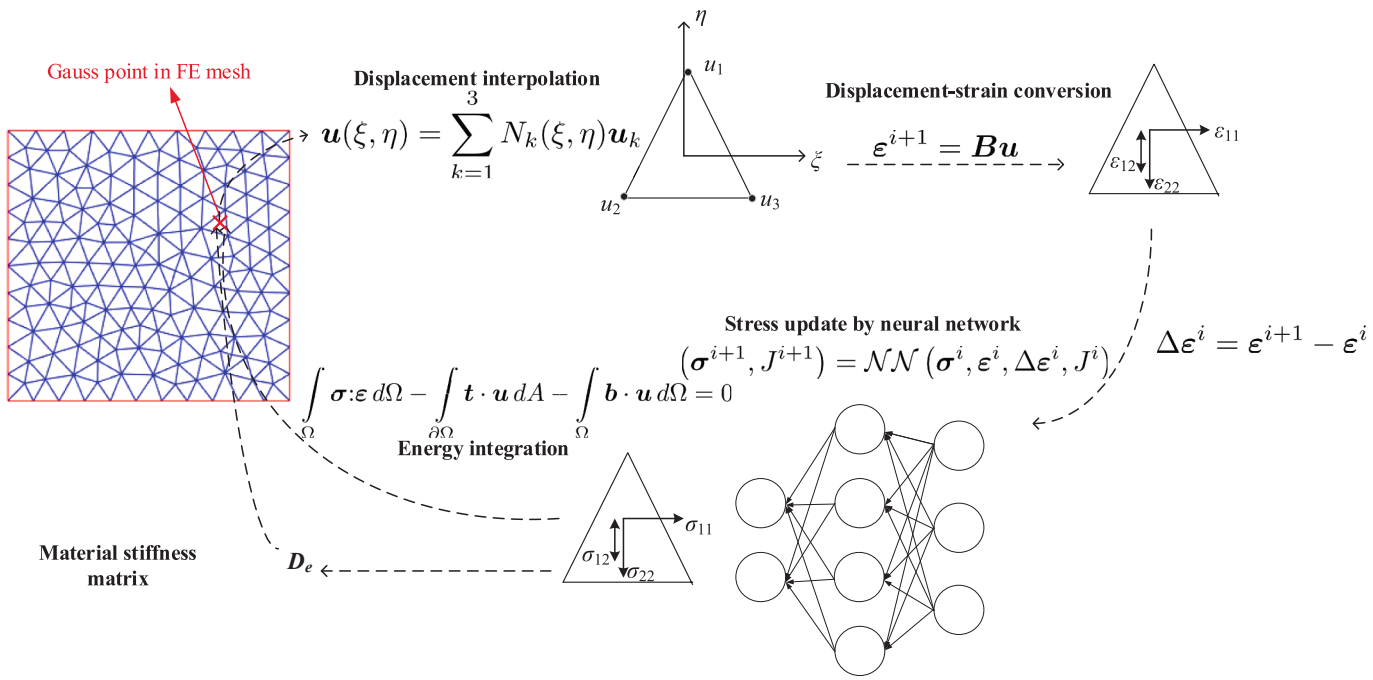


Fig. 12. Schema of the incorporation of neural network into finite element framework. N_k denotes the shape function; $\boldsymbol{t}$ denotes the traction force vector on the surface boundary $\partial\Omega$; $\boldsymbol{b}$ denotes the body force vector.

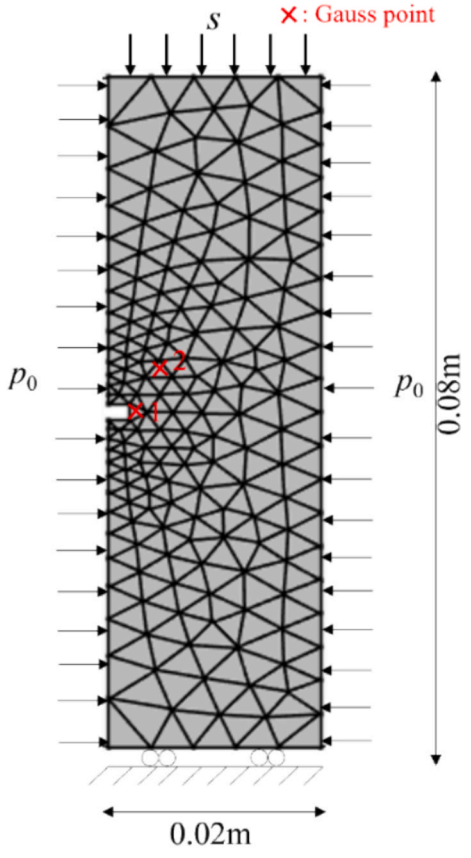


Fig. 13. FE domain of the biaxial shear test, as well as two selected Gauss points.

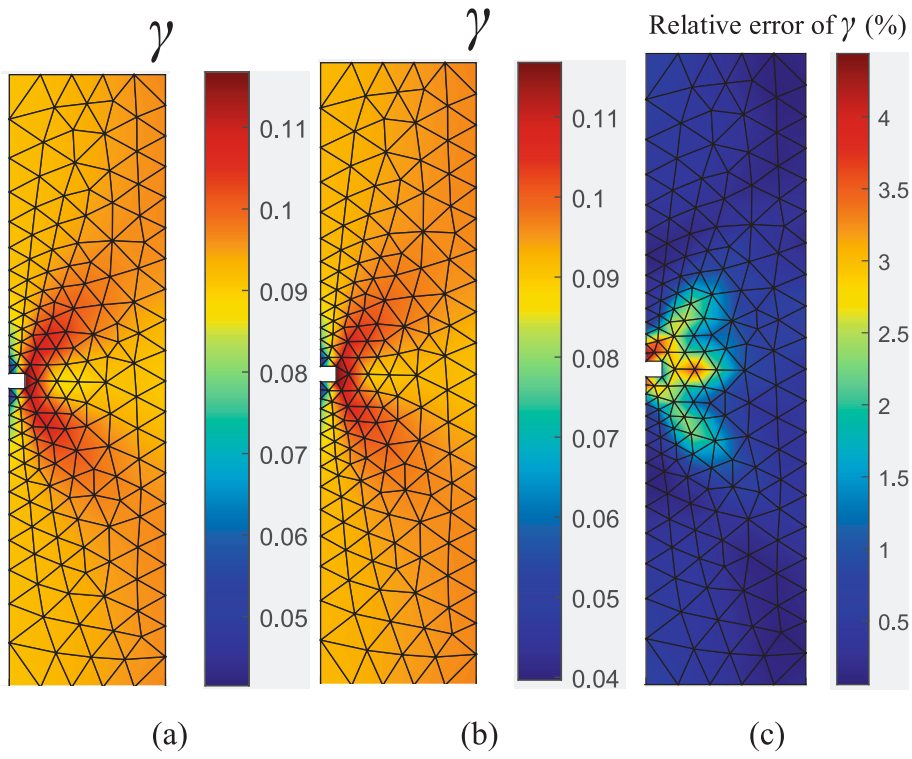


Fig. 14. Contours of shear strain γ for biaxial compression using (a) BSM and (b) PINN, as well as (c) their relative error.

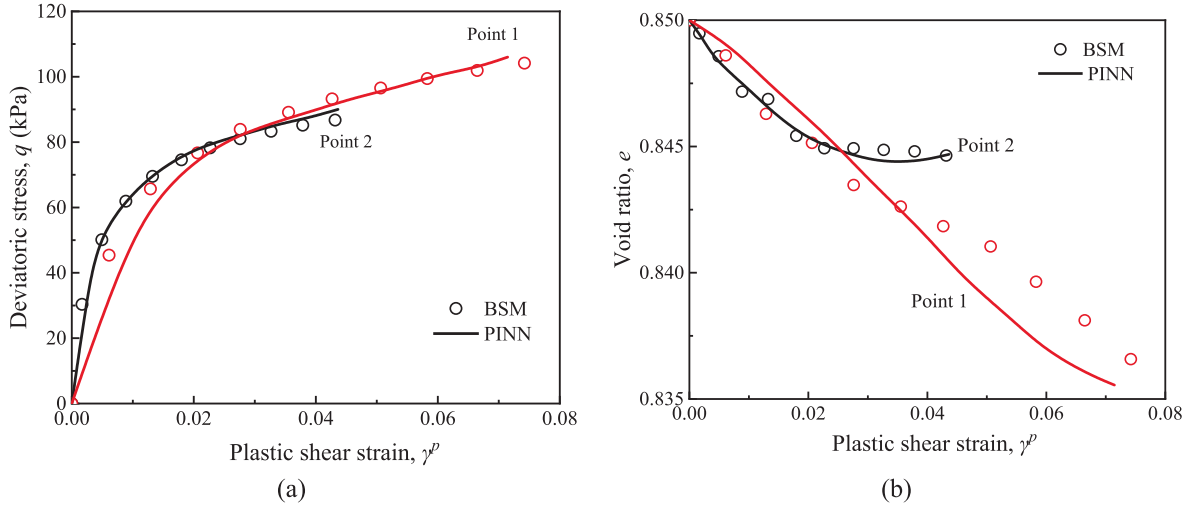


Fig. 15. Results of biaxial compression for two Gauss points. The variation of (a) deviatoric stress q and (b) void ratio e, with plastic shear strain γ^p using BSM and PINN.

Table 2
Comparison of R^2 between PINN and MLP for physical variables.

BVP	Stress, σ		Void ratio, e		Plastic strain, ϵ^p	
	PINN	MLP	PINN	MLP	PINN	MLP
Biaxial compression	94.32 %	90.98 %	95.61 %	91.64 %	92.43 %	90.97 %
Cavity expansion	83.57 %	74.64 %	84.38 %	74.92 %	80.94 %	74.17 %
Foundation loading	87.35 %	78.92 %	88.76 %	79.36 %	84.64 %	81.30 %

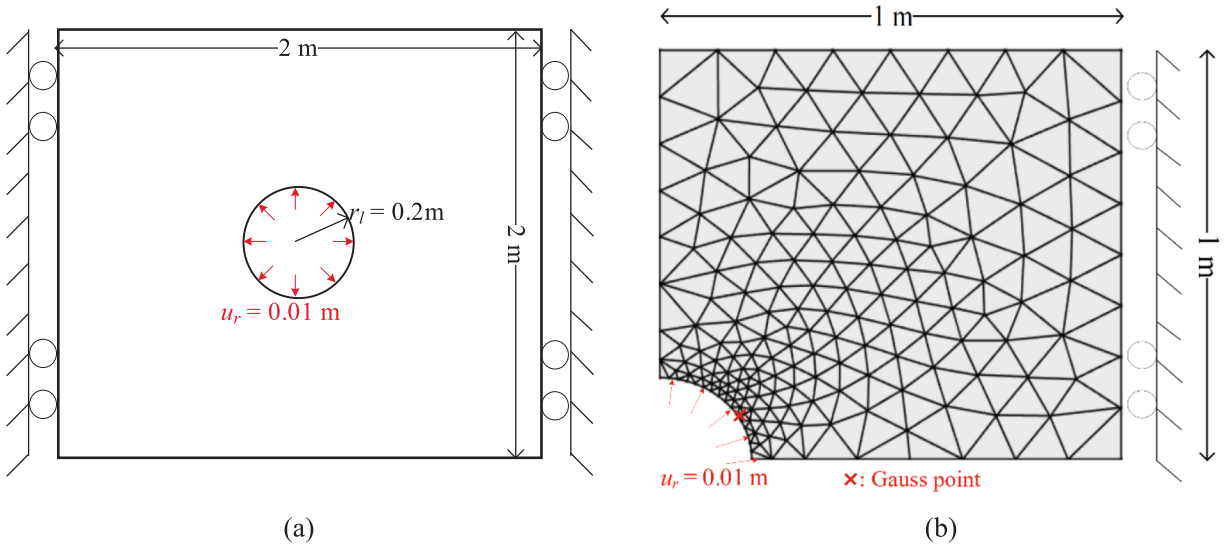


Fig. 16. FE domain of the cavity expansion: (a) geometry and boundary conditions; (b) meshing and selected Gauss point.

is fixed in the vertical direction. A small hole is used to induce in the left centre as an imperfection to trigger shear strain concentration (He and Semnani, 2023). Two Gauss points are selected to display the process of loading in detail.

The contours plot of shear strain at the end of loading are displayed in Fig. 14. It can be observed in Fig. 14(a) and (b) that the shear strain field of the PINN matches the BSM well. Besides, both methods successfully describe the expected development of a shear band around the imperfection. Furthermore, Eq. (13) is used to evaluate the relative error (RE) between the real and predicted values:

$$\mathrm{RE} = \left|\frac{y^* - y}{y^*}\right| \tag{13}$$

As present in Fig. 14(c) the relative error at all Gauss points is less than 4.5 %, indicating an excellent agreement between BSM and FEM-PINN calculations. Note that excessive strain localization would have an adverse effect on the PINN's performance because there would be less data of larger strain level for the trained model to effectively capture the material behaviour.

The evolution of deviatoric stress and void ratio with plastic shear strain at both selected Gauss points is shown in Fig. 15. Point 1 has a larger stress as it is closer to the imperfection hole. The void ratio in point 1 shows some discrepancies between the BSM and PINN, as displayed in Fig. 15(b). We argue that the reason might be that the neural network characterizes the void ratio at point 1 as linear distribution. For Gauss point 2, its void ratio almost remains constant after the shear strain of 3 %, in which the FEM-PINN also reproduced this feature.

Additionally, the MLP was also run in the FEM framework for a comparison. Table 2 shows the R^2 values of the PINN and MLP over all Gauss points for different physical state variables. It shows that the performance of the PINN is better than the MLP for all three physical variables, which demonstrates the superiority of the proposed PINN framework.

4.3. Case 2: Cavity expansion–contraction

Fig. 16(a) shows the geometry and boundary conditions of the cavity expansion–contraction problem. A cavity inside the centre of a 2 × 2 m square soil domain has a radius of $r_l = 0.2$ m. The cavity first expands in the radial direction by a displacement of $u_r = 0.01$ m, then contracts by $u_r = 0.005$ m, and finally re-expands to $u_r = 0.01$ m. The left and right boundaries are fixed to horizontal displacements, but the top and bottom boundaries are free. Only a quarter of the domain was used, as displayed in Fig. 16(b). One Gauss point on the edge of the cavity is selected to analyse as a representative.

Fig. 17 compares the result contours at maximal displacement (u_r = 0.01 m) in first expansion for total displacement, deviatoric stress, and plastic shear strain, where a very good match is observed with values of relative error within 1.5 %, 5.5 % and 7.5 % for the displacement, deviatoric stress and plastic shear strain respectively. Stress concentration emerges around the cavity edge, which has been reported by Guan et al. (2023), and a realistic larger stress region is present near the right hand side of the cavity because of the constraints of the boundary displacement. The FEM-PINN also well reproduces the localization of plastic shear strain near the cavity edge area, where the error is actually negligible. Besides, the comparison between PINN and MLP (Table 2) shows higher R^2 values of the PINN in stress tensor, void ratio, and plastic strain.

Fig. 18 shows the evolution of stress and plastic strain components with the cavity radial displacement in expansion, contraction, and re-expansion phases for the selected Gauss point. It can be observed that the stress and plastic strain during cavity expansion and contraction would undergo different loading paths for every component. Considering this fact, the FEM-PINN framework still provides well-acceptable results and successfully captures the evolution tendency for each stress and strain component. In the phase of first expansion, soil will experience an elastoplastic deformation, showing the increase of stress and plastic strain. When it contracts from the maximum radial displacement, soil presents almost an elastic behaviour with no changes in plastic strain. Finally, for the secondary expansion, soil produces the plastic deformation again as expected.

4.4. Case 3: Foundation loading

The third case aims to test the local loading on the foundation, where the geometry and boundary conditions are depicted in Fig. 19(a). A homogeneous soil foundation with a size of 5 m in length and 4 m in height is subjected to a local vertical loading at the surface on the top left side. This vertical loading has a width $B_s = 0.8$ m and is applied as a vertical settlement of $S = 0.08$ m (corresponding to $S/B_s = 10$ %). The left and right boundaries have constrained horizontal displacements, while the bottom boundary is constrained on vertical displacements. The mesh of the FE model is shown in Fig. 19(b). Similarly, we selected one Gauss point located in the shear band to analyse in detail, as shown in Fig. 19(b).

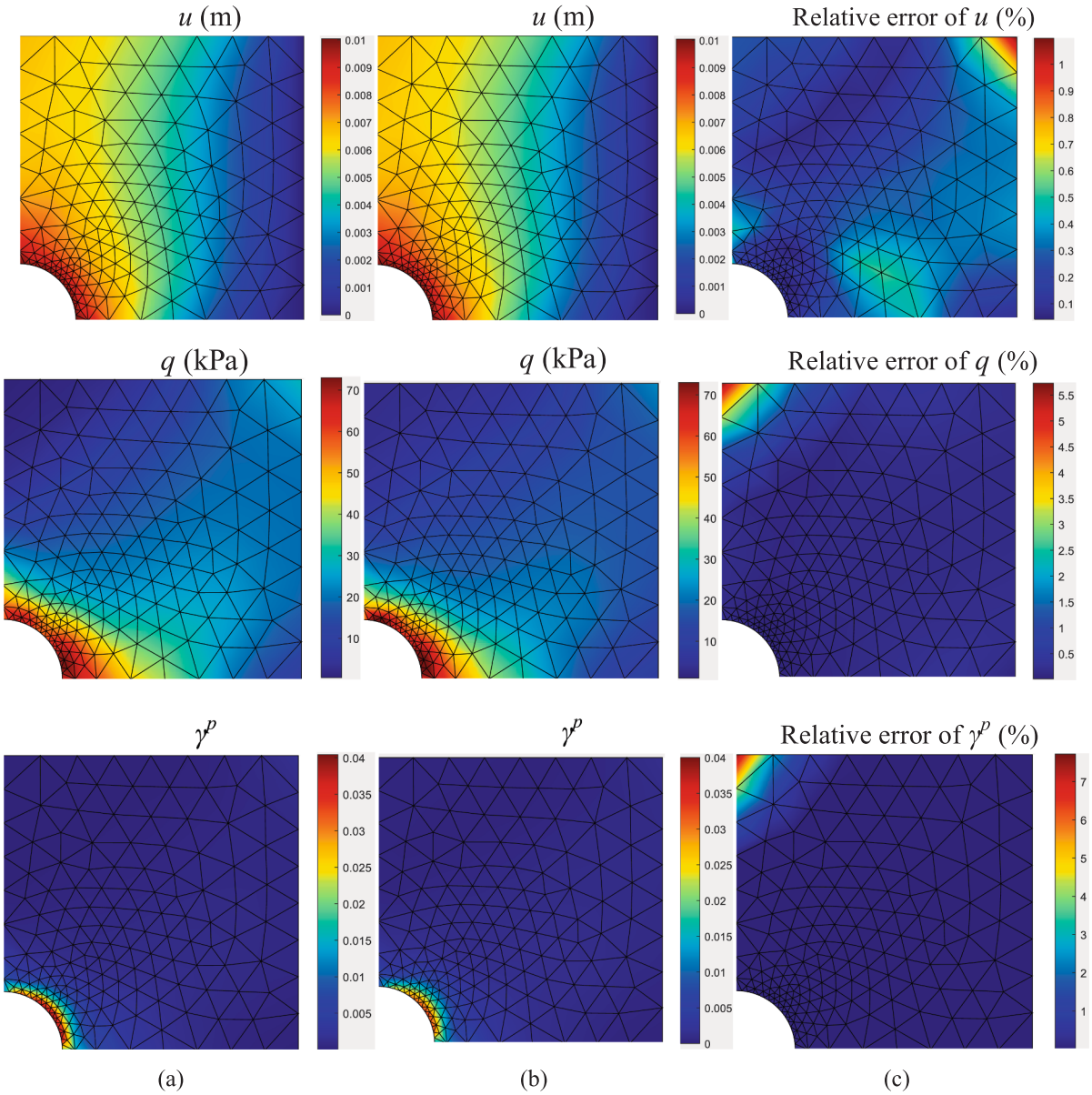


Fig. 17. Contours of FEM results for cavity expansion using (a) BSM and (b) PINN in the fields of total displacement u, deviatoric stress q, and plastic shear strain γ^p, as well as (c) their relative error.

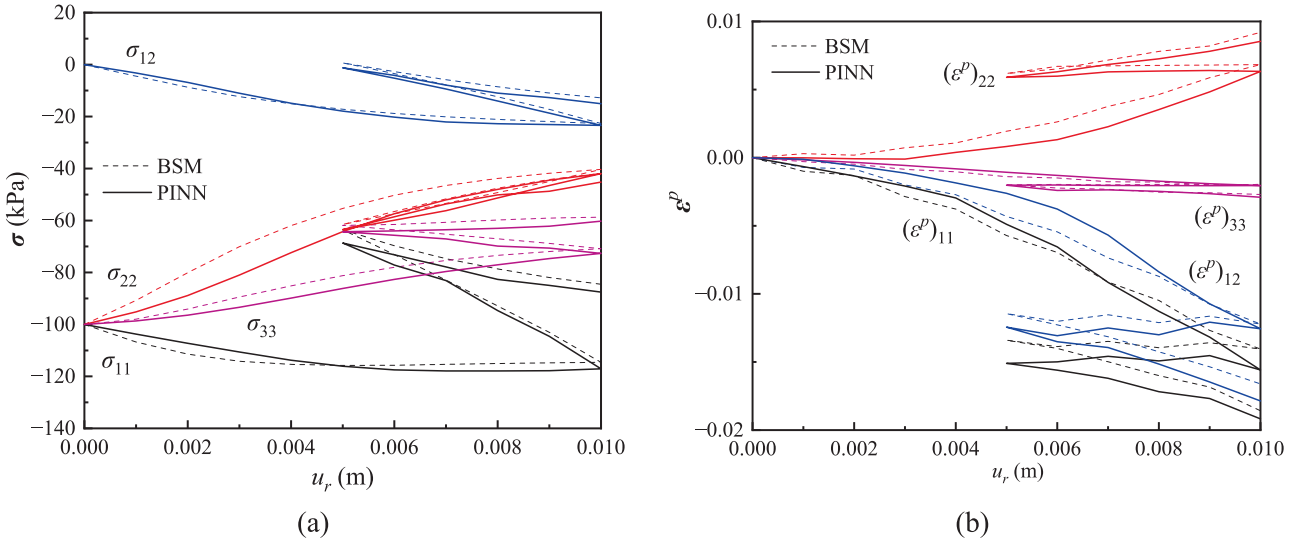


Fig. 18. Results of cavity expansion–contraction for selected Gauss point in evolution of (a) stress components; and (b) plastic shear strain components, using bounding surface model and PINN.

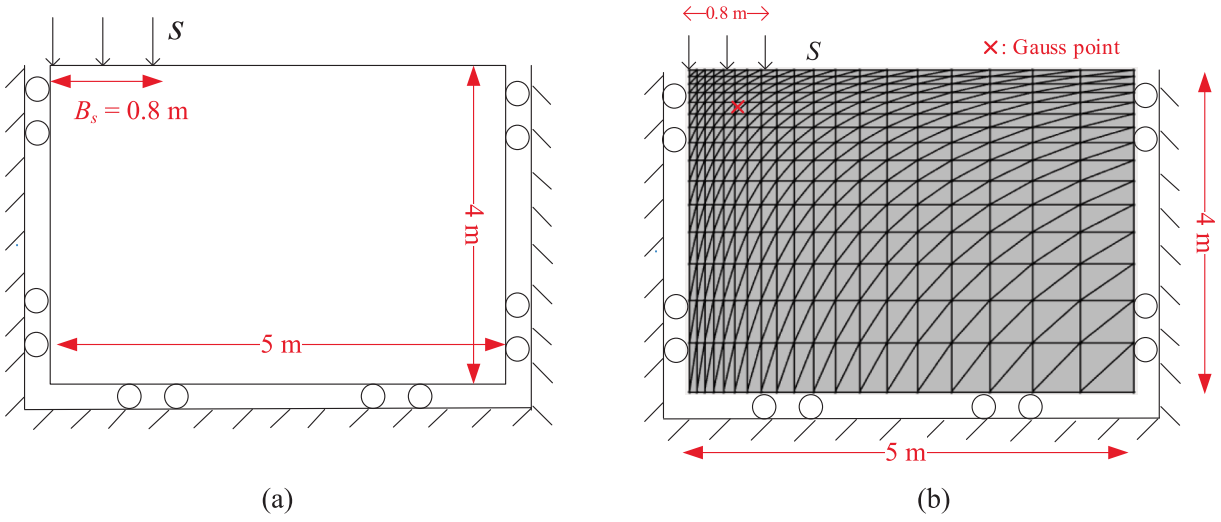


Fig. 19. FE domain of the local loading on the foundation: (a) geometry and boundary conditions; (b) meshing and selected Gauss point.

Fig. 20 compares the result contours of the BSM and FEM-PINN with values of relative error within 0.8 %, 7.0 % and 6.0 % for the vertical displacement, deviatoric stress, and plastic shear strain respectively. The right edge of vertical loading presented the maximal deviatoric stress and plastic shear strain in BSM. This feature was also reproduced by the FEM-PINN framework. In addition, Fig. 21 displays the accurately predicted evolution of q and γ^p on selected Gauss point with the loading process of S/B_s. An acceptable match between the BSM and FEM-PINN can be captured as well. Besides, Table 2 shows that the performance of the PINN in foundation loading is better than the MLP with higher R^2 values for three physical variables.

5. Conclusion

This study proposed a physical-informed neural network approach to capture the elastoplastic behaviour of soils under different strain and stress paths. The general elastoplastic stress–strain relationship with strain decomposition was used as an additional loss function to constrain the outputs to a framework of elastoplasticity. The results demonstrated that the loss function which includes the restriction of physical information, could achieve a better performance than traditional deep networks based on MLP. The PINN provided a more stable prediction in predicting stress and strain in complex loading paths. It could also accurately capture the evolution of plastic strain components and void ratio. Critically, it achieved this without using any prior knowledge on plasticity, such as a yield function or hardening laws.

The PINN was then incorporated into the FEM framework as a replacement for the constitutive model to test its performance in unseen boundary value problems. Three cases were examined: biaxial compression, cavity expansion–contraction, and foundation loading. The results indicated that the FEM-PINN calculation exhibited excellent performance of deviatoric stress, plastic shear strain, and void ratio in comparison to the benchmark bounding surface model. In addition, the stress concentration and unloading–reloading paths were also reproduced by the FEM-PINN framework.

The PINN for elastoplastic soil in this study is an interesting approach to model complex soil behaviours in engineering applications because it uses only easily measurable input parameters, such as void ratio and stress. This makes the framework easily implementable with laboratory measurements and also for practising engineers in cases where complex

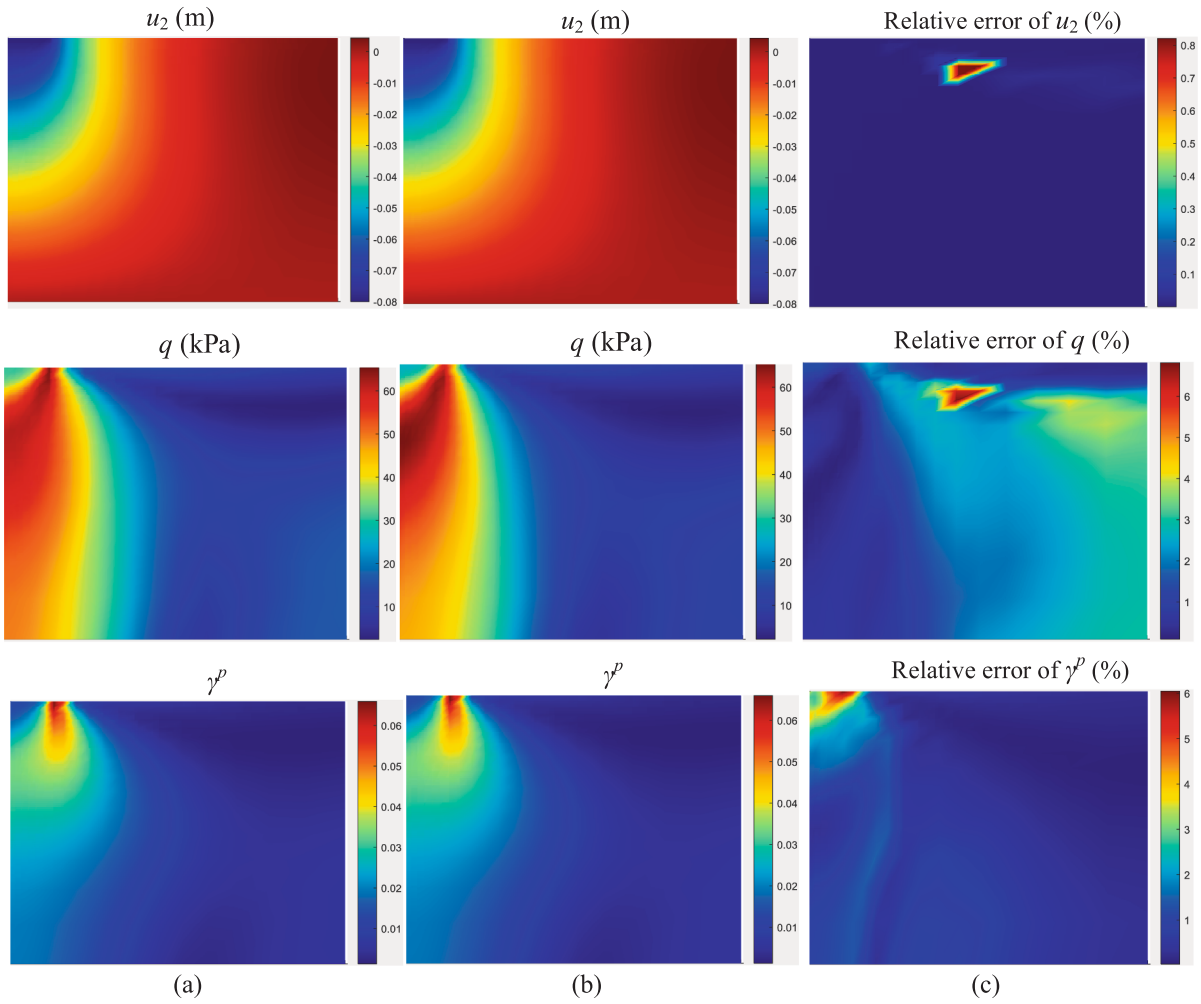


Fig. 20. Contours of FEM results for local loading on foundation using (a) BSM and (b) PINN in the fields of vertical displacement u_2, deviatoric stress q, and plastic shear strain γ^p, as well as (c) their relative error.

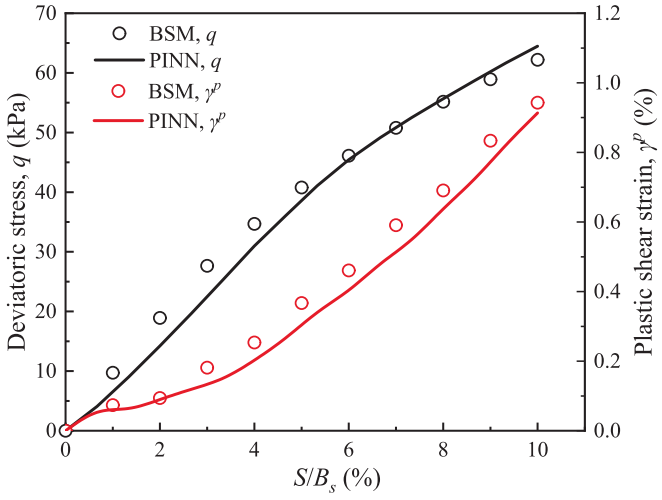


Fig. 21. Results of local loading on foundation for selected Gauss point in deviatoric stress q and plastic shear strain γ^p using BSM and PINN.

constitutive models are not easily accessible.

Future work will involve a stricter number of physical constraints, such as full enforcement of thermodynamics compliance. We will also explore stress-driven calculations and cyclic and rapid loading conditions.

CRediT authorship contribution statement

Mingpeng Liu: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Qinghua Zhang:** Writing – review & editing, Investigation, Formal analysis. **Raul Fuentes:** Writing – review & editing, Supervision, Project administration, Investigation.

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.

Appendix:. Bounding surface model

The bounding surface model adopted in this research is based on the CASM by Yu (1998) and the model by Zhou et al. (2017). The elastic volumetric ($d\varepsilon_V^e$) and deviatoric strain ($d\varepsilon_q^e$) increments are calculated using K and G (Yang et al., 2023):

$$d\varepsilon_V^e = \frac{dp}{K} = \frac{\kappa dp}{(1+e)p} \tag{A.1}$$

$$d\varepsilon_q^e = \frac{dq}{3G} = \frac{2(1+\mu)dq}{9(1-2\mu)K} \tag{A.2}$$

The normal consolidation line (NSL) of soil is expressed as:

$$\nu = N - \lambda \ln p \tag{A.3}$$

where $\nu = (1+e)$ is known as specific volume; λ is the slope of NSL in the $\nu - \ln p$ space; N denote the ν value of NSL at $p = 1$ kPa. The critical state line (CSL) is shown as follows:

$$q = Mp \tag{A.4}$$

$$\nu = \Gamma - \lambda \ln p \tag{A.5}$$

where M is the slope of CSL in p-q space; Γ denotes the ν value of CSL at $p = 1$ kPa.

Fig. A depicts the bounding surface in p-q space. The formula of the bounding surface is shown as:

$$f = \left(\frac{q}{Mp}\right)^n + \frac{\ln(p/p_c)}{\ln r} \tag{A.6}$$

where n and r are parameters. r can be expressed as a function of Γ, N, λ, and κ with respect to the critical state of soil:

$$r = \exp\left(\frac{N-\Gamma}{\lambda-\kappa}\right) \tag{A.7}$$

The radius mapping rule is employed, in which the mapping centre point is the original point in stress space. The mapping rule is expressed as:

$$\frac{p}{p_m} = \frac{q}{q_m} = \frac{\rho}{\rho_m} \tag{A.8}$$

where p_m and q_m are the stress components in the mapping point; ρ and ρ_m are the distance to the mapping centre from the real point and mapping point, respectively.

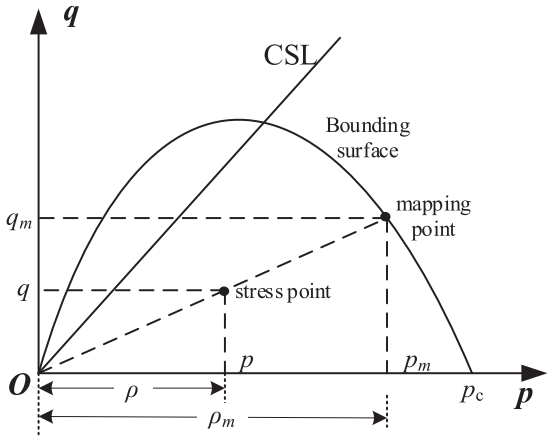


Fig. A. Schema of bounding surface model.

The bounding surface model adopts the non-associated flow rule, in which the stress-dilation function D is expressed as follows:

$$D = \frac{\mathrm{d}\varepsilon_V^p}{\mathrm{d}\varepsilon_q^p} = \frac{M^2(\rho/\rho_m) - \eta^2}{2\eta} \tag{A.9}$$

where $\mathrm{d}\varepsilon_V^p$ and $\mathrm{d}\varepsilon_q^p$ are plastic volumetric and deviatoric strain increments, respectively; $\eta = q/p$ is the stress ratio. The following isotropic hardening law is adopted to relate the variation of the bounding surface $\mathrm{d}p_c$ to the internal variable $\mathrm{d}\varepsilon_V^p$ as follows:

$$\mathrm{d}p_c = \frac{\nu}{\lambda - \kappa} p_c \, \mathrm{d}\varepsilon_V^p \tag{A.10}$$

The stress state should remain on the bounding surface during plastic evolution, as follows:

$$\frac{\partial f}{\partial p}\mathrm{d}p + \frac{\partial f}{\partial q}\mathrm{d}q + \frac{\partial f}{\partial p_c}\frac{\partial p_c}{\partial \varepsilon_V^p}\mathrm{d}\varepsilon_V^p = 0 \tag{A.11}$$

Substituting $\mathrm{d}\varepsilon_V^p$ by the $D\mathrm{d}\varepsilon_q^p$ can deduce to:

$$\mathrm{d}\varepsilon_q^p = \frac{1}{K_p}\left\{\frac{\partial f}{\partial \boldsymbol{\sigma}}\right\}^{\mathrm{T}}\{\mathrm{d}\boldsymbol{\sigma}\} \tag{A.12}$$

where K_p is the plastic modulus and expressed as:

$$K_p = -\frac{\partial f}{\partial p_c}\frac{\partial p_c}{\partial \varepsilon_V^p} D = \frac{\nu}{(\lambda - \kappa)\ln r}\frac{M^2 - \eta^2}{2\eta} \tag{A.13}$$

The above formula is only suitable to the point on the bounding surface. For the points within the bounding surface, Zhou et al. (2017) used a relatively simple form:

$$K_p = \frac{\nu}{(\lambda - \kappa)\ln r}\frac{M^2(\rho/\rho_m) - \eta^2}{2\eta} \tag{A.14}$$

in which the introduction of ρ/ρ_m reflects that the state points closer to the bounding surface will have a smaller plastic modulus.

The Eqs. (A-1) to (A-14) formulate the theoretical basis of the bounding surface model. As mentioned above, there are seven parameters that need to be calibrated: λ, κ, μ, M, N, Γ, and n, by which other parameters and corresponding constitutive law can be derived from these seven parameters.

Data availability

Data will be made available on request.

References

Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E., 2021. Physics-informed neural networks (PINNs) for fluid mechanics: a review. Acta Mech. Sin. 37 (12), 1727–1738.

Eghbalian, M., Pouragha, M., Wan, R., 2023. A physics-informed deep neural network for surrogate modeling in classical elasto-plasticity. Comput. Geotech. 159, 105472.

Guan, Q.Z., Yang, Z.X., 2023. Hybrid deep learning model for prediction of monotonic and cyclic responses of sand. Acta Geotech. 18, 1447–1461.

Guan, Q.Z., Yang, Z.X., Guo, N., Hu, Z., 2023. Finite element geotechnical analysis incorporating deep learning-based soil model. Comput. Geotech. 154, 105120.

Haghighat, E., Raissi, M., Moure, A., Gomez, H., Juanes, R., 2021. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. Comput. Methods Appl. Mech. Eng. 379, 113741.

Haghighat, E., Abouali, S., Vaziri, R., 2023. Constitutive model characterization and discovery using physics-informed deep learning. Eng. Appl. Artif. Intel. 120, 105828.

Hashash, Y.M.A., Jung, S., Ghaboussi, J., 2004. Numerical implementation of a neural network based material model in finite element analysis. Int. J. Numer. Meth. Eng. 59 (7), 989–1005.

He, G.-F., Zhang, P., Yin, Z.-Y., Jin, Y.-F., Yang, Y., 2023. Multi-fidelity data-driven modelling of rate-dependent behaviour of soft clays. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 17(1):64-76.
He, Y., Semnani, S.J., 2023. Machine learning based modeling of path-dependent materials for finite element analysis. Comput. Geotech. 156, 105254.
Kirchdoerfer, T., Ortiz, M., 2016. Data-driven computational mechanics. Comput. Methods Appl. Mech. Eng. 304, 81–101.
Lai, F., Liu, S., Shiau, J., Liu, M., Cai, G., Huang, M., 2025a. Data-driven modeling for evaluating deformation of a deep excavation near existing tunnels. Underground Space 24, 162–179.
Lai, F., Tschuchnigg, F., Schweiger, H., Liu, S., Shiau, J.S., Cai, G., 2025b. A numerical study of deep excavations adjacent to existing tunnels: Integrating CPTU and SDMT to calibrate soil constitutive model. Can. Geotech. J.
Linka, K., Kuhl, E., 2023. A new family of Constitutive Artificial Neural Networks towards automated model discovery. Comput. Methods Appl. Mech. Eng. 403, 115731.
Liu, M., Sun, E., Zhang, N., Lai, F., Fuentes, R., 2024a. A virtual calibration chamber for cone penetration test based on deep-learning approaches. J. Rock Mech. Geotech. Eng.
Liu, M., Zhuang, P., Lai, F., 2024b. A Bayesian optimization-genetic algorithm-based approach for automatic parameter calibration of soil models: Application to clay and sand model. Comput. Geotech. 176, 106717.
Masi, F., Stefanou, I., 2022. Multiscale modeling of inelastic materials with Thermodynamics-based Artificial Neural Networks (TANN). Comput. Methods Appl. Mech. Eng. 398, 115190.
Masi, F., Einav, I., 2024. Neural integration for constitutive equations using small data. Comput. Methods Appl. Mech. Eng. 420, 116698.
Meng, X., Karniadakis, G.E., 2020. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. J. Comput. Phys. 401, 109020.
Mozaffar, M., Bostanabad, R., Chen, W., Ehmann, K., Cao, J., Bessa, M.A., 2019. Deep learning predicts path-dependent plasticity. Proc. Natl. Acad. Sci. 116 (52), 26414–26420.
Pan, Y., Zhang, L., 2022. Mitigating tunnel-induced damages using deep neural networks. Autom. Constr. 138, 104219.
Potts, D.M., Gens, A., 1985. A critical assessment of methods of correcting for drift from the yield surface in elasto-plastic finite element analysis. Int. J. Numer. Anal. Meth. Geomech. 9 (2), 149–159.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.
Roscoe, K.H., Burland, J.B., 1968. On the generalised stress-strain behaviour of 'wet' clay. Cambridge University Press, Cambridge, Engineering Plasticity, pp. 535–609.
Roy, A.M., Guha, S., 2023. A data-driven physics-constrained deep learning computational framework for solving von Mises plasticity. Eng. Appl. Artif. Intel. 122, 106049.
Su, M.M., Yu, Y., Chen, T.H., Guo, N., Yang, Z.X., 2024. A thermodynamics-informed neural network for elastoplastic constitutive modeling of granular materials. Comput. Methods Appl. Mech. Eng. 430, 117246.
Vahab, M., Haghighat, E., Khaleghi, M., Khalili, N., 2022. A Physics-Informed Neural Network Approach to Solution and Identification of Biharmonic Equations of Elasticity. J. Eng. Mech. 148 (2), 04021154.
Wang, Z., Cudmani, R., Olarte, A.P., A.,, 2024. Tensor-based physics-encoded neural networks for modeling constitutive behavior of soil. Comput. Geotech. 170, 106173.
Wu, M., Xia, Z., Wang, J., 2023. Constitutive modelling of idealised granular materials using machine learning method. J. Rock Mech. Geotech. Eng. 15 (4), 1038–1051.
Yang, H., Yu, H.-S., Chen, X., Zhuang, P.-Z., 2023. Rigorous solution for drained analysis of spherical cavity expansion in soils of finite radial extent. Comput. Geotech. 160, 105516.
Yu, H.S., 1998. CASM: a unified state parameter model for clay and sand. Int. J. Numer. Anal. Meth. Geomech. 22 (8), 621–653.
Zhang, A., Mohr, D., 2020. Using neural networks to represent von Mises plasticity with isotropic hardening. Int. J. Plast 132, 102732.
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. Meth. Geomech. 44 (9), 1315–1335.
Zhang, P., Yin, Z.-Y., Jin, Y.-F., 2021. State-of-the-Art Review of Machine Learning Applications in Constitutive Modeling of Soils. Arch. Comput. Meth. Eng. 28 (5), 3661–3686.
Zhang, P., Yin, Z.-Y., Jin, Y.-F., Sheil, B., 2022a. Physics-constrained hierarchical data-driven modelling framework for complex path-dependent behaviour of soils. Int. J. Numer. Anal. Meth. Geomech. 46 (10), 1831–1850.
Zhang, P., Yin, Z.-Y., Jin, Y.-F., Yang, J., Sheil, B., 2022b. Physics-Informed Multifidelity Residual Neural Networks for Hydromechanical Modeling of Granular Soils and Foundation Considering Internal Erosion. J. Eng. Mech. 148 (4), 04022015.
Zhang, P., Yin, Z.-Y., Sheil, B., 2023. Interpretable data-driven constitutive modelling of soils with sparse data. Comput. Geotech. 160, 105511.
Zhang, P., Sheil, B., Girolami, M., Yaji, K., Yin, Z.-Y., 2024. A novel consolidation analysis framework: universal function approximators regularized by physical principles. Can. Geotech. J.
Zhang, P., Karapiperis, K., Weeger, O., 2025. t-PiNet: A thermodynamics-informed hierarchical learning for discovering constitutive relations of geomaterials. J. Mech. Phys. Solids 197, 106049.
Zhou, C., Fong, K.Y., Ng, C.W.W., 2017. A new bounding surface model for thermal cyclic behaviour. International Journal for Numerical and Analytical Methods in Geomechanics 41 (16), 1656–1666.
Zhou, X., Lu, D., Zhang, Y., Du, X., Rabczuk, T., 2022. An open-source unconstrained stress updating algorithm for the modified Cam-clay model. Comput. Methods Appl. Mech. Eng. 390, 114356.