International Journal of Plasticity 171 (2023) 103786

Contents lists available at ScienceDirect

International Journal of Plasticity

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

Data-driven multiscale modelling of granular materials via knowledge transfer and sharing

Tongming Qu [a,*], Jidong Zhao [a,*], Shaoheng Guan [b,c], Y.T. Feng [b]

[a] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong SAR, China
[b] Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Swansea, Wales SA1 8EP, UK
[c] Institute of Theoretical Physics-Computational Physics, Graz University of Technology, Petersgasse 16, Graz 8010, Austria

ARTICLE INFO

Keywords:
Granular materials
DEM
Machine learning
Transfer learning
Data-driven material modelling
Hierarchical multiscale modelling

ABSTRACT

Machine learning approaches have found immense potential to revolutionise the constitutive modelling of granular materials. However, data scarcity poses a significant challenge to this emerging paradigm. This study aims to tackle this issue by presenting two transfer learning-based strategies that harness well-established constitutive knowledge and similar material data to reduce data demands for data-driven material modelling. The first approach utilises phenomenological constitutive models to generate massive synthetic data which reflect the targeted material behaviour to train a base model. This base model is then repurposed for a new task based on numerical simulation data via transfer learning. The other approach involves using available material data to train a base model, which is then applied to other new materials that are similar but with limited data. The proposed transfer learning methods are tested on both particle-scale simulations of representative volume elements (RVEs) and hierarchical multiscale modelling of boundary value problems (BVPs) of granular materials. The trained data-driven material model is embedded in numerical simulations with the finite element method (FEM) to validate its accuracy, efficiency, and stability. The results demonstrate that transfer learning can effectively achieve high-quality machine learning predictions with limited data. The transfer learning strategy presented in this study is expected to be widely applicable to small data-driven material modelling.

1. Introduction

Granular materials, such as soils, are primary constitutions of the earth's land surface. Almost all engineering activities of our human beings involve the handling of this type of material. The constitutive relation describes how a material deforms when subjected to external loads. As the bedrock of modern computational geomechanics, constitutive models of granular media are essential for analysing diverse geotechnical problems.

The existing research on constitutive modelling of granular materials can be roughly categorised into four paradigms, as shown in Fig. 1. The earliest paradigm is experimentally based and involves laboratory element tests, such as direct or simple shear tests and triaxial compression tests, as the primary means to understand the material behaviour of granular soils. Even in the present day,

* Corresponding authors.
E-mail addresses: tongmingqu@ust.hk (T. Qu), jzhao@ust.hk (J. Zhao).

https://doi.org/10.1016/j.ijplas.2023.103786
Received 29 June 2023; Received in revised form 14 October 2023;
Available online 30 October 2023

experimental approaches continue to be a standard way for understanding soil properties, as well as validating and improving new models.

The valuable testing data and interpretation gained from experimental observations further inspire the development of theoretical constitutive models, constituting the *analytical* paradigm as shown in Fig. 1. To date, phenomenological constitutive modelling remains the predominant approach utilized for characterizing the strength-deformation behaviour of granular soils, as well as for computational applications in geotechnical engineering problems. The success of a phenomenological constitutive theory is underpinned by some fundamental assumptions. For instance, in classical plasticity theory, these assumptions include the explicit partition of elastic and plastic domains via the concept of yield surface, the determination of the direction of plastic flow with associated or non-associated flow rule, and the evolutional description of the yield surface through hardening laws. Although remaining popular, plasticity theory has been found to have assumptions un-conforming to the actual experimental or numerical observations of granular soils (Kawamoto et al., 2018; Kuhn and Daouadji, 2018; Pouragha and Wan, 2017). In addition, it often mandates the introduction of an excessive number of phenomenological model parameters to describe complex soil behaviour such as anisotropy and liquefaction.

A relatively new, pure computationally-based strategy has prevailed for the past decade where a micromechanics-based approach such as the discrete element method (DEM) is used to produce constitutive responses and is further coupled with a continuum-based numerical approach, such as the finite element method (FEM) or material point method (MPM) in hierarchical multiscale structure to bypass the use of phenomenological constitutive models for direct simulation of a boundary value problem (BVP) (Guo and Zhao, 2014; Liang and Zhao, 2019). However, high computational costs have limited wide applications of the multiscale modelling approaches since each Gauss or material point has an RVE consisting of a packing of DEM particles that needs to be solved synchronously for an entire continuum domain at each loading step (Guo and Zhao, 2016).

More recently, data-driven constitutive modelling with machine learning has emerged as a new paradigm to complement the previously discussed laboratory experiments, phenomenological models, and hierarchical multiscale modelling (Abueidda et al., 2021; Gorji et al., 2020; Guan et al., 2023b; Jang et al., 2021; Mozaffar et al., 2019; Zhang and Mohr, 2020). In contrast to phenomenological models, the data-driven method has some unique merits: (1) no phenomenological assumptions, such as yield surface, flow rules and hardening laws; (2) no need to calibrate free parameters; (3) The ability of continuous improving the model when new data associated materials are obtained. The earliest application of neural networks for predicting stress-strain relations of concrete and sands can be traced back to the 1990s (Ellis et al., 1995; Ghaboussi et al., 1991; Ghaboussi and Sidarta, 1998). Together with the revolutionary progress of deep learning over recent years, data-driven constitutive modelling is receiving increasing attention (Guan et al., 2023a; Ibragimova et al., 2021; Jordan et al., 2020; Tancogne-Dejean et al., 2021).

Unlike conventional numerical or analytical approaches, the predictive capability of a data-driven model is often determined by both the *model* itself and the available *data* sets. On the *model* aspect, one of the goals is to explore the potential of existing models or strategies, such as long-short term memory (LSTM) (Ma et al., 2022; Wu and Wang, 2022) and its variant (Zhang et al., 2020), gated recurrent units (GRU) (Qu et al., 2021a), convolutional neural networks (CNNs) (Ibragimova et al., 2022), and temporal convolutional neural networks (TCN) (Wang et al., 2022), or develop new models for constitutive modelling, e.g. minimal state cell (MSC) (Bonatti and Mohr, 2021, 2022), reinforcement learning (Wang and Sun, 2019), and adversarial learning (Wang et al., 2019). Another goal is to incorporate domain knowledge or constraints into model training, e.g. frame-indifference (Heider et al., 2020; Qu et al., 2021b), thermodynamics (Masi and Stefanou, 2022; Masi et al., 2021), micromechanics (Qu et al., 2021a). The importance of *data* in machine learning-based constitutive modelling cannot be overstated, yet limited attention has been given to this aspect. Some notable attempts to tackle the challenges associated with data include: multi-fidelity predictions (Su et al., 2023; Zhang et al., 2022), and deep active learning strategy (Qu et al., 2023). In addition, machine learning is used to replace some components in the framework of plasticity, for example, the yield surface (Nascimento et al., 2023; Vlassis and Sun, 2021), hardening laws (Li et al., 2019; Wen et al., 2021), and traditional stress-integration scheme (Fazily and Yoon, 2023).

Within the machine learning family, transfer learning offers a potential solution to mitigating the challenge of data scarcity. A unique attraction of transfer learning is its ability to exploit knowledge gained from one task and dataset to improve the performance of a model in a related but different task and dataset. Successful applications of transfer learning in the field of solid mechanics include

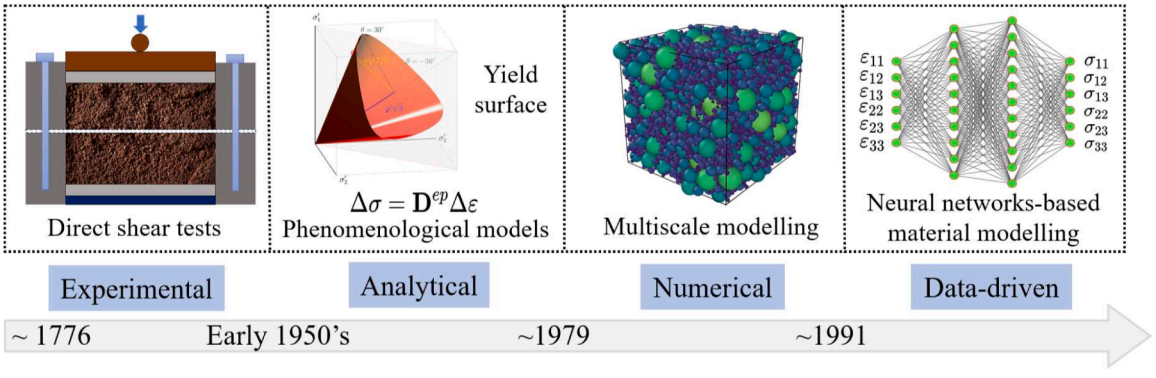


Fig. 1. Paradigm shift on constitutive modelling of granular materials.

fault slip prediction (Wang et al., 2021), the training of physics-informed neural networks (PINN) (Bahmani and Sun, 2021) and fracture mechanics problems (Liu et al., 2021; Perera and Agrawal, 2022). However, to the best of the authors' knowledge, no comprehensive attempt has been made to apply transfer learning for the constitutive modelling of granular materials.

In this study, we aim to unleash the potential of transfer learning for material modelling through two distinct strategies. The first strategy attempts to integrate phenomenological constitutive models with data-driven multiscale modelling. Specifically, the deviatoric hardening (DH) model is used as a prototype model to generate synthetic material response data. Deep learning models are first trained using the synthetic data before being transferred to new tasks based on high-fidelity stress-strain pairs from particle-scale DEM simulations. The second strategy aims to explore the transfer of knowledge across different materials, i.e. a base model is trained with one type of material first, and then the trained model is repurposed to learn the material behaviour of other materials.

The remainder of this study is organised as follows. Section 2 gives a brief revisit on transfer learning and its application to multiscale modelling, followed by a detailed investigation of using a phenomenological constitutive model for data-driven material modelling in Section 3. Section 4 explores the knowledge sharing across distinct granular materials based on DEM simulation data of true triaxial testing. In Section 5, the value of transfer learning in data-driven multiscale modelling is instantiated by embedding the deep learning-based material model in FEM computations. The significance and limitations of transfer learning-based constitutive modelling are discussed in Section 6, and the concluding remarks are summarised in Section 7.

2. Transfer learning for constitutive modelling

2.1. Fundamentals of transfer learning

Transfer learning is a training strategy in machine learning that aims to enhance the predictive performance of target learners on specific domains by leveraging knowledge acquired from previously trained models on related domains. The concept of transfer learning can be defined in terms of domains and tasks. A domain D consists of feature space χ and a marginal probability $\mathbb{P}^X$ of the input X, where $X = \in \chi$. A task is composed of a label space $\mathscr{Y}$ and a surrogate function f: $\chi \rightarrow \mathscr{Y}$. The output $Y=\{y_1, \ldots, y_n\} \in \mathscr{Y}$. The mapping f serves to forecast the output $f(X)$, which is learned from training datasets consisting of pairs $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$. Suppose we have a source domain $D_s = \{(X_{si}, Y_{si})\}_{i=1}^{n_s}$, and a target domain $D_t = \{(X_{ti}, Y_{ti})\}_{i=1}^{n_t}$, where $X_{si} \in \chi_s$ *and* $X_{ti} \in \chi_t$ are the inputted data and $Y_{si} \in \mathscr{Y}_s$, $Y_{ti} \in \mathscr{Y}_t$ are corresponding outputs, respectively. Normally, $0 \leq n_t \leq n_s$. Transfer learning can improve the target mapping $f_t(\cdot)$ in D_t by employing the knowledge learned from D_s. A schematic procedure for transfer learning is shown in Fig. 2.

For artificial neural networks, the knowledge learned from the source data is stored in the weights and biases parameters of a trained model. A common approach in transfer learning is by freezing the weights and biases of the trained model and then unfreezing specific layers and/or adding a few layers for retraining on the new datasets. "Freezing" refers to making parameters untrainable, thereby preserving the previously learned information during future training. This transfer learning scheme is inspired by the finding that the first few layers of CNNs learn some general features that are common across all domains, and thus there is no need to learn the same information again while training on similar data. The number of layers to freeze and/or fine-tune depends on the similarities between the source and target deep-learning models.

When no general patterns have been learned in the source model, freezing the weight parameters may hinder the learning for the target model. In such situations, it is advisable to re-train the entire neural network. Unlike an unlearned model with randomly initialised variables, a pre-trained neural network has been tuned based on similar source data. These pre-trained weights and bias parameters provide a much better starting point for the target model, leading to a faster training convergence and a lower risk of getting stuck in local optima. This type of transfer learning is referred to as "full model fine-tuning" in this study.

2.2. Two strategies of transfer learning for constitutive modelling

As shown in Fig. 2, the premise of transfer learning can be summarised in three aspects: (1) Task A and Task B share the same input

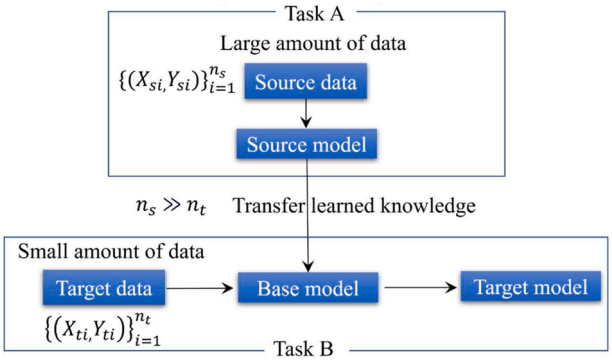


Fig. 2. A schematic procedure for transfer learning.

X; (2) much more data is available for Task A than for Task B, and (3) the mapping between the two tasks are similar, and the knowledge learned from Task A can help learn Task B. Here two strategies of transfer learning for multiscale modelling are explored.

The first application involves knowledge extraction and transfer from phenomenological constitutive models to data-driven modelling with new data. Although phenomenological models are based on certain empirical assumptions and thus the generated strain-stress data pairs may not accurately capture the actual material behaviour, they can still provide relatively reliable approximations. Using the data generated from these models, a deep learning model can be pre-trained and further used as the basis for developing a high-fidelity model with new data.

The second application of transfer learning is that the data-driven models trained on one type of material can be repurposed for other types. In real-life scenarios, it is often the case that we may have abundant data for certain types of materials while very little data for other types. For example, the stress-strain data of granular sands on Earth can be tested and collected readily, while data for lunar sands are fairly scarce. Despite variations in stress-strain responses amongst different materials, they are expected to follow certain underlying material laws. Thus the learned knowledge from one type of material will be useful for the constitutive modelling of other similar materials.

3. Constitutive knowledge transfer from phenomenological models for data-driven modelling

Phenomenological models are probably the most well-studied approach for formulating the constitutive behaviour of granular materials. A remarkable advantage is their ability to quickly compute the stress/strain response for a specific strain/stress path. In this section, we employ a deviatoric hardening model to generate a substantial amount of stress-strain data by numerically integrating stress-controlled conventional triaxial strain paths. The basic formulation for the deviatoric hardening model can be found in Appendix A. A data-driven model is trained using artificially generated data and is subsequently re-trained using data from virtual triaxial testing. Several learning schemes based on both transfer learning and committee-based active learning are explored and compared. Note that active learning can prioritise selecting the most informative specimens to train a well-behaved model from a labelled data pool. More details on active learning can refer to our previous work (Qu et al., 2023).

3.1. Parameter calibration of macroscopic material parameters based on DEM simulations

Considering that the effectiveness of transfer learning directly relies on the similarity between the source and target data, the free parameters of the DH model are calibrated using DEM simulations, aiming to generate synthesised source data that aligns with the DEM target data as closely as possible. In this section, the particle-scale parameters used for DEM modelling in this study are the same as those described in Qu et al. (2021a). Small-strain isotropic compression tests and true-triaxial tests with a constant mean principal stress are performed to determine K and G, respectively. The elastic volumetric and deviatoric strain increments can be calculated by:

$$\begin{cases} dp = Kd\varepsilon_V^e \\ dq = 3Gd\varepsilon_d^e \end{cases} \tag{1}$$

where dp, dq, $d\varepsilon_V^e$ and $d\varepsilon_d^e$ are increments of mean stress, deviatoric stress, elastic volumetric strain, and deviatoric strain, respectively; and K and G are the elastic bulk modulus and shear modulus, respectively. Note that K and G are dependant on the void ratio of the specimen and the mean effective stress. In this case, the specimens with the same void ratio but different confining pressures are considered. The general form of K and G in the elastoplastic constitutive model proposed by Guo and Li (2008) is adopted as it accurately captures the pressure-dependence of K and G:

$$K = K_0 P, \;\; G = G_0 P_a \left(\frac{\sigma_3}{p_a}\right)^n \tag{2}$$

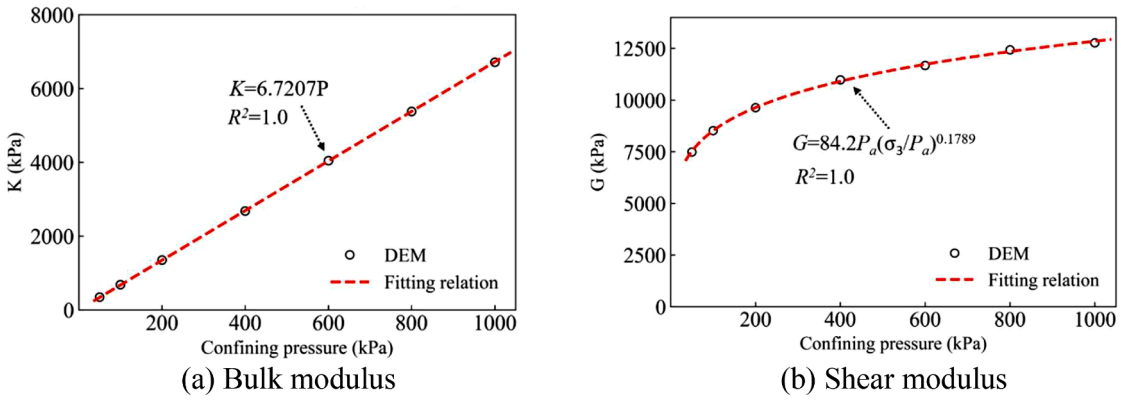


Fig. 3. Determination of elastic deformation parameters.

where p, σ_3 and p_a are the mean stress, minor principal stress, and standard atmospheric pressure (=101.4 kPa), respectively; and K_0, G_0 and n are constants with values found to be 6.7207, 84.2, and 0.1789, respectively, as shown in Fig. 3.

Conventional triaxial tests are simulated to determine other hardening parameters in DH constitutive models. The hardening parameter A is fitted from the η-ε_q relation (see Fig. 4a). The parameter η_c (slope of the zero-dilatancy line) is determined by fitting the ratio of q/p under different confining pressures at the zero-dilatancy point where the volumetric deformation transits from compaction to dilatancy. The parameter η_f (slope of the failure line) is obtained from the $p-q$ evolution under different confining pressures, as shown in Fig. 4b.

3.2. Transfer learning from phenomenological models to virtual triaxial experiments

3.2.1. Generating stress-strain sequences via numerical integration

In an elastoplastic model, the path-dependant constitutive responses are not analytically integrable. A numerical integration is normally required to compute the resulting mechanical responses when the materials are subjected to a specified stress/strain history. Here our objective is to generate a series of stress-strain curves with the stress and strain states defined in $\{p, q\}$ and $\{\varepsilon_v, \varepsilon_q\}$ spaces, respectively. Taking a stress-controlled conventional triaxial loading path as an example, the strain responses along some specified stress paths can be determined as follows:

$$\boldsymbol{\varepsilon} = \int_t \dot{\boldsymbol{\varepsilon}} dt = \int_{\boldsymbol{\varepsilon}} d\boldsymbol{\varepsilon} = \int_{\boldsymbol{\varepsilon}} (d\boldsymbol{\varepsilon}^e + d\boldsymbol{\varepsilon}^p) = [\boldsymbol{C}^e]\boldsymbol{\sigma} + \frac{1}{H_p}\int_{\boldsymbol{\sigma}} \left(\frac{\partial \psi}{\partial \boldsymbol{\sigma}}\right)^T \frac{\partial f}{\partial \boldsymbol{\sigma}} d\boldsymbol{\sigma} \tag{3}$$

where an incremental strain $d\boldsymbol{\varepsilon}$ can be decomposed into an elastic part $d\boldsymbol{\varepsilon}^e$ and a plastic part $d\boldsymbol{\varepsilon}^p$; $[\boldsymbol{C}^e]=[\boldsymbol{D}^e]^{-1}$ is the elastic compliance matrix; and H_p, ψ and f are the plastic hardening modulus, plastic potential function, and yield function, respectively, as defined in Appendix A.

The model adopts a deviatoric hardening law, i.e., using plastic distortions for the characterisation of hardening effects, and a non-associated flow rule to formulate stress-strain relations of materials. Considering an initial active loading history where the deformation process is described by the evolution of the bounding surface $f(p^{(n)}, q^{(n)}, \varepsilon_q^{p^{(n)}}) = 0$, the algorithm for a stress-controlled explicit integration can be found in Algorithm 1.

Using this algorithm, we generate a series of stress-strain curves of the granular materials and compare them with the results of DEM simulations of triaxial testing, as shown in Fig. 5. All numerical integrations are performed on the p-q space following stress-dominated monotonic triaxial loading paths. The increments of mean stress (dp) and deviatoric stress (dq) are used as independent variables. The corresponding increments of volumetric and deviatoric strain (dε_v and dε_q) are outputs, with dp=0.3 and dq=0.9 being step sizes during the numerical integration. The iteration continues until ε_q is larger than 0.25 or a total of 2500 iteration steps are conducted.

Note that the DH model is one of the phenomenological elastoplastic models using the deviatoric strain as an internal variable to infer the evolution of hardening laws. Its formulation is constructed upon a set of assumptions derived from experimental observations. In contrast, the DEM model simulates macroscopic stress-strain and volumetric strain behaviour based on particle-scale simulations. The strain localisation at the critical state is naturally captured (Qu et al., 2019). It is understandable that the DEM data and the DH data will differ slightly. Despite some discrepancies, the results show that the stress-strain sequences obtained from the DEM simulations and DH models are in satisfactory agreement. Subsequently, we leverage the DH model to generate a large number of stress-strain sequence pairs by incrementally increasing the magnitude of the confining pressure. A total of 450 groups of sequences are generated, as depicted in Fig. 6.

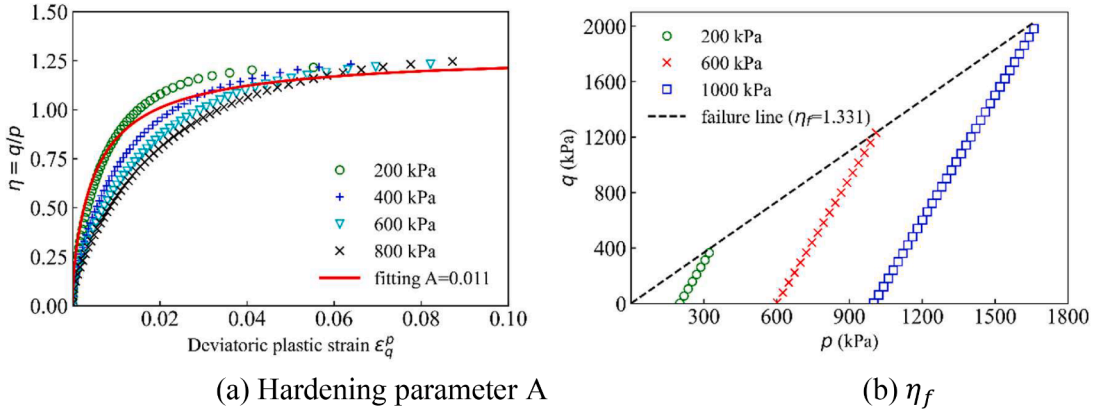


Fig. 4. Determination of hardening parameters A and η_f.

Algorithm 1
Stress-controlled explicit integration procedures with deviatoric hardening.

Known $\{p, q, \varepsilon_v, \varepsilon_q, \varepsilon_v^p, \varepsilon_q^p\}$ at the step n, and $f(p^{(n)}, q^{(n)}, \varepsilon_q^{p^{(n)}}) = 0$

1. Given: $\{\Delta p, \Delta q\}$
2. Update: $p^{(n+1)} = p^{(n)} + \Delta p$, $q^{(n+1)} = q^{(n)} + \Delta q$
3. Evaluate hardening moduli:

$$\eta^{(n+1)} = q^{(n+1)}/p^{(n+1)}$$

$$H_e = \frac{\partial f}{\partial p} K \frac{\partial \Phi}{\partial p} + \frac{\partial f}{\partial q} 3G \frac{\partial \Phi}{\partial q} = \eta^{(n+1)}(\eta^{(n+1)} - \eta_c)\, K + 3G$$

$$H_p = -\frac{\partial f}{\partial \varepsilon_q^p} \frac{\partial \psi}{\partial q} = \frac{p^{(n+1)}(\eta_f - \eta^{(n+1)})^2}{A\eta_f}$$

4. Compute strain increments:

$$\begin{bmatrix} d\varepsilon_v \\ d\varepsilon_q \end{bmatrix} = \begin{bmatrix} \left(K - \frac{\frac{\partial f}{\partial p} K^2 \frac{\partial \Phi}{\partial p}}{H_e + H_p}\right) & \left(-\frac{\frac{\partial f}{\partial p} 3GK \frac{\partial \Phi}{\partial q}}{H_e + H_p}\right) \\ \left(-\frac{\frac{\partial f}{\partial p} 3GK \frac{\partial \Phi}{\partial q}}{H_e + H_p}\right) & \left(3G - \frac{\frac{\partial f}{\partial q} 9G^2 \frac{\partial \Phi}{\partial q}}{H_e + H_p}\right) \end{bmatrix}^{-1} \begin{bmatrix} dp \\ dq \end{bmatrix}$$

5. Update strain values:

$$\varepsilon_v^{p^{(n+1)}} = \varepsilon_v^{p^{(n)}} + d\varepsilon_v, \quad \varepsilon_q^{p^{(n+1)}} = \varepsilon_q^{p^{(n)}} + d\varepsilon_q$$

6. Set n=n+1; go back to step 1 for the next iteration

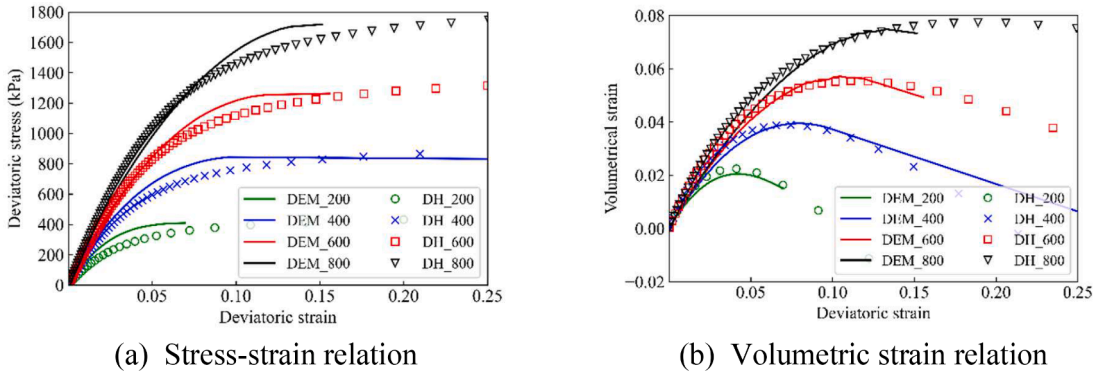


Fig. 5. A comparison between DH models and DEM simulations.

3.2.2. *Transfer learning from phenomenological models to discrete element modelling*

The data acquired from the DH constitutive model is employed as preliminary datasets to train a base GRU network. The fundamentals of GRU can be found in Qu et al. (2021a). Deviatoric strain and confining pressure serve as the inputs to the network, while deviatoric stress and volumetric stress are the outputs. In the pretraining stage, the available data are partitioned to training and

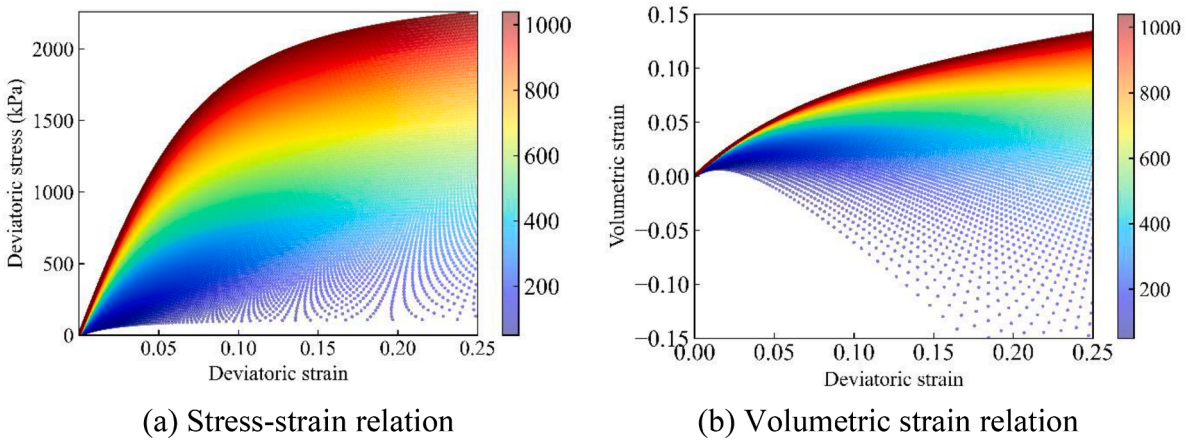


Fig. 6. Artificial stress-strain sequence pairs produced from the DH model.

validation sets with a proportion of 80 % and 20 %, respectively.

A total of 40 groups of stress-dominated triaxial loading cases, with confining stress ranging from 25 kPa to 1000 kPa (an interval of 25 kPa) are simulated via DEM, for later learning. Three transfer learning schemes, including (i) freezing+fine tuning, (ii) full model fine-tunning or pre-training, and (iii) transfer learning+active learning, and one contrastive scheme without transfer learning, are designed to examine the capability of different transfer learning strategies. The idea of "transfer learning+active learning" is explained as follows. A base model is initially trained using the data from phenomenological models and is then transferred to the new training by leveraging active learning to prioritize the selection of DEM data. The forecast accuracy of a deep neural network (DNN) is evaluated by the mean squared error (MSE), which is computed over the entire stress-strain paths by:

$$MSE = \frac{1}{N^{ij}} \sum_{j=1}^{N^j} \sum_{i=1}^{N^i} \left(y_{ij}^{True} - y_{ij}^{Prediction} \right)^2 \tag{4}$$

where y_{ij}^{True} and $y_{ij}^{Prediction}$ are the actual and predicted values of the i^{th} data point of the j^{th} stress-strain curve (or path), respectively; and N^{ij} is the total number of data points of all the stress-strain curves involved. The hyperparameters are selected by a trial-and-error method, and the final combination is presented in Table 1. A total of 6121 trainable parameters are used, with 5280, 820, and 42 parameters for the GRU layer, dense layer, and output layer, respectively. As a total of 20 GRU units are used, only 20 internal memory variables (vectors) are employed. A schematic representation of the GRU model can be found in Appendix B. The size of the moving window, also called the timesteps, refers to the number of historical sequential points considered in each input sequence. The *sigmoid* activation function is applied to both the GRU and dense layers, and the linear activation function is used for the output layer. The network is trained using the Adam optimizer. In the transfer learning stage, the training number is progressively selected at each training case with the remaining data being partitioned with a proportion of 20 % and 80 %, respectively for validation and testing.

Fig. 7 illustrates the average MSE over the unseen data as the number of strain-stress paths used for training increases. A lower MSE represents a better prediction performance. The training schemes with transfer learning remarkably outperform those without transfer learning. In addition, the "freeze+fine tuning" scheme outperforms "pre-training" when only a small number of training data is adopted, whereas the situation is reversed when more training data are available. The reasons behind the phenomenon are that the freezing method permanently inherits the knowledge of the DH model and thus has a good prediction capability at the preliminary stage. However, with more training numbers, these fixed weights and biases may hinder the improvement of the model, demanding more layers to be added to these neural networks. Note that only the weight and bias parameters in the GRU layer have been frozen in the "freeze+fine tuning" case. The transfer learning plus active learning scheme achieves the best performance amongst the examined schemes. Note that one premise of using active learning is that we should know the loading paths (unlabelled inputs, either strain or stress paths). Thus it is recommended to add active learning to transfer learning schemes, provided that the strain paths are known *a priori* or can be artificially constructed.

Figs. 8 and 9 depict the stress and volumetric forecasts of the four transfer learning schemes on stress-dominated triaxial testing paths with varied confining pressures when only 6 groups of training datasets are available. The predictive performance of all models is consistent with the MSE results shown in Fig. 7. The non-transfer learning scheme, although appearing to capture the primary material behaviour, exhibits significantly lower prediction accuracy than those with transfer learning. The results also demonstrate that the pre-training transfer learning outperforms the "freeze+fine tuning" scheme.

Another interesting observation is that all the cases without active learning fail to match the stress responses under the higher confining pressure cases. One possible reason is that training specimens are randomly selected, while the pressure-dependant material nonlinearity under the higher confining pressure may belong to extrapolation if these data points are not selected for training. In contrast, active learning can discover the most hard-to-predict specimens and include them in the training datasets. This speculation is supported by Fig. 10, which demonstrates that active learning prefers the data points at the boundary in the data space.

An apparent disparity in the outcomes, particularly the volumetric responses of DEM and GRU models, is observed in the large deviatoric strain stage. The phenomenon can be attributed to the data imbalance issue due to a limited number of data points available at that stage, because a minor stress increment in the critical stage can even lead to a substantial increase in strain in our implementation of a stress-dominated triaxial loading case. This data imbalance-caused training issue arises from the fact that typical loss functions used in supervised machine learning, such as MSE in this study, calculate the average results across all training data. Consequently, the learned model encounters greater difficulty in capturing patterns within the small data regime. Possible solutions include active learning, synthetic minority oversampling technique (SMOTE), and custom loss functions that consider different weights for the uneven distribution of training data. In this study, the capability of active learning in addressing the data imbalance problem has been confirmed as shown in Figs. 8c and 9c.

Table 1
Network architecture and key hyperparameters.

Item	Value
Network architecture	GRU:40-Dense:20
The moving window size (timesteps)	20
Learning rate	0.01
Batch size	128
Epoch number	200

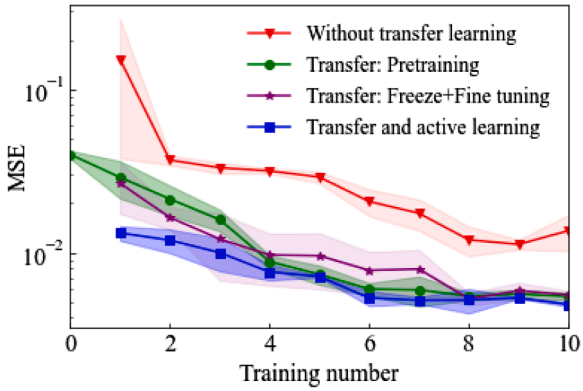


Fig. 7. Forecast performance of several training strategies over unseen test datasets.

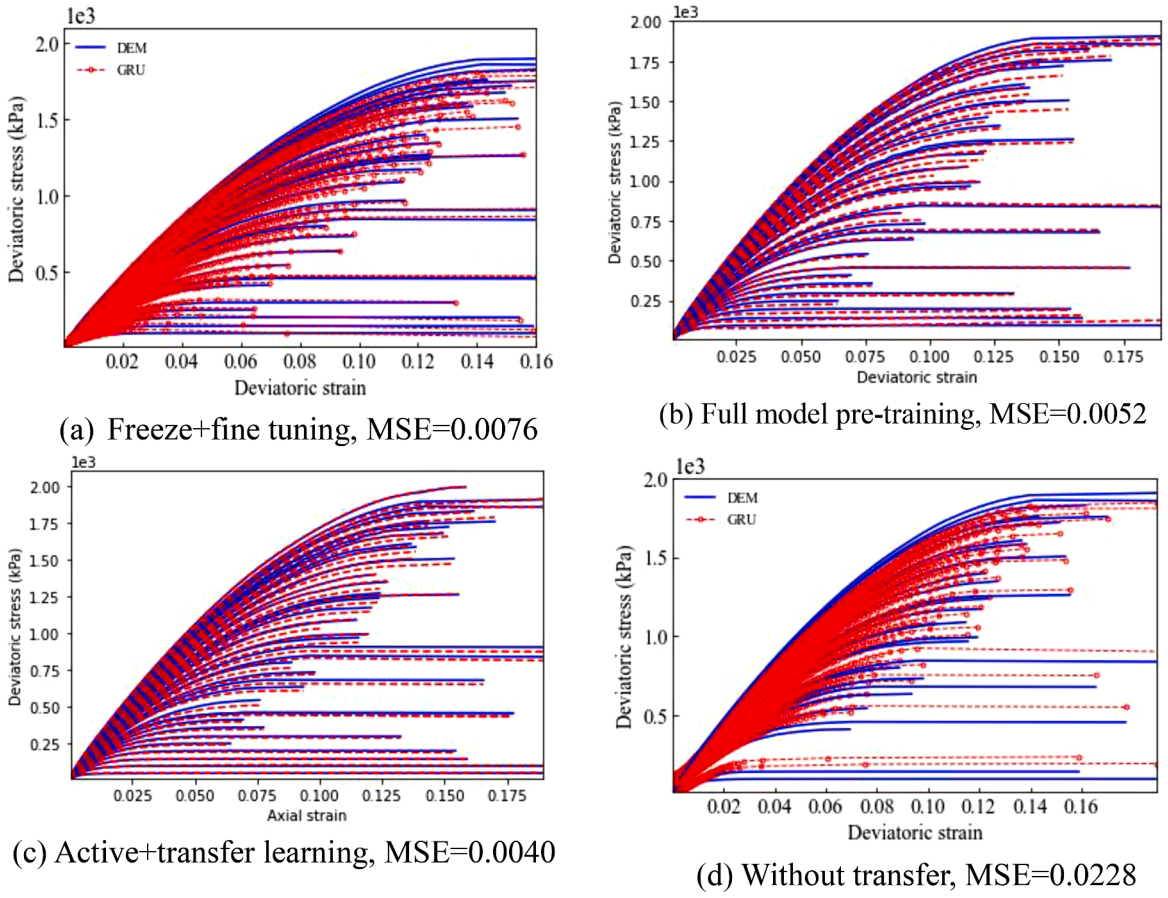


Fig. 8. Stress-strain predictions of stress-controlled triaxial testing with only 6 groups of training datasets.

Furthermore, although the GRU model can predict a wide range of strain paths satisfactorily, it is inherently incapable of properly predicting cases where the step sizes/intervals of inputs in the training datasets significantly differ from those in the forecasts. A straightforward but not fundamental method to improve the incremental size dependence in traditional recurrent neural networks (RNNs) is by artificially enriching strain paths with various strain intervals (Jung and Ghaboussi, 2006). The other method includes the linearised MSC (or LMSC) neural networks proposed by Bonatti and Mohr (2022). More investigations on this aspect will be pursued elsewhere.

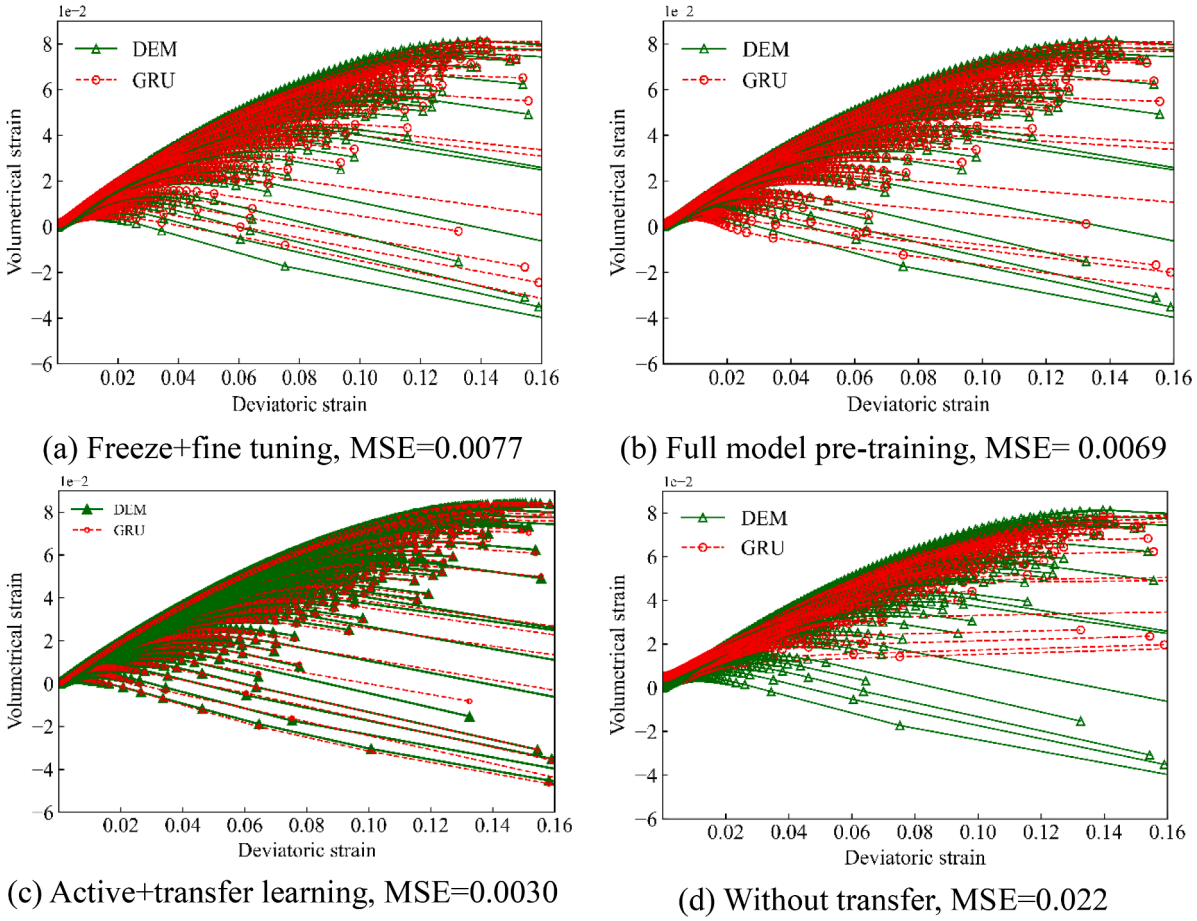


Fig. 9. Volumetric predictions of stress-controlled triaxial testing with only 6 groups of training datasets.

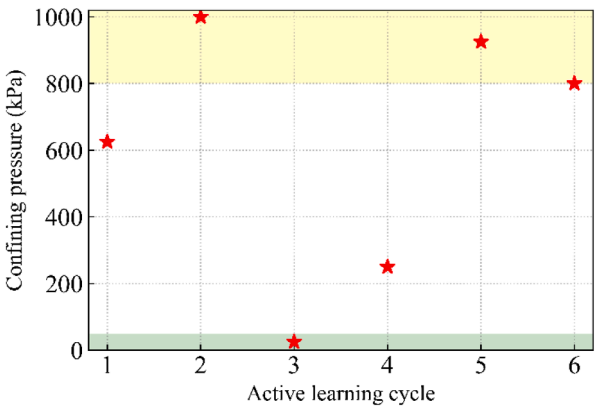


Fig. 10. The prioritised specimens via transfer and active learning.

4. Constitutive knowledge transfer across different materials

Transfer learning has the potential to facilitate the transfer of constitutive knowledge from one material to other materials, which can be valuable in the development of data-driven material models with small data sets. In this section, the effectiveness of transfer learning in using learned constitutive knowledge from similar materials to develop reliable material models will be investigated.

4.1. Discrete element modelling of granular materials with distinct gradation and loading paths

A series of true triaxial tests for virtual granular materials are simulated via DEM. Two types of granular materials with a total of 10,000 particles but different gradations and external loading paths are considered. Material 1 has a bell shape gradation with particle diameters ranging from 4.8 to 13.9 mm. A total of 60 groups of true triaxial loadings with a constant mean pressure (1 MPa, 2 MPa, 4 MPa) and a constant coefficient of middle principal stress (0, 0.25, 0.5, 0.75, 1.0) are considered. In contrast, Material 2 has a gradation ranging from 3.39 to 28.27 mm, subjected to a constant coefficient of middle principal stress (0, 0.25, 0.5, 0.75, 1.0), following a constant minor principal stress (1 MPa, 2 MPa, 4 MPa) as well. The two specimens and their gradation curves can be found in Fig. 11, and their material properties are listed in Table 2. The principal stress and volumetric responses of both specimens are demonstrated in Fig. 12.

4.2. Transfer learning from one material to other types of materials

To explore the transfer of knowledge across different materials, three learning schemes are investigated to obtain a well-behaved constitutive model for Material 2. Both volumetric and principal stress responses are considered in the analysis. The first scheme involves conventional learning using only the data from Material 2. The second scheme trains a base model using the data from Material 1 and then transfers this model to learn the data from Material 2. The knowledge learned from Material 1 is adopted to improve the performance of the model when applied to Material 2. The third scheme utilises committee-based active learning to prioritise the selection of training data from Material 2. The network architecture and hyperparameters of the DNN model are inherited from the one used in Section 3. The network structure requires 6544 trainable parameters in total, with 5640, 820 and 84 parameters for the GRU, dense, and output layers, respectively. In addition, the data split criterion utilized in Section 3 remains in use for this section.

The performance of the three learning schemes is demonstrated in Fig. 13. As the number of training specimens or the training number increases, the value of the loss function over testing data decreases significantly, especially when a small number of training specimens are used. The two schemes with transfer learning outperform the scheme without transfer learning. Also in the two schemes involving transfer learning, the one incorporating active learning performs better. These results highlight the effectiveness of transfer learning in achieving high-fidelity predictions with fewer training specimens. They also suggest that the network has acquired a certain knowledge of granular materials during its training on the first material, enabling it to learn the constitutive behaviour of other materials more easily through transfer learning.

Although the "transfer + active learning" scheme obtains the optimal performance, the application of active learning is limited to situations that the strain paths are known. Thus, when the strain paths can be artificially constructed, the active learning scheme is the

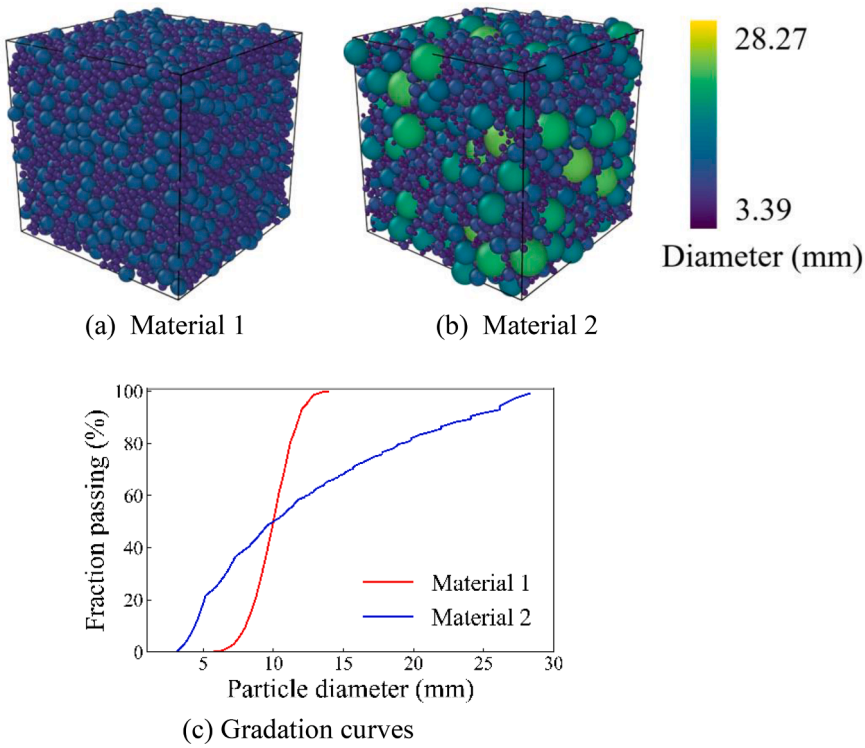


Fig. 11. Two different granular materials.

Table 2
Material properties.

Material properties	Value
Particle density (kg/m^3)	2600
Young's modulus (GPa)	0.8
Poisson's ratio	0.12
Friction coefficient	0.4
Restitution coefficient	0.95
Rolling coefficient	0.05

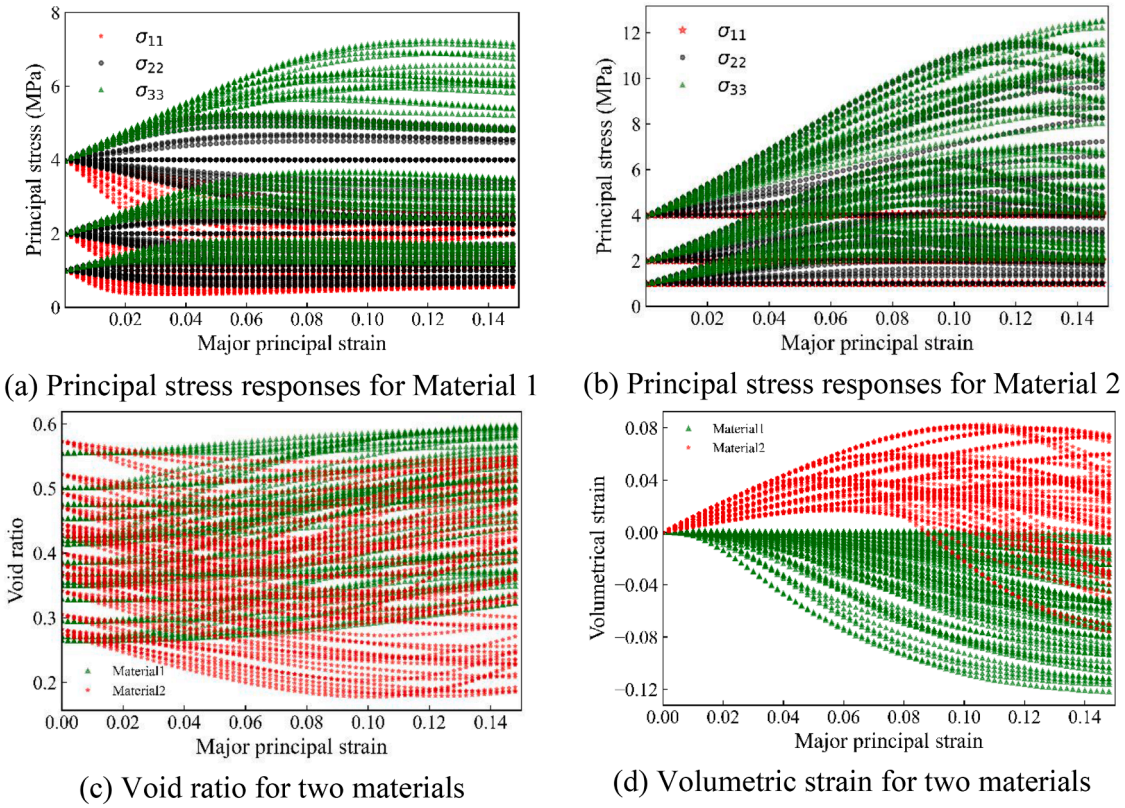


Fig. 12. Mechanical responses for the two different granular materials.

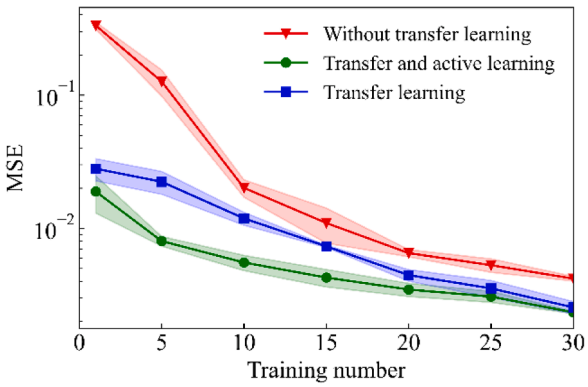


Fig. 13. Predictive performance for three transfer schemes across different materials.

preferred choice. However, in simulations of BVPs, it is often difficult to construct realistic path-dependant strain paths for granular materials.

The "transfer+active learning scheme" with 20 groups of training specimens is selected as an example to show the actual prediction performance of the trained DNN models. Fig. 14 demonstrates the selected training specimens and forecast performance for major principal stress and volumetric strain with increasing major principal (axial) strain. The colour map in Fig. 14(a) and (b) show the order (ranking) of the selected specimens for training the data-driven constitutive model. Fig. 14(a) suggests that the specimens with the largest and smallest confining pressures are preferred via the active learning algorithm, as these specimens represent a more challenging extrapolation when the original base models make forecasts. In contrast, the corresponding volumetric strain responses for these specimens do not significantly reside on the boundaries of the stress data. This is probably because the error indicator for prioritizing training specimens in the adopted active learning is equally weighted considering three principal stress components and one volumetric relation, and therefore the stress predictions have a higher priority in terms of improving the overall prediction performance.

The major principal stress and volumetric predictions are given in Fig. 14(c) and (d). The results show that the constitutive behaviour (i.e. stress hardening or softening, volumetric dilatancy or contraction) of granular materials with different initial states (varied confining pressures and void ratios) can be perfectly learned by the deep neural network. Note both interpolation and extrapolation are incorporated into Fig. 14(c) and (d). The capability of GRU models in capturing unloading-reloading strain paths has been confirmed by our previous work (Qu et al., 2021a, 2021b) and is thus not pursued here.

5. Transfer learning-assisted data-driven multiscale modelling

This section aims to investigate the application of deep transfer learning in data-driven hierarchical multiscale modelling (HMM) of granular materials, specifically within the computational framework which couples FEM and DEM (hereafter referred to as FD-HMM). This approach replaces phenomenological constitutive models in conventional FEM computations with DEM simulations, as proposed by Guo and Zhao (2014). By using HMM, we generate a substantial amount of stress-strain sequence pairs with diverse strain paths in BVPs. The resulting data serve to train deep learning-based constitutive models, which are subsequently integrated into macroscopic FEM computations, leading to accelerated HMM computations.

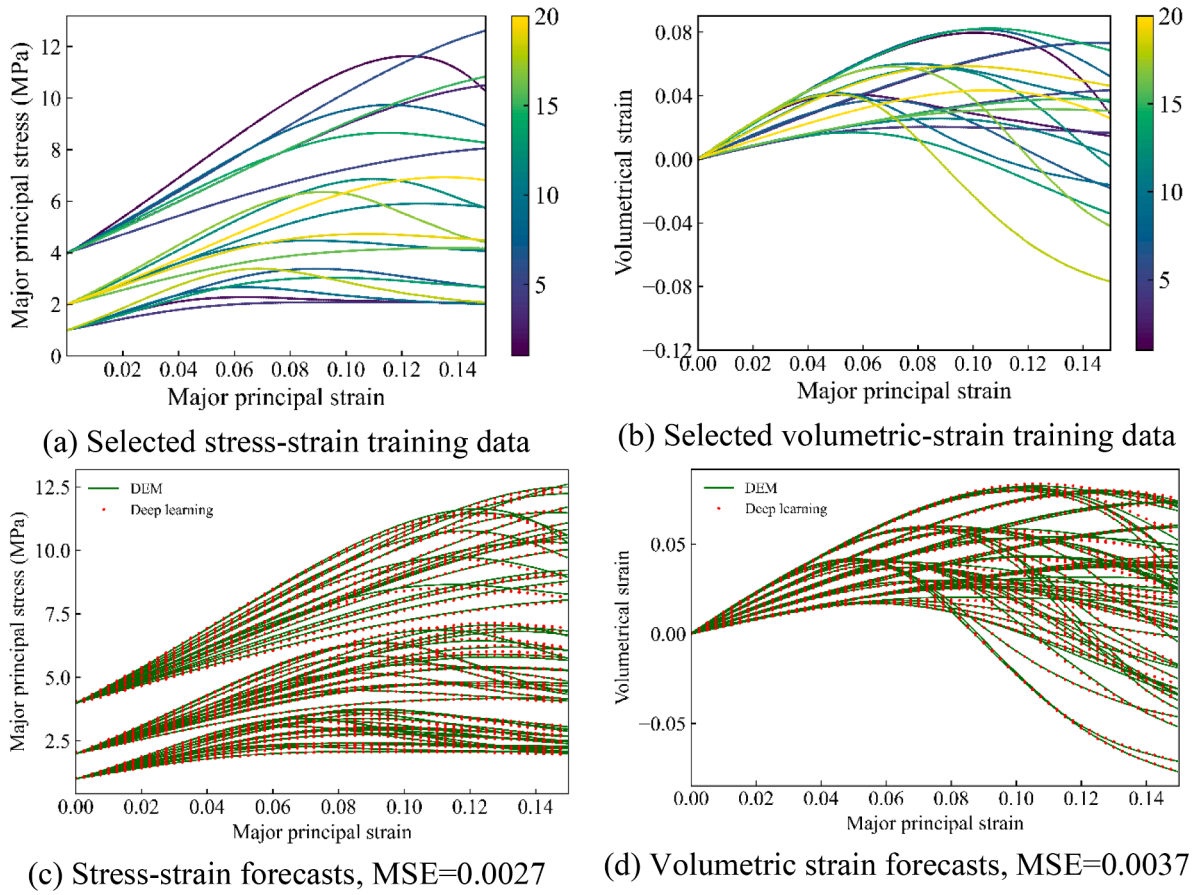


Fig. 14. Selected training specimens and forecast performance for the transfer+active learning scheme with 20 groups of training specimens.

5.1. Hierarchical multiscale modelling coupling FEM and DEM

The FD-HMM framework utilises FEM to simulate macroscopic BVPs, while the material constitutive relation is determined by particle-scale DEM simulations of RVEs. Fig. 15 depicts the solution procedure, wherein the global deformation at each Gauss point serves as boundary conditions for the corresponding RVE treated by DEM to solve and homogenise the resultant stress responses and tangent operators. These values are then transmitted back to the Gauss points for FEM computations.

In FEM, a Gauss point refers to a specific material point within a finite element to carry on the material properties. These points do not possess a physical volume but are essential for the evaluation of the mechanical responses at different locations of a physical domain. The term of RVE is used to describe the essential volume of particles to be modelled by the DEM to provide material responses at a Gauss point to feed the FEM solution. The RVE size is equivalent to the number of particles to be included in an RVE and is purely chosen to produce relatively reliable material responses while offering affordable computational costs. Once a DEM solution is found for an RVE under a specific boundary condition for a Gauss point, required macroscopic quantities, such as stress, strain, tangential stiffness, and fabric tensors, will be homogenized from the deformed RVE configuration to feed the Gauss point for FEM calculation, through either the assumption of constant deformation gradient (first-order) or constant strain gradient (second-order). In this sense, the scale of the macroscopic FEM domain is separated from the *meso*/micro-scopic RVE domain for the multiscale modelling, constituting a basic principle for the hierarchical multiscale approach (Geers et al., 2010).

The FD-HMM approach avoids the need for phenomenological assumptions in the constitutive relation of materials, but it requires solving a large number of RVEs with DEM. This feature presents a significant challenge to computational costs and limits the potential application of FD-HMM to engineering-scale BVPs. With the development of the data-driven material model, data-driven multiscale modelling offers an important alternative with significantly reduced computational costs (Karapiperis et al., 2021).

5.2. Data-driven multiscale modelling with an implicit FEM solver

RNNs are known for their capability to capture path-dependant material behaviour (Bonatti et al., 2022). However, as many trainable parameters may be involved, significant computational costs are often required for both training and prediction. Recent work has demonstrated that multi-layer perceptron (MLP) incorporating additional internal variables cannot only reproduce the stress-strain behaviour of materials satisfactorily but also improve inference efficiency. One strategy for introducing internal variables is to utilise the accumulation of absolute strain increments $\boldsymbol{\varphi}$, which provides an implicit description of the loading history (Guan et al., 2023c; Huang et al., 2020):

$$\varphi_j^{(t)} = \sum |\Delta \varepsilon_j^{(t)}| = \begin{cases} \left|\varepsilon_j^{(t)}\right|, t = 1 \\ \left|\varepsilon_j^{(1)}\right| + \sum_{t=2}^{(T)} \left|\varepsilon_j^{(t)} - \varepsilon_j^{(t-1)}\right|, \quad t \subseteq [2, T] \end{cases} \tag{5}$$

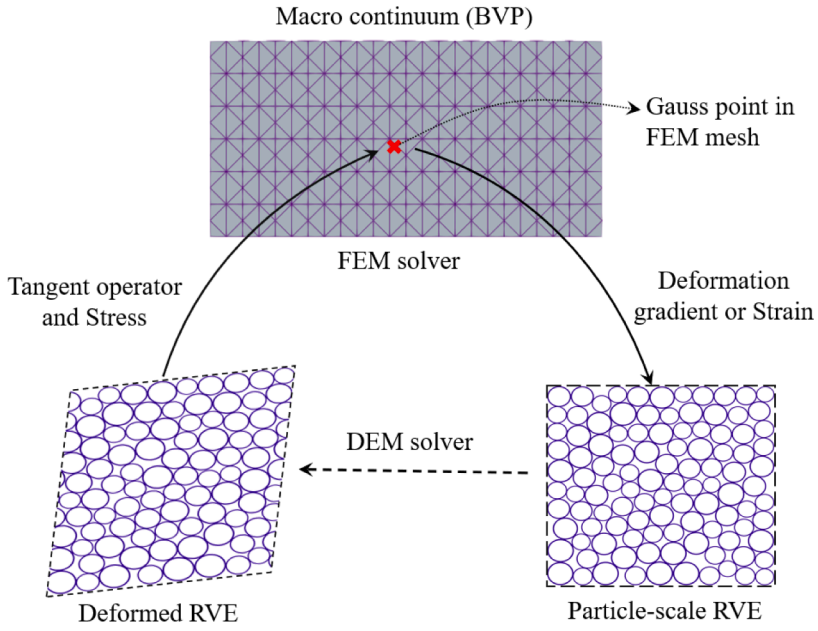


Fig. 15. Coupling procedure of FEM/DEM hierarchical multiscale modelling (adapted from Guo and Zhao 2014).

where $\varphi_j^{(t)}$, $\Delta\varepsilon_j^{(t)}$ and $\varepsilon_j^{(t)}$ represent the jth components of the internal variable, the strain increment, and the absolute strain, respectively, at the t^{th} time step $t \subseteq [1, T]$. In this study, an MLP is trained as a surrogate constitutive model in numerical computations.

For implicit FEM computations, the tangent operator **D** must be calculated at each Gauss point to determine the element stiffness matrix $\mathbf{K}^{\text{e}}$:

$$\boldsymbol{K}^e = \int_{\Omega^e} \mathbf{B}^T \mathbf{D} \mathbf{B} \mathrm{d}\Omega^e \tag{6}$$

where **B** represents the gradient of the element shape functions; Ω^{e} refers to the element domain. Within the FD-HMM framework, the tangent operator **D** can be estimated using the microstructure and particle-scale stiffnesses of the RVE in DEM through various homogenisation techniques. An analytical expression for the **D** matrix based on Vogit's homogenisation is provided below:

$$\mathbf{D} = \frac{1}{V} \sum_{N_c} (k_n \mathbf{n}^c \otimes \mathbf{d}^c \otimes \mathbf{n}^c \otimes \mathbf{d}^c + k_s \mathbf{t}^c \otimes \mathbf{d}^c \otimes \mathbf{t}^c \otimes \mathbf{d}^c) \tag{7}$$

where V denotes the volume of the RVE; N_c represents the total number of mechanical contacts within the RVE specimen; k_{n} and k_{s} refer to normal and tangential contact stiffnesses between particles, respectively; $\boldsymbol{n}^c$ and $\mathbf{t}^c$ denote unit vectors in the normal and tangential directions of a contact plane, respectively; $\mathbf{d}^c$ represents the branch vector connecting the centroids of two contacting particles.

To enable data-driven multiscale modelling with an implicit FEM solver, **D** should be incorporated as an additional prediction. Instead of training a large model with both stress responses and tangent operator **D** as outputs, we adopt two independent networks, i.e. NN1 for stress and NN2 for **D** predictions as follows:

$$\widehat{\sigma}^{(t)} = \mathcal{NN}_1\left(\epsilon^{(t)}, \boldsymbol{\varphi}^{(t)}\right) \tag{8}$$

$$\widehat{\mathbf{D}}^{(t)} = \mathcal{NN}_2\left(\epsilon^{(t)}, \boldsymbol{\varphi}^{(t)}\right) \tag{9}$$

where $\widehat{\boldsymbol{\sigma}}^{(t)}$ and $\widehat{\mathbf{D}}^{(t)}$ denote the predicted stress responses and tangent operator at the time t, respectively, for the given (total) strain tensor $\epsilon^{(t)}$ and internal variables $\boldsymbol{\varphi}^{(t)}$.

5.3. Transfer learning for data-driven multiscale modelling

5.3.1. Data-driven multiscale modelling of biaxial shear tests

This part focuses on embedding data-driven constitutive models into FEM simulations of BVPs to examine the capability of transfer learning in multiscale modelling. Specifically, biaxial shear testing of granular soils is simulated with FEM-DEM hierarchical multiscale modelling techniques. The test specimen has a dimension of 0.5 m × 1 m and is discretised into 8 × 16 eight-node quadrilateral elements as shown in Fig. 16. A frictional loading condition is implemented on the specimens so that the horizontal movement at the end

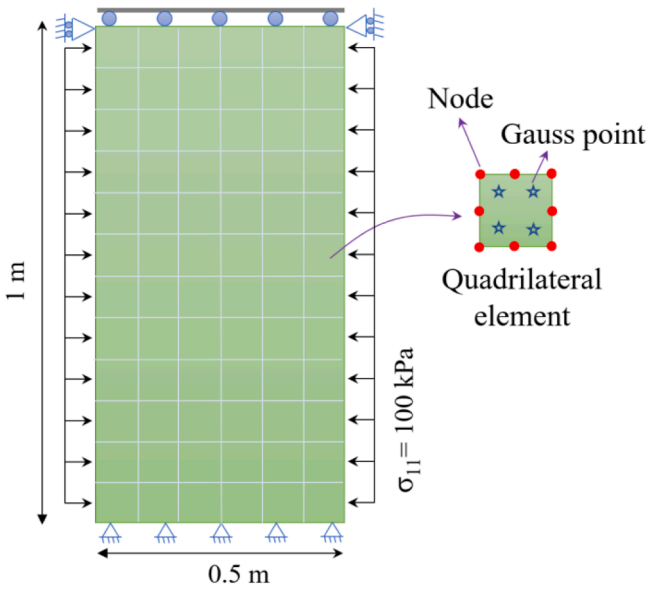


Fig. 16. FEM mesh partition for biaxial shear testing with rough loading ends.

of the specimen is not allowed. A total of 288 Gauss points are involved with the same number of identical DEM RVEs assigned to the simulation. Previous research has demonstrated that at least 400 particles in a 2D RVE should be incorporated to represent the material behaviour of granular media properly (Guo and Zhao, 2014). Here a total of 500 particles are considered in the RVE specimen, with each being isotropically consolidated to a mean effective pressure of 100 kPa. Two types of specimens with different material properties are prepared with the detailed parameters given in Table 3.

To examine the role of transfer learning in such data-driven multiscale modelling, three different modelling cases are considered:

- Case 1: FD-HMM simulations
- Case 2: Data-driven multiscale modelling of M2 with material knowledge transferring from M1
- Case 3: Data-driven multiscale modelling of M2 without knowledge transfer

For the FD-HMM simulations, it requires around 112 min to finish a total of 100 loading steps to reach a final axial strain of 12 %, accelerated by parallel computing on a server computer with 12 identical CPU processors. The computation yields a total of 288 strain-stress sequences associated with the material behaviour. During non-linear iterative solution procedures, a variety of Newton-Rapson iterations, ranging from several to dozens of updates, may be required. In this study, only the final or converged strain-stress and strain-**D** pairs at each loading step are leveraged for training models, which greatly reduces data redundancy and computational costs of training networks.

As using an excessive number of learnable parameters often causes problems such as overfitting and computational inefficiency, we adopt relatively small network architectures to train material models in Case 2 and Case 3. The architectures for NN1 and NN2 are [5-5-5-3] and [5-8-8-6], resulting in a total of 78 and 174 learnable weight and bias parameters, respectively. These parameters are optimised with Adam optimiser with a learning rate of 0.001. One million training epochs with an early-stopping mechanism are performed. *Sigmoid* and *linear* activation functions are used for the middle layers and output layer of the networks, respectively. The entire datasets are split as training/validation/testing sets following a 70 %, 15 %, and 15 % proportion. All the computations and training of neural networks are conducted in an Intel Core i7-8700 CPU 3.20 GHz x12. It takes around 16.43 and 23.9 min to train the strain-stress and strain-**D** networks, respectively. When embedding these networks in FEM computations, around 3.71 min are required to perform a total of 100 loading steps in the biaxial testing simulations.

5.3.2. *Comparative analysis of simulation results*

All three computational cases are successfully executed without encountering any convergence issues. Deviatoric strain is a measure of the deformation of a soil specimen under shear stress. For the 2D case, the volumetric strain ε_v, deviatoric strain tensor $\boldsymbol{\varepsilon}_{dev}$, and deviatoric strain ε_{q} are calculated as below:

$$\varepsilon_v = tr(\boldsymbol{\varepsilon}) \tag{10}$$

$$\boldsymbol{\varepsilon}_{dev} = \boldsymbol{\varepsilon} - \frac{1}{2} tr(\boldsymbol{\varepsilon}) * \mathbf{I} \tag{11}$$

$$\varepsilon_{\mathrm{q}} = \sqrt{2\boldsymbol{\varepsilon}_{dev} : \boldsymbol{\varepsilon}_{dev}} \tag{12}$$

where $\boldsymbol{\varepsilon}$ and $\mathbf{I}$ are the strain tensor and identity tensor, respectively.

Fig. 17 presents a comparison of the distribution of accumulated deviatoric strain during biaxial shear testing. The emergence and development of strain localisation can be identified. The results indicate that all three cases can reproduce the strain localisation phenomenon, while the transfer learning case significantly outperforms the one without knowledge transfer. Some discrepancies are indeed observed at the beginning of strain localisation (axial strain $\varepsilon_{22} = 0.03$). Here two cross-shear bands emerge in both the coupled FEM/DEM simulation and the transfer learning-based FEM simulation, whereas only a single shear band is observed for the data-driven multiscale modelling without any pre-trained knowledge. As the axial strain increases, another shear band emerges in the case without transfer learning, but the two ends of the band are not coaxial. The intensity of the two crossing bands also appears imbalanced. In contrast, the transfer-learning-based FEM simulation case closely resembles the FEM/DEM modelling throughout the entire loading process.

Fig. 18 displays the contours of the stress normal defined by $\sqrt{\boldsymbol{\sigma} : \boldsymbol{\sigma}}$ for the biaxial shearing simulation in the three cases. Although the three cases show comparable results, the distribution of stress intensity for the FEM simulation without transfer learning differs slightly from the other two cases, especially in the middle of the sheared specimen.

Fig. 19 depicts the macroscopic sample responses in terms of top-end reaction force versus the volumetric strain of the specimen. The evolution of the top force for all three cases exhibits a similar overall pattern and the magnitudes of axial strain when the top forces

Table 3
Material properties for the two RVE specimens.

Material type	Young's modulus (MPa)	Poisson's ratio	Density (kg/m^3)	Frictional angle	Radius (m)
M 1	200	0.3	2600	33	0.003–0.007
M 2	800	0.12	2600	21.8	0.0045–0.0055

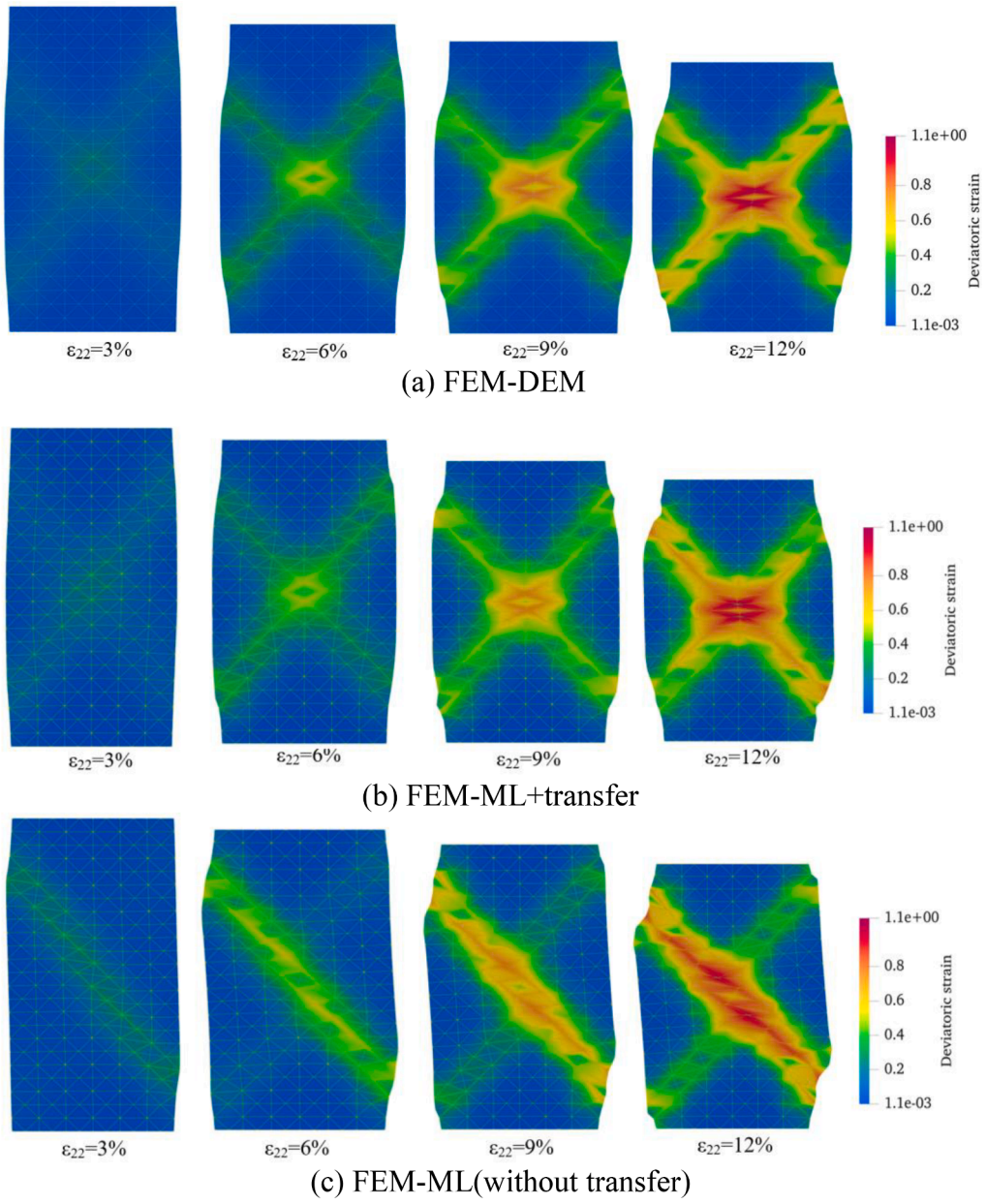


Fig. 17. Contours of the accumulated deviatoric strain.

reach their respective peak value are also similar to each other. However, it appears that the variation of top force at the initial stage ($\varepsilon_{22} < 0.03$) for the machine learning cases slightly deviates from the FEM/DEM simulation. For the evolution of volumetric strain, the transfer learning-based FEM simulation perfectly matches the result from FEM/DEM simulation, until the axial strain reaches around 6 %. However, both deep learning-based FEM simulations deviate from the FEM/DEM modelling when the axial strain exceeds 6 %. At this stage, significant strain localisation (see Fig. 17) occurs, which imposes a challenge for deep learning models to accurately reproduce the actual material responses.

The results indicate that a pure data-driven surrogate constitutive model is not necessarily capable of reproducing the real material behaviour when embedding it into a BVP problem, while the model's predictive performance can be enhanced by incorporating pre-trained knowledge via transfer learning.

Fig. 20 illustrates the relationship between the major principal stress and principal strains experienced by Gauss points in the biaxial testing simulations. The stress-strain paths obtained from FEM-DEM and FEM-ML simulations exhibit marked differences. Specifically, the stress-strain paths in FEM-DEM multiscale modelling show distinct discrete features, while those in FEM-ML simulations appear more continuous. This disparity indicates that the deep learning-based surrogate constitutive models are obliged to

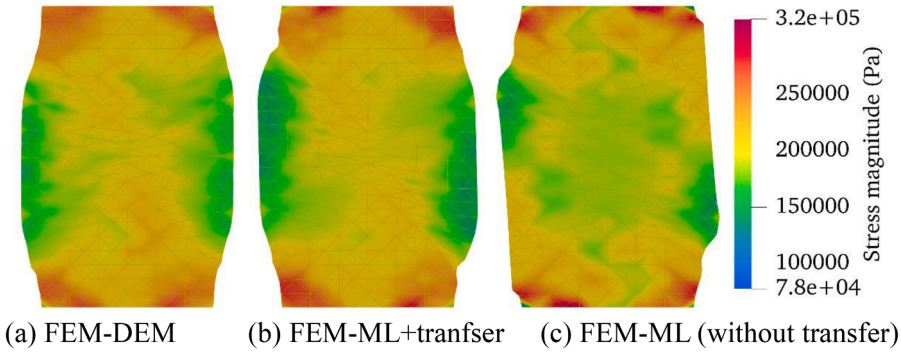


Fig. 18. Contours of the stress intensity for the biaxial specimens under a 12 % axial strain.

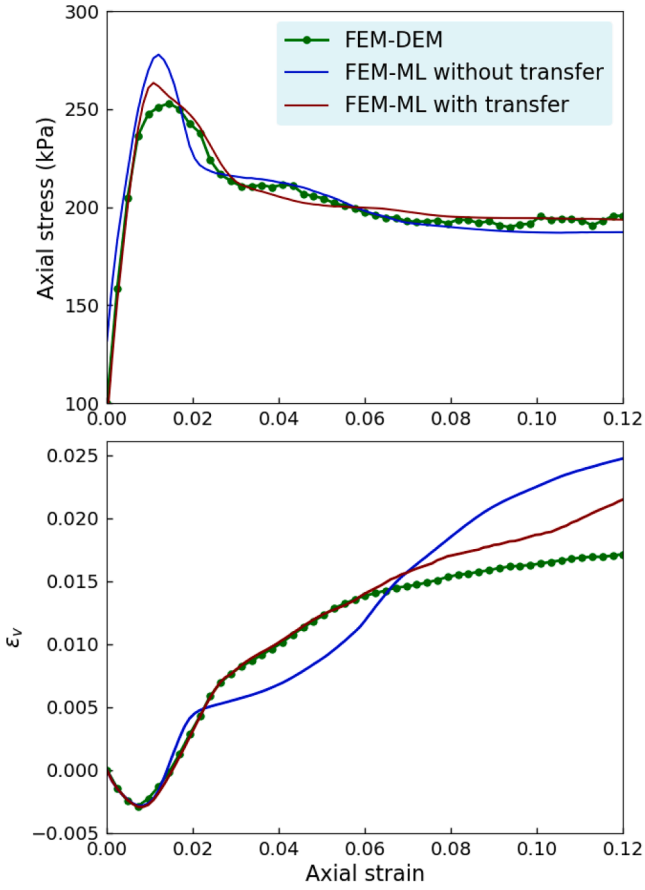


Fig. 19. The evolution of axial stress and volumetric strain against axial strain.

make extrapolated predictions to a certain extent when being integrated into FEM computations. The reason can be attributed to the fact that the deep learning model attempts to make a continuous fit based on available data, resulting in differences observed between the trained surrogate model and the raw data. During FEM computations, different stress responses at Gauss points can cause various nodal displacements and strain increments, thereby giving rise to different strain paths at the same Gauss point.

6. Discussion

In the community of machine learning, transfer learning is known to bring two advantages: one is improving the prediction accuracy and the other is reducing the training time. This study mainly covers the former merit, as we consider “full model fine-tuning” method, instead of the “freezing method” as the dominant transfer strategy. The reasons behind this are two-fold: (1) the time-

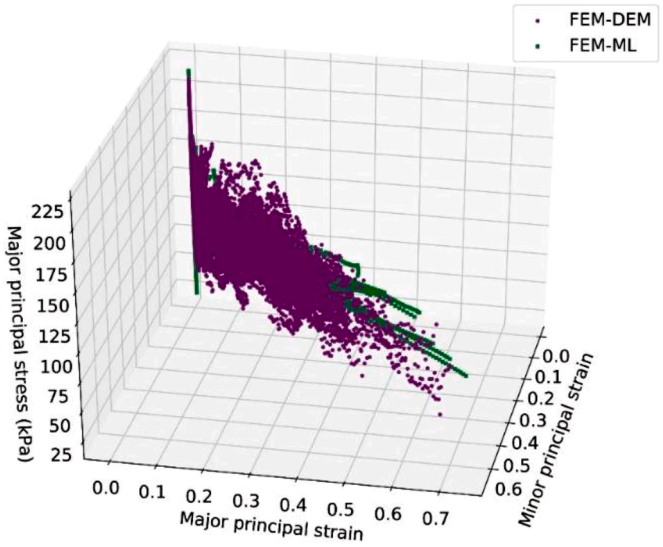


Fig. 20. Major principal stress versus principal strain experienced by Gauss points during biaxial testing.

sequence model does not have an explicit representation for CNN-type neural networks where the first few layers control some general features while the last layers are about the specific features of the fitting; (2) the material modelling problem entails a relatively small dataset and training cost, and thus the bottleneck of data-driven constitutive modelling is about the accuracy, rather than training costs. The advantages and limitations of transfer learning for data-driven constitutive modelling are further elaborated below.

6.1. Knowledge transfer across various paradigms

Within traditional constitutive modelling frameworks, material knowledge cannot be readily transferred across different independent paradigms. In the background of data-driven modelling, the most prevalent approach for utilizing existing knowledge is to enforce physics-based governing equations into the loss function when training neural networks. However, for constitutive modelling, phenomenological models are underpinned by certain assumptions and no governing equations can be theoretically derived for granular media.

Instead, transfer learning offers a more feasible pathway to bridge existing constitutive knowledge in the form of data. As demonstrated in Section 3, we can extract the data representation from well-established analytical models by performing numerical integration to generate substantial data, which will be used to train base deep neural networks. Since these neural networks have learned fundamental features of material behaviour through synthetic stress-strain pairs, less demand for high-fidelity data is required to train deep neural networks, enabling them to adapt to new materials effectively.

The same concept applies to various scenarios across different research paradigms. Particularly for experimental data, the stress-strain responses that can be easily measured are often limited to a few selected strain or stress paths, for example, triaxial loading paths. Although a large amount of experimental data can be prepared, the data produced are still in a small regime of loading paths, in contrast to all the potential stress/strain states and paths which a material may experience in a typical BVP simulation. In these situations, the cross-paradigm transfer learning strategy may become crucial. For example, using numerical simulations can mitigate the difficulty of failing to experimentally generate complex stress-strain paths.

6.2. Knowledge sharing across various materials

In addition to knowledge transfer across analytical models, numerical simulations, and experimental observations, transfer learning can achieve knowledge sharing between similar materials, as investigated in Sections 4 and 5. By learning existing data representations for certain materials, a well-trained deep-learning model can be applied to the development of constitutive models for new materials. On the one hand, the model can achieve continuous improvement if new data about the material are available for re-training the model. On the other hand, the preserved model can be repurposed for new materials, instead of building a model from scratch. The study highlights the importance of preserving all trained data-driven models for future use. For material modelling, it is possible to develop a general or universal data-driven material model by making full use of existing data in the community. Future development can start by restoring this general model for other customised uses with transfer learning, rendering the new training process less data-intensive.

6.3. The scenario that transfer learning may fail

It is noteworthy that transfer learning is not always effective. The premise for successful transfer learning is that the base model can bring additional information which is useful for the new material model but inaccessible through the available data about the new material. Therefore, transfer learning may fail in two cases: (1) The data on the new materials already include a sufficient variety of strain paths, and the data used to train the base model does not contain other new information which is beneficial for training new tasks; (2) The dataset used for pre-training does not share similarities with the training task to be transferred. Currently, the similarity evaluation amongst different data purely comes from domain knowledge. It remains a future attempt to automatically quantify the similarity using data science methods and then indicate whether transfer learning may contribute to training.

7. Concluding remarks

This study presents two transfer learning strategies for data-driven multiscale modelling of granular materials, aiming to achieve high forecast accuracy for deep learning-based material models with limited data availability. The value of transfer learning is first examined in particle-scale DEM simulations of triaxial testing, and is then extended to FEM/DEM multiscale modelling scenarios of biaxial shear tests. Two types of transfer learning models, together with an active learning scheme, are compared and discussed. The deep learning-based material model is embedded in FEM simulations for data-driven multiscale modelling. Two distinct applications of transfer learning for constitutive modelling are demonstrated and the results confirm that transfer learning can enable machine learning in material model discovery even with limited data. Some concluding remarks can be drawn below:

(1) Data-driven model is not a substitute for phenomenological models or multiscale modelling. Instead, by making full use of these well-established models to offer data representations associated with material behaviour, transfer learning can help to fusion these existing research paradigms effectively. Specifically, the stress-strain data from numerical integration based on analytical models can be used to train a base model, which can be transferred to train new tasks according to numerical or experimental data, thereby improving the forecast accuracy of trained models.
(2) The primary advantage of transfer learning is that it facilitates building relatively reliable data-driven constitutive models on small material data conditions where large datasets are unavailable. For the scenario of data-driven material modelling, the full model fine-tuning approach is recommended as a basic transfer learning strategy. Otherwise, transfer learning combined with active learning is the most preferred choice, provided that strain or stress paths are known or can be artificially constructed.
(3) Transfer learning can enable effective knowledge sharing amongst similar materials. A pre-trained model for a certain type of material can be adapted to train models for a new type of material with limited data via transfer learning, as long as both materials share similar strength and deformation mechanisms.
(4) When embedding DNNs into FEM simulations of a BVP as material models, it is found that the strain paths experienced by each Gauss point in DNN-based FEM are different from the training data in FEM-DEM multiscale modelling even when the same BVP is simulated. The finding highlights the importance of developing a data-driven material model with a certain extrapolation ability that can be generalised to new scenarios beyond the training data.
(5) Although this study focused on the application of transfer learning in hierarchical multiscale modelling of granular materials where the DEM model is used on the RVE level, the same strategy can be applied to other multiscale models for materials experiencing plastic and material nonlinearities.

CRediT authorship contribution statement

Tongming Qu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jidong Zhao:** Funding acquisition, Formal analysis, Project administration, Resources, Supervision, Writing – review & editing. **Shaoheng Guan:** Investigation, Methodology, Software, Writing – review & editing. **Y.T. Feng:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jidong Zhao reports financial support was provided by National Natural Science Foundation of China. Jidong Zhao reports financial support was provided by the Research Grants Council of Hong Kong.

Data availability

Data will be made available on request.

Acknowledgment

The study was financially supported by the National Natural Science Foundation of China (via General Project #11972030) and the Research Grants Council of Hong Kong (under GRF #16208720).

Appendix A. deviatoric hardening model

The DH model, like other elastic-plastic models, necessitates the specification of a yield surface, flow rule, and hardening laws to establish an incremental stress-strain analytical relationship. However, the DH model diverges from other models by using the deviatoric plastic strain ε_{q} as the hardening parameter, which is incorporated into hardening equations to depict the evolution of the yield surface. In the *p-q* stress space, the yield function *f* is defined as follows:

$$f = q - \eta p = 0 \tag{A-1}$$

where $p = \frac{1}{3}(\sigma_x + \sigma_y + \sigma_z)$, $q = \sigma_y - \sigma_x$ are the mean and deviatoric stresses, respectively; η is the hardening parameter and its evolution is described by a hyperbolic hardening law:

$$\eta = \eta_f \frac{\varepsilon_q^p}{A + \varepsilon_q^p} \tag{A-2}$$

where A is a material constant associated with the hardening law; ε_q^p is the deviatoric plastic strain, and η_f is the slope of the failure line.

The plastic potential function Φ is

$$\Phi = q + \eta_c p \ln\left(\frac{p}{p_0}\right) = 0 \tag{A-3}$$

where p_0 is the initial mean stress; and η_c is the slope of the zero-dilatancy line.

The constitutive relation in the $p - q$ stress space can be described by

$$dp = \left(K - \frac{\frac{\partial f}{\partial p} K^2 \frac{\partial \Phi}{\partial p}}{H_e + H_p}\right) d\varepsilon_v - \frac{\frac{\partial f}{\partial q} 3GK \frac{\partial \Phi}{\partial p}}{H_e + H_p} d\varepsilon_q \tag{A-4}$$

$$dq = \left(-\frac{\frac{\partial f}{\partial p} 3GK \frac{\partial \Phi}{\partial q}}{H_e + H_p}\right) d\varepsilon_v + \left(3G - \frac{\frac{\partial f}{\partial q} 9G^2 \frac{\partial \Phi}{\partial q}}{H_e + H_p}\right) d\varepsilon_q \tag{A-5}$$

where K and G are, respectively, the small-strain bulk and shear moduli; $d\varepsilon_v$ and $d\varepsilon_q$ are the increments of volumetric strain and deviatoric strain, respectively; H_e and H_p are the elastic and plastic hardening moduli, respectively.

$$H_e = \frac{\partial f}{\partial p} K \frac{\partial \Phi}{\partial p} + \frac{\partial f}{\partial q} 3G \frac{\partial \Phi}{\partial q} \tag{A-6}$$

$$H_p = -\frac{\partial f}{\partial \varepsilon_q^p} \frac{\partial \Phi}{\partial q} \tag{A-7}$$

In the framework of a non-associated flow rule, the increments of the deviatoric plastic strain ε_q^p and volumetric strain ε_v^p can be described by:

$$d\varepsilon_q^p = \mathrm{d}\lambda \frac{\partial \Phi}{\partial q},\ d\varepsilon_v^p = \mathrm{d}\lambda \frac{\partial \Phi}{\partial p},\ \Phi(p, q) = \mathrm{Const} \tag{A-8}$$

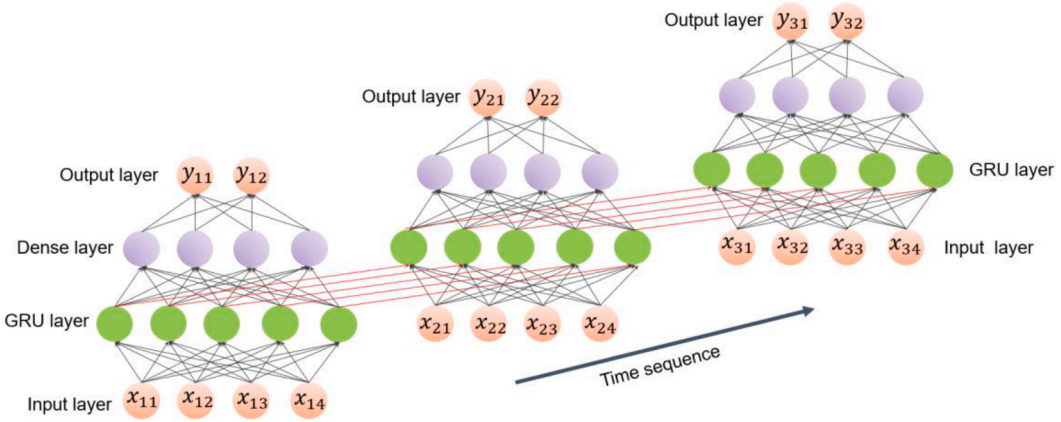
where $\mathrm{d}\lambda$ is a plastic multiplier, and its value can be computed as follows:

$$\mathrm{d}\lambda = \frac{1}{H_P}\left(\frac{\partial \mathrm{f}}{\partial \mathrm{p}} \mathrm{d}p + \frac{\partial \mathrm{f}}{\partial \mathrm{q}} \mathrm{d}q\right) \tag{A-9}$$

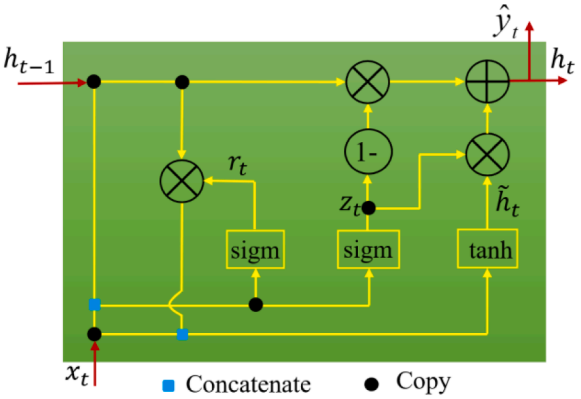
Appendix B. a standard GRU model

A schematic representation of the standard GRU model is shown in Fig. B.1, where (a) demonstrates the basic GRU network architecture incorporating GRU and dense layers, and (b) shows the mathematical operation that happened in each GRU cell. Note the

input x_{tn} and output y_{tm} represent the n^{th} input and the m^{th} output variable at the t^{th} time step, respectively.



(a) A standard GRU network structure



(b) The information flow within a standard GRU cell

Fig. B.1. A schematic of the GRU model.

References

Abueidda, D.W., Koric, S., Sobh, N.A., Sehitoglu, H., 2021. Deep learning for plasticity and thermo-viscoplasticity. Int. J. Plast. 136, 102852.

Bahmani, B., Sun, W., 2021. Training multi-objective/multi-task collocation physics-informed neural network with student/teachers transfer learnings. arXiv preprint.

Bonatti, C., Berisha, B., Mohr, D., 2022. From CP-FFT to CP-RNN: recurrent Neural Network Surrogate Model Of Crystal Plasticity. Int. J. Plast., 103430

Bonatti, C., Mohr, D., 2021. One for all: universal material model based on minimal state-space neural networks. Sci. Adv. 7, eabf3658.

Bonatti, C., Mohr, D., 2022. On the importance of self-consistency in recurrent neural network models representing elasto-plastic solids. J. Mech. Phys. Solids 158, 104697.

Ellis, G., Yao, C., Zhao, R., Penumadu, D., 1995. Stress-strain modeling of sands using artificial neural networks. J. Geotech. Eng. 121, 429–435.

Fazily, P., Yoon, J.W., 2023. Machine learning-driven stress integration method for anisotropic plasticity in sheet metal forming. Int. J. Plast. 166, 103642.

Geers, M.G., Kouznetsova, V.G., Brekelmans, W., 2010. Multi-scale computational homogenization: trends and challenges. J. Comput. Appl. Math. 234, 2175–2182.

Ghaboussi, J., Garrett Jr, J., Wu, X., 1991. Knowledge-based modeling of material behavior with neural networks. J. Eng. Mech. 117, 132–153.

Ghaboussi, J., Sidarta, D., 1998. New nested adaptive neural networks (NANN) for constitutive modeling. Comput. Geotech. 22, 29–52.

Gorji, M.B., Mozaffar, M., Heidenreich, J.N., Cao, J., Mohr, D., 2020. On the potential of recurrent neural networks for modeling path dependent plasticity. J. Mech. Phys. Solids 143, 103972.

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

Guan, S., Feng, Y., Ma, G., Qu, T., Wang, M., Zhou, W., 2023b. An explicit FEM-NN framework and the analysis of error caused by NN-predicted stress. Acta Geotech. 1–20.

Guan, S., Qu, T., Feng, Y., Ma, G., Zhou, W., 2023c. A machine learning-based multi-scale computational framework for granular materials. Acta Geotech. 18, 1699–1720.

Guo, N., Zhao, J., 2014. A coupled FEM/DEM approach for hierarchical multiscale modelling of granular media. Int. J. Numer Methods Eng. 99, 789–818.

Guo, N., Zhao, J., 2016. Multiscale insights into classical geomechanics problems. Int. J. Numer. Anal. Methods Geomech. 40, 367–390.

Guo, R., Li, G., 2008. Elasto-plastic constitutive model for geotechnical materials with strain-softening behaviour. Comput. Geosci. 34, 14–23.

Heider, Y., Wang, K., Sun, W., 2020. SO (3)-invariance of informed-graph-based deep neural network for anisotropic elastoplastic materials. Comput. Methods Appl. Mech. Eng. 363, 112875.

Huang, D., Fuhg, J.N., Weißenfels, C., Wriggers, P., 2020. A machine learning based plasticity model using proper orthogonal decomposition. Comput. Methods Appl. Mech. Eng. 365, 113008.

Ibragimova, O., Brahme, A., Muhammad, W., Connolly, D., Lévesque, J., Inal, K., 2022. A convolutional neural network based crystal plasticity finite element framework to predict localised deformation in metals. Int. J. Plast. 157, 103374.
Ibragimova, O., Brahme, A., Muhammad, W., Lévesque, J., Inal, K., 2021. A new ANN based crystal plasticity model for FCC materials and its application to non-monotonic strain paths. Int. J. Plast. 144, 103059.
Jang, D.P., Fazily, P., Yoon, J.W., 2021. Machine learning-based constitutive model for J2-plasticity. Int. J. Plast. 138, 102919.
Jordan, B., Gorji, M.B., Mohr, D., 2020. Neural network model describing the temperature-and rate-dependent stress-strain response of polypropylene. Int. J. Plast. 135, 102811.
Jung, S., Ghaboussi, J., 2006. Neural network constitutive model for rate-dependent materials. Comput. Struct. 84, 955–963.
Karapiperis, K., Stainier, L., Ortiz, M., Andrade, J., 2021. Data-driven multiscale modeling in mechanics. J. Mech. Phys. Solids 147, 104239.
Kawamoto, R., Andò, E., Viggiani, G., Andrade, J.E., 2018. All you need is shape: predicting shear banding in sand with LS-DEM. J. Mech. Phys. Solids 111, 375–392.
Kuhn, M.R., Daouadji, A., 2018. Multi-directional behavior of granular materials and its relation to incremental elasto-plasticity. Int. J. Solids Struct. 152, 305–323.
Li, X., Roth, C.C., Mohr, D., 2019. Machine-learning based temperature-and rate-dependent plasticity model: application to analysis of fracture experiments on DP steel. Int. J. Plast. 118, 320–344.
Liang, W., Zhao, J., 2019. Multiscale modeling of large deformation in geomechanics. Int. J. Numer. Anal. Methods Geomech. 43, 1080–1114.
Liu, X., Athanasiou, C.E., Padture, N.P., Sheldon, B.W., Gao, H., 2021. Knowledge extraction and transfer in data-driven fracture mechanics. Proc. Natl. Acad. Sci. 118, e2104765118.
Ma, G., Guan, S., Wang, Q., Feng, Y., Zhou, W., 2022. A predictive deep learning framework for path-dependent mechanical behavior of granular materials. Acta Geotech. 17, 3463–3478.
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., Stefanou, I., Vannucci, P., Maffi-Berthier, V., 2021. Thermodynamics-based Artificial Neural Networks for constitutive modeling. J. Mech. Phys. Solids 147, 104277.
Mozaffar, M., Bostanabad, R., Chen, W., Ehmann, K., Cao, J., Bessa, M., 2019. Deep learning predicts path-dependent plasticity. Proc. Natl. Acad. Sci. 116, 26414–26420.
Nascimento, A., Roongta, S., Diehl, M., Beyerlein, I.J., 2023. A machine learning model to predict yield surfaces from crystal plasticity simulations. Int. J. Plast. 161, 103507.
Perera, R., Agrawal, V., 2022. A generalized machine learning framework for brittle crack problems using transfer learning and graph neural networks. Mech. Mater. 181 (2023), 104639.
Pouragha, M., Wan, R., 2017. Non-dissipative structural evolutions in granular materials within the small strain range. Int. J. Solids Struct. 110, 94–105.
Qu, T., Di, S., Feng, Y., Wang, M., Zhao, T., 2021a. Towards data-driven constitutive modelling for granular materials via micromechanics-informed deep learning. Int. J. Plast. 144, 103046.
Qu, T., Di, S., T Feng, Y., Wang, M., Zhao, T., Wang, M., 2021b. Deep learning predicts stress–strain relations of granular materials based on triaxial testing data. Comput. Model. Eng. Sci. 128, 129–144.
Qu, T., Feng, Y., Wang, Y., Wang, M., 2019. Discrete element modelling of flexible membrane boundaries for triaxial tests. Comput. Geotech. 115, 103154.
Qu, T., Guan, S., Feng, Y., Ma, G., Zhou, W., Zhao, J., 2023. Deep active learning for constitutive modelling of granular materials: from representative volume elements to implicit finite element modelling. Int. J. Plast. 164, 103576.
Su, M., Guo, N., Yang, Z., 2023. A multifidelity neural network (MFNN) for constitutive modeling of complex soil behaviors. Int. J. Numer. Anal. Methods Geomech. https://doi.org/10.1002/nag.3620.
Tancogne-Dejean, T., Gorji, M.B., Zhu, J., Mohr, D., 2021. Recurrent neural network modeling of the large deformation of lithium-ion battery cells. Int. J. Plast. 146, 103072.
Vlassis, N.N., Sun, W., 2021. Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. Comput. Methods Appl. Mech. Eng. 377, 113695.
Wang, K., Johnson, C.W., Bennett, K.C., Johnson, P.A., 2021. Predicting fault slip via transfer learning. Nat. Commun. 12, 7319.
Wang, K., Sun, W., 2019. Meta-modeling game for deriving theory-consistent, microstructure-based traction–separation laws via deep reinforcement learning. Comput. Methods Appl. Mech. Eng. 346, 216–241.
Wang, K., Sun, W., Du, Q., 2019. A cooperative game for automated learning of elasto-plasticity knowledge graphs and models with AI-guided experimentation. Comput. Mech. 64, 467–499.
Wang, M., Qu, T., Guan, S., Zhao, T., Liu, B., Feng, Y., 2022. Data-driven strain–stress modelling of granular materials via temporal convolution neural network. Comput. Geotech. 152, 105049.
Wen, J., Zou, Q., Wei, Y., 2021. Physics-driven machine learning model on temperature and time-dependent deformation in lithium metal and its finite element implementation. J. Mech. Phys. Solids 153, 104481.
Wu, M., Wang, J., 2022. Constitutive modelling of natural sands using a deep learning approach accounting for particle shape effects. Powder Technol. 404, 117439.
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., Sheil, B., 2022. Physics-constrained hierarchical data-driven modelling framework for complex path-dependent behaviour of soils. Int. J. Numer. Anal. Methods Geomech. 46, 1831–1850.
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, 1315–1335.