International Journal of Plasticity 164 (2023) 103576

Contents lists available at ScienceDirect

International Journal of Plasticity

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

Deep active learning for constitutive modelling of granular materials: From representative volume elements to implicit finite element modelling

Tongming Qu [a,b], Shaoheng Guan [b,c,*], Y.T. Feng [b,*], Gang Ma [c,d], Wei Zhou [c,d], Jidong Zhao [a]

[a] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong Special Administrative Region, China
[b] Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Swansea, Wales SA1 8EP, UK
[c] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, Hong Kong Special Administrative Region, China
[d] Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education, Wuhan University, Wuhan 430072, Hong Kong Special Administrative Region, China

ARTICLE INFO

Keywords:
Granular materials
Machine learning
Constitutive models
Active learning
Discrete element model
Finite element model

ABSTRACT

Constitutive relation remains one of the most important, yet fundamental challenges in the study of granular materials. Instead of using closed-form phenomenological models or numerical multiscale modelling, machine learning has emerged as an alternative paradigm to revolutionise the constitutive modelling of granular materials. However, deep neural networks (DNNs) require massive training data and often fail to make credible extrapolations. This study aims to develop a deep active learning strategy to (i) identify unreliable forecasts without knowing the ground truth; and (ii) continuously improve and verify a data-driven constitutive model until the desired generalisation is satisfied. The role of active learning in constitutive modelling is instantiated through three scenarios: (i) off-line strain-stress data pool of granular materials; (ii) interactive constitutive training and strain-stress data labelling; and (iii) finite element modelling (FEM) driven by deep learning-based constitutive models. The results confirm the capability of active learning in advancing data-driven constitutive modelling of granular materials toward developing a faithful surrogate constitutive model with less data. The same active learning strategy can also be applied to other data-centric applications across various science and engineering fields.

1. Introduction

The description of the deformation and instability of materials under external loads has been an open scientific challenge for human beings since the origin of modern science. As a mathematical approximation of material behaviour the constitutive relation serves as the cornerstone not only for understanding the mechanical performance of materials but also for performing macroscale numerical computations (e.g. by FEM).

* Corresponding authors at: Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Swansea, Wales SA1 8EP, UK.
E-mail addresses: guanshaoheng@whu.edu.cn (S. Guan), y.feng@swansea.ac.uk (Y.T. Feng).

https://doi.org/10.1016/j.ijplas.2023.103576
Received 14 September 2022; Received in revised form 29 January 2023;
Available online 28 February 2023

For granular media the strength and stiffness are strongly rooted in grain-scale interactions: loads are supported by and transferred via interparticle contacts. Any contact-scale sliding will cause irreversible plastic deformation, giving rise to evolving strength and stiffness characteristics (Qu et al., 2022; Zhao et al., 2016). These seemingly simple particle motions can trigger ample and complicated mechanical behaviour of granular materials, such as state and path dependence, anisotropy and strain localisation (Gong et al., 2019; Peng et al., 2020; Peng et al., 2022; Qu et al., 2019; Zhao and Guo, 2015).

The most prevalent approach to formulating constitutive relations of granular materials until now has been the phenomenological constitutive theory of plasticity (Gao and Zhao, 2017; Shi et al., 2019; Zhang et al., 2021). Its success is underpinned by four primary assumptions: (1) an explicit partition of elastic and plastic zones in a stress-strain sequence; (2) the yield surface distinguishing the boundary between elasticity and plasticity; (3) the associative or non-associative flow rule describing the direction of plastic deformation; and (4) the hardening rule characterising the evolution of the yield surface. This classical plasticity theory has received extensive applications, but it also confronts dilemmas owing to the *prior* assumptions and increasing model and state parameters to be calibrated. In addition, some researchers found that many basic assumptions that are taken for granted in phenomenological models are not in agreement with meticulous numerical and experimental observations (Kawamoto et al., 2018; Kuhn and Daouadji, 2018a, 2018b; Pouragha and Wan, 2017).

The other paradigm for constitutive modelling is hierarchical multiscale modelling (HMM) (Guo and Zhao, 2014). In this framework, continuum-based numerical models (e.g. FEM) can be computed without resorting to phenomenological assumptions. The constitutive relation required at each Gauss point is approximated with particle-scale discrete element modelling (DEM) of a representative volume element (RVE) (Guo and Zhao, 2014, 2016a; Qu et al., 2021c). A boundary value problem (BVP) simulated with FEM normally incorporates many Gauss points, thus massive DEM models need to be concurrently solved, resulting in extremely high computational costs.

Machine learning offers a promising alternative to tackle the challenge of constitutive modelling (Abueidda et al., 2021; Jang et al., 2021; Li et al., 2022; Wu et al., 2022; Xi et al., 2021; Zhang and Mohr, 2020; Zhang et al., 2021). Unlike the phenomenological constitutive theory and multiscale modelling, data-driven models do not require parameter calibration and phenomenological assumptions, neither do they request unaffordable computational resources to infer stress responses from strain paths. Although it is not new to apply neural networks to model the stress-strain relations of concrete and sands (Ellis et al., 1995; Ghaboussi et al., 1991; Ghaboussi and Sidarta, 1998), the revolutionary development of deep learning over recent years re-inspires extensive explorations in data-driven constitutive models (Guan et al., 2023; Ibragimova et al., 2022; Ibragimova et al., 2021; Jordan et al., 2020; Tancogne-Dejean et al., 2021). For example, Wang and Sun (2019) and Wang et al. (2019) developed reinforcement learning and game theory-based deep learning models for the constitutive modelling of granular materials. Qu et al. (2021a) developed a micromechanics-informed stress-strain modelling strategy by integrating evolving elastic stiffness matrix and physics-invariant material properties for constitutive training. Ma et al. (2022) proposed an improved long-short term memory (LSTM) model considering the initial state of granular materials. Wu and Wang (2022) considered the effect of particle shape on LSTM-based constitutive modelling. Bonatti and Mohr (2021, 2022) proposed a minimal state cell (MSC) to reduce the number of state variables (memory units) and developed linearised MSC (or LMSC) neural networks to improve incremental size dependence in traditional recurrent neural networks (RNNs). The capability of LMSC is further instantiated by the crystal plasticity of an aluminium alloy and the learned model demonstrates satisfactory performance when embedded in FEM (Bonatti et al., 2022).

The importance of involving domain-knowledge into data-driven constitutive modelling has been widely recognised. Heider et al. (2020) and Qu et al. (2021b) discussed frame-indifference properties in data-driven constitutive modelling. Vlassis and Sun (2021) and Zhang et al. (2022) embedded the first and second-order gradients of the loss function into training with Sobolev training. Masi et al. (2021) proposed a thermodynamics-based deep learning strategy by taking the energy potential and a dissipation pseudo-potential as the direct outputs of neural networks. The final stress responses are obtained by differentiating the outputs of neural networks with respect to their inputs based on the automatic differentiation technique. Masi and Stefanou (2022) further extend their TANN framework to materials with inelastic and complex microstructures by using encoder-decoder to automatically search for the internal state variables of complex inelastic materials. Wen et al. (2021) developed a physics-driven machine learning model for constitutive modelling incorporating temperature and time-dependent effects in lithium metal. Following the classical elastoplastic framework, a tree-based machine learning model is adopted to bypass the use of the traditional power-law-based hardening rule. As most of the plasticity theory remains, only limited experimental results are required to train the data-driven model.

Deep learning-based constitutive modelling includes two indispensable components: *model* and *data*. In contrast to the above-reviewed progress from the perspective of *model*, rather few efforts have been taken to guarantee that the dataset used to train neural networks is sufficiently representative. Yet data is crucial to developing a successful machine learning model and a high-quality dataset should cover all possible scenarios that the surrogate model is intended for. Data-driven constitutive modelling tends to suffer from two challenges pertaining to data: (1) the training process is data-demanding while it is costly to obtain high-fidelity data, via either laboratory experiments or microscale numerical simulations; (2) as data-driven models are not underpinned by physical principles, the reliability of a data-driven model is always a concern.

Active learning can provide a useful perspective to tackle the above challenges. Any form of learning model in which the algorithm can partially control input specimens can be regarded as active learning (Lughofer, 2012). For conventional supervised learning with a "passive learning" framework, where the training dataset is generated before the training process is performed, the model has to be prepared blindly and inefficiently. The idea behind active learning is that not all data is created equally, because different data may carry a different information intensity for the model. Provided that the most instructive data is identified, labelled, and employed to fit a model, it can be expected that such data may enable a trained model to reach the desired generalisability. However, incorporating active learning in the constitutive modelling of materials remains a relatively unexplored territory to date.

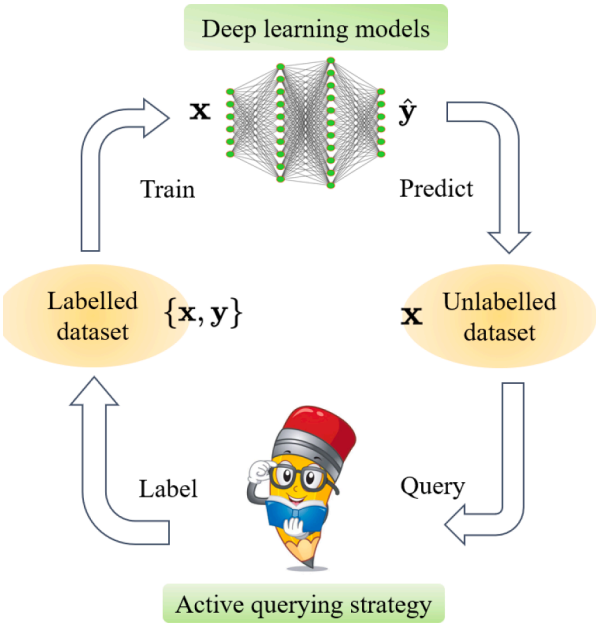


Fig. 1. Procedures of deep active learning.
(The cartoon is accessed from https://kidsread.wordpress.com/2022/06/03/3-steps-for-teaching-kids-to-read/).

This study aims to fill the above gap by developing a deep learning committee-based active constitutive learning strategy. Through three different scenarios, the unique advantages of deep active learning are revealed. On the one hand, active learning can prioritise selecting the most informative data when training a surrogate model. On the other hand, active learning can serve as an effective tool to detect potential predictive blind points and improve the model continuously. We demonstrate that active learning works for both multilayer perceptron (MLP) and RNNs, but its application can go beyond any type of data-driven surrogate model. We also confirm that deep active learning is suitable not only for the data from common conventional or true triaxial tests but also for complex strain-stress paths experienced in a BVP problem. Furthermore, the trained DNNs are embedded into FEM computations to bypass phenomenological constitutive models or particle-scale DEM simulations.

The remainder of the paper is structured as follows. Section 2 introduces the basic idea of active learning in general and the committee-based active learning strategy in particular. In Section 3, a comprehensive examination is conducted to understand different active learning strategies in data-driven constitutive modelling. An interactive constitutive training scheme is demonstrated in Section 4, where data generation and the training of DNNs are integrated seamlessly. Such a procedure allows models to identify the most informative data to make DNNs learn faster in a cost-effective manner. Section 5 performs a data-driven hierarchical multiscale modelling case to exemplify the capability of active learning in discovering unreliable forecasts in macroscopic modelling. The significance and limitations of active learning in constitutive modelling are discussed in Section 6. Some concluding remarks are made in Section 7.

2. A deep active learning strategy for constitutive modelling

Active learning is originally motivated by many machine learning scenarios where input-output data pairs are difficult or expensive to collect, whereas most supervised learning models are data-greedy. An important issue arises: can we use as small datasets as possible to train a reliable predictive model? By identifying the most instructive data for a surrogate model, active learning can maximise the performance of models while minimising the amount of data to be labelled. Note that active learning is a strategy, rather than a specific model. When combining deep learning models with the active learning strategy, the model is called deep active learning.

The basic procedure of deep active learning is shown in Fig. 1. For a mapping relation connecting inputs $\mathbf{X}$ and outputs $\mathbf{Y}$, we have a labelled data pool $\boldsymbol{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ and an unlabelled data pool $\boldsymbol{D}_u = \{x_i\}_{i=1}^{N_u}$ before training. The first step is to build deep neural networks $F^{(0)}$ with prescribed $\mathbf{X}$ and $\mathbf{Y}$. These models can be either trained or untrained with labelled datasets $\boldsymbol{D}_l$. Second, these deep neural networks are utilised to infer outputs $\widehat{\mathbf{Y}}$ based on the unlabelled data pool $\boldsymbol{D}_u$. Third, an active querying strategy is adopted to identify some unlabelled data which is most likely to be mis-predicted in $\boldsymbol{D}_u$. Then the fourth step is to label/resample these recognised datasets via performing laboratory experiments or numerical simulations. These data are moved from $\boldsymbol{D}_u$ to $\boldsymbol{D}_l$ and new deep neural networks $F^{(1)}$ are fitted based on the updated labelled data $\boldsymbol{D}_l$. The workflow is an iterative procedure including "exploration—labelling—training". The procedure is repeated until the trained DNNs reach the desired accuracy over a targeted stress-strain space. Over these active learning rounds $t = \{0, 1, 2, \ldots, T\}$, a series of DNNs $F^{(0)}, F^{(1)}, F^{(2)}, \ldots F^{(T)}$ are trained until the mapping relation from $\mathbf{X}$ to $\mathbf{Y}$ has been well captured by the latest DNN.

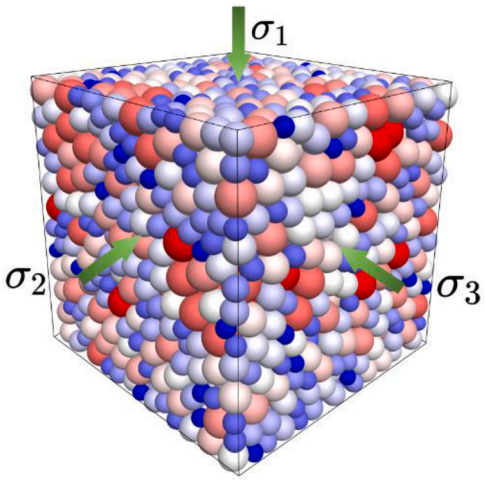


Fig. 2. DEM models for generating stress-strain data.

The core of active learning lies in the querying strategy. Many active query strategies are reported in the literature, such as uncertainty sampling (Lewis and Catlett, 1994), expected error reduction (Settles, 2009), density-weighted methods (Settles, 2009) and committee-based query (CBQ) strategy (Seung et al., 1992). However, most active learning strategies are domain-dependent and cannot be adapted to other fields. Nevertheless, the CBQ approach can be a universal option which also works for sequence-based regression problems, such as the constitutive modelling of granular materials. In addition, the CBQ method is conceptually simple and easy to be implemented, and therefore this strategy is adopted in this study.

The CBQ strategy requires a committee $\boldsymbol{C} = \{\boldsymbol{M}^{(1)}, \ldots, \boldsymbol{M}^{(N)}\}$ of N surrogate models ($\boldsymbol{M}$) which are fitted based on an available labelled dataset $\boldsymbol{D}_l$. Each committee member provides forecasts on the data points extracted from the unlabelled data pool $\boldsymbol{D}_u$. Then the predictive disagreement among different committee members, i.e. surrogate models, can serve as an indicator of the degree of uncertainty. Specifically, the sample with the largest degree of disagreement among all the committee members is selected as the most informative one for the current surrogate model.

This approach is established based on the idea that the model should make a better prediction in the domain that is sufficiently covered by the training dataset, compared to the region where the training data is sparsely distributed. It is expected that the forecast in the region with sufficient data coverage will converge to the ground truth and thus the inferences made by the committee surrogate models are relatively close to each other. By contrast, the predictions around the region where the training data is insufficient will scatter with a measurable variance. The disagreement of forecasts can be quantified by the standard deviation s_N for a given specimen:

$$s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \overline{z})^2} \tag{1}$$

where z_i is the forecast of the i^{th} committee member with $i \in [1,N]$, and $\overline{z}$ is the mean value of these forecasts. By ranking the standard deviations of the predictions from the unlabelled pool in their magnitude, the highest variances in the output forecasts indicate the most instructive data.

One more question is how to choose committee members. This study adopts a group of DNN models, as committee members, with the only difference consisting in the random initialisation of weights and biases before training. The predictive capability of DNNs is empowered by weight and bias parameters. As a high-dimensional mapping of the objective function, different (random) weight initialisations give rise to marginally different DNNs after training, even though the network architecture and all other hyperparameters are the same. Thus, one can leverage the active learning strategy regardless of what types of DNNs are in use.

For the constitutive modelling of path-dependent materials, currently there are two typical DNNs available: One is a multilayer perceptron (MLP) and the other one is a time-series model, such as RNNs and temporal convolutional neural networks (TCN) (Wang et al., 2022). MLP can capture the point-to-point mapping efficiently but must introduce some internal variables to encode the loading history. In contrast, time-sequence models can capture extremely complex stress-strain relations with multiple unloading-reloading cycles but must rely on a large number of parameters and complex network structures to achieve good performance.

In the following sections, we will consider different constitutive training scenarios which may be encountered in path-dependent materials. Also, both MLPs and time-series-based DNNs will be investigated in the subsequent sections. The detailed road map is as follows:

- *The role of active learning*. Section 3 aims to clarify the applicability of active learning in the constitutive modelling of granular materials by performing parametric and comparative examinations of three different scenarios of active learning in a stress-strain data pool. The conclusions drawn will guide the study in Sections 4 and 5.

Table 1
DEM simulation parameters.

Radii (mm)	Density (kg/m^3)	Normal stiffness (N/m)	Tangential stiffness (N/m)	Frictional coefficient	Local damping coefficient
2-4	2,600	1×10^5	5×10^4	0.5	0.5

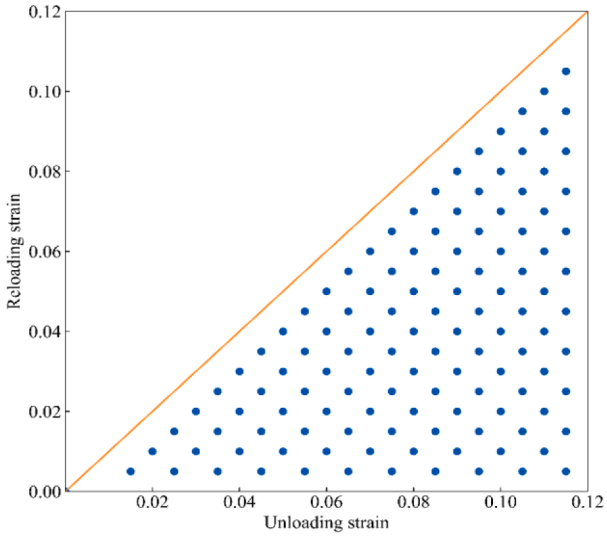


Fig. 3. Sampling points for constant-p true triaxial testing.

- *On-the-fly active learning*. Section 4 demonstrates the on-the-fly active learning where data generation and model training are performed interactively. During each round, only the optimal data which can minimise the prediction uncertainty is labelled. This part focuses on stress-strain predictions of granular materials under conventional triaxial testing scenarios. Special attention is paid to the advantage of active learning in identifying unreliable forecasts of a data-driven surrogate model and adaptively improving the model until it behaves satisfactorily in the desired stress-strain space.
- *Data-driven multiscale model*. Section 5 develops a data-driven multiscale model by harnessing the stress-strain data produced from the FEM×DEM hierarchical multiscale modelling of granular materials. These stress-strain sequence pairs are homogenised from DEM simulations of RVE specimens which serve as material models at Gauss points in FEM modelling of a BVP. Active learning is used to prioritise the selection of training specimens. The trained DNN is also embedded into FEM computations.

3. A comprehensive examination of active learning-assisted data-driven constitutive modelling based on a data pool

3.1. Discrete element models to generate data

Although many simplifications are made in DEM, some research has shown that even the simplest sphere-based modelling can reproduce the primary behaviour (stress-strain relation, volumetric behavior, and critical state) of real granular materials (Qu et al., 2022). Here the stress-strain data of granular materials is generated through offline true triaxial testing via DEM (Fig. 2). A total of 122 groups of triaxial tests are simulated with mutually different unloading-reloading loops. These specimens are composed of 4,037 spherical particles within a cubic volume subjected to an isotropic consolidation pressure of 200 kPa. The simulation properties are listed in Table 1. During triaxial testing, the lateral boundary of the whole specimen is controlled to maintain a constant mean pressure (constant-p) with identical middle and minor principal stresses, namely:

$$\sigma_2 = \sigma_3 = (3p - \sigma_1)/2 \tag{2}$$

where $\sigma_1, \sigma_2, \sigma_3$ and p denote the major, middle, and minor principal stress, and the mean pressure, respectively. Such a constant-p condition is achieved by dynamically controlling the magnitude of lateral pressures, in the manner of moving the position of boundaries adaptively via a servo-boundary algorithm. The influence of this servo-boundary on the macroscopic stress-strain behavior is rather limited according to the previous research (Qu et al., 2019), as the servo-wall boundary is a type of stress boundary, in essence.

As the lateral boundary is governed by the selected loading condition (for example, the constant-p condition here), the only discrepancy for the triaxial specimens consists in the axial loading path. In the current case, all the specimens are loaded by imposing a constant rate of 1×10^{-4}m/s in the vertical direction but with mutually different unloading-reloading loops. Fig. 3 describes the

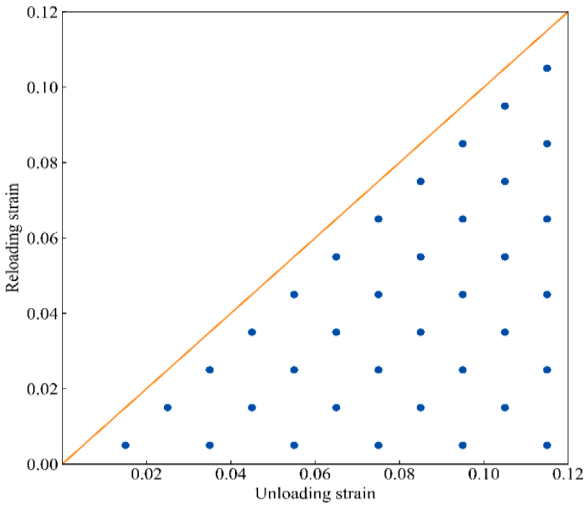


Fig. 4. Specimen distribution used for preliminary training.

distribution of these unloading-reloading points in each strain path, where each point represents a specific loading path, with its horizontal and vertical coordinates representing the unloading and reloading strains, respectively.

The specimens are selected with three restrictions: (1) the unloading strain is greater than the subsequent reloading strain in a single loop; (2) the reloading strain remains in a compressive state (positive value); and (3) all loading processes terminate at a maximum compressive strain of 12%. The first stipulation is a mandatory requirement following some physical rules, while the latter two are used to restrict the constitutive responses within a limited space.

To examine the role of active learning in constitutive modelling, three different training cases are considered:

Case 1: DNN models are initially trained with uniformly but relatively sparse sampled data. A total of 36 groups of specimens, as shown in Fig. 4, are used for the preliminary training. Then active learning is harnessed to detect blind predictions and improve the DNN models continuously by adding the most important points during each round. The workflow of Case 1 can be found in Fig. 5a.
Case 2: DNN models start from nothing, i.e. no data is used to train the initial DNNs. These DNNs are improved incrementally by adding the data points recognised by active learning. This case is designed to investigate whether active learning can select data judiciously at the very beginning of training. The workflow of Case 2 is exhibited in Fig. 5b.
Case 3: DNNs are trained with randomly selected stress-strain samples. This is the conventional training scenario without the involvement of active learning and thus can be regarded as a passive learning example.

3.2. The adopted GRU neural network and accuracy evaluation

In this section, the gated recurrent units (GRU) neural network is adopted as the base DNN model. In our previous work (Qu et al., 2021a), this type of model has been proved to have an extraordinary fitting ability for path-dependent granular materials. Also, the GRU model is not very sensitive to network architectures and hyperparameters for relatively small-scale training data. Diverse architectures and hyperparameters can yield similar predictions owing to the strong fitting capability. In this work, the adopted architectures and hyperparameters are determined by the experience drawn from our previous work. The details can be found in Table 2. Such an architecture yields 5523 parameters to be trained. The GRU networks are built based on Tensorflow (Abadi et al., 2016) and Keras (Chollet, 2015).

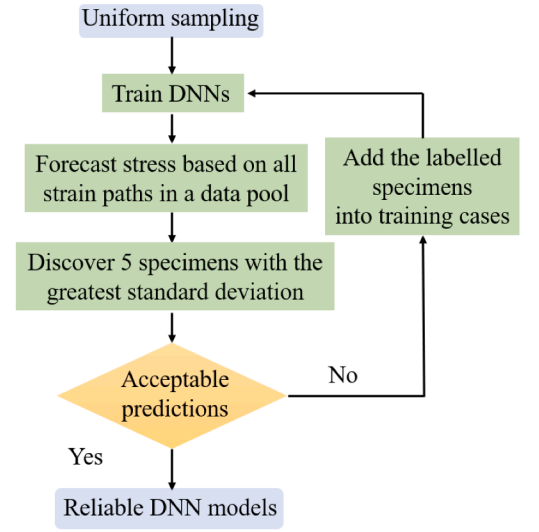
The inference accuracy of a model is evaluated by quantifying the overall difference between the forecasts and the ground truth. Two evaluation metrics are employed in this study. One is the mean absolute error (MAE) (see Eq. (3)) and the other is the score metric (refer to Qu et al. 2021a). The former can offer a relatively rigorous evaluation of the predictive ability while the latter is more intuitive to understand how good these predictions are as a whole.

The MAE of the jth stress-strain path is given by

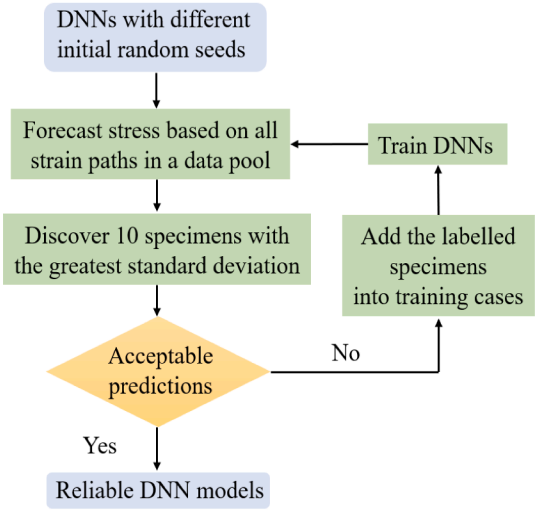
$$MAE_j = \frac{1}{N^j} \sum_{i=1}^{N^j} \left| y_{ij}^{True} - y_{ij}^{Prediction} \right| \tag{3}$$

where y_{ij}^{True} and $y_{ij}^{Prediction}$ are the actual and predicted values of the ith data point of the jth stress-strain curve, respectively.

Fig. 6 shows the loss values of training datasets in Case 1. The results show that the MAE values converge to a steady state after 200 epochs. In each active learning round, new data is added to training datasets. With increasing active learning rounds, the values of the loss function decrease gradually. Note that as the network architectures have been predetermined, no validation is conducted during training.



(a) Case 1



(b) Case 2

Fig. 5. Basic active learning procedures for Case 1 and Case 2.

Table 2
The adopted network architecture and some key hyperparameters.

Item	Value
Network architecture	GRU: 40
The length of moving windows	40
Batch size	64
Epoch number	200
Learning rate	0.001

3.3. Examination and verification of the role of active learning in constitutive modelling

A series of parametric investigations are performed to understand the potential and limitations of deep active learning for data-driven constitutive modelling of granular materials. First, for Cases 1 and 2, four different DNNs are prepared and three committees with two, three, and four DNNs as members are considered. In Case 1, each DNN is pre-trained based on available datasets, while

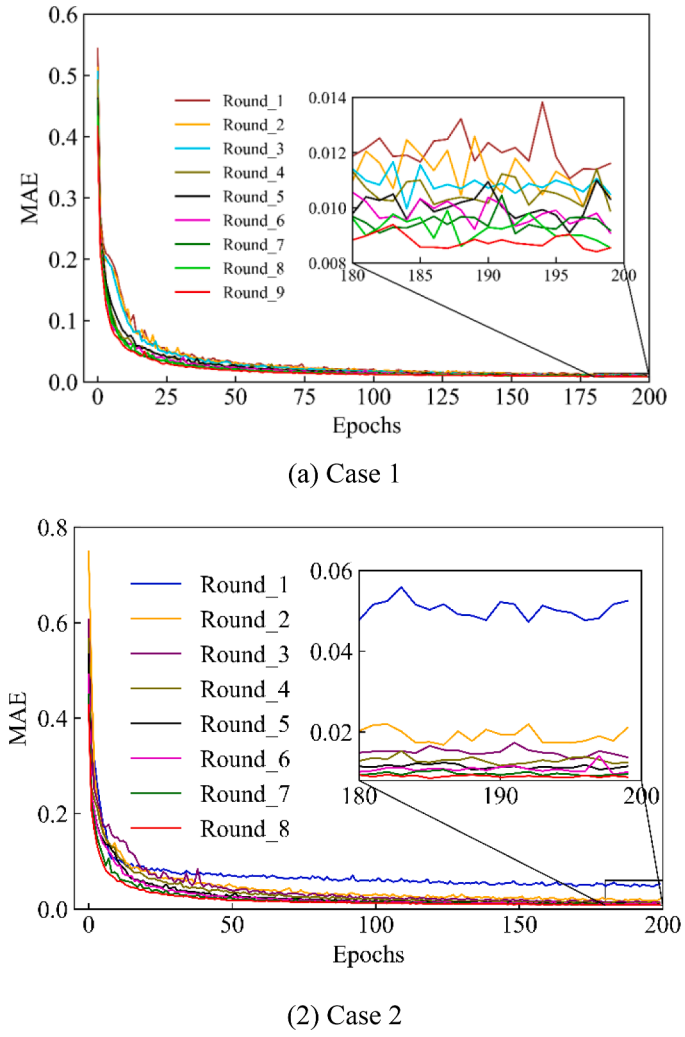


Fig. 6. Learning curves during training.

Table 3
The first-round active learning-assisted forecasts and verifications for Case 1.

Ranking	Estimated MAE ranking with active learning			Actual MAE ranking			
	2 DNNs	3 DNNs	4 DNNs	NN1	NN2	NN3	NN4
1	52	52	52	52	52	1	52
2	1	1	1	54	1	52	60
3	53	53	53	60	53	53	1
4	55	54	74	4	4	60	74
5	6	4	4	74	54	74	53
6	74	74	54	1	8	65	65
7	65	55	60	69	66	55	4
8	60	8	55	55	69	54	54
9	4	60	65	53	15	4	69
10	56	6	6	8	57	63	15

in Case 2, all the DNNs are simply initialised with different random seeds and no training is conducted at the beginning.

Active learning is utilised to detect all the forecast performance over the entire labelled data pool. Each committee member is employed to infer stress responses of all the strain paths in the data pool. Ten groups of data specimens with the largest standard deviations in Case 1 are listed in descending order in Table 3, where each index number in the table represents a certain strain path. To verify the capability of active learning, the actual predictive accuracy for each strain path is quantified by comparing it with the ground truth stress response. Then the ten worst predicted strain paths by each DNN are also listed.

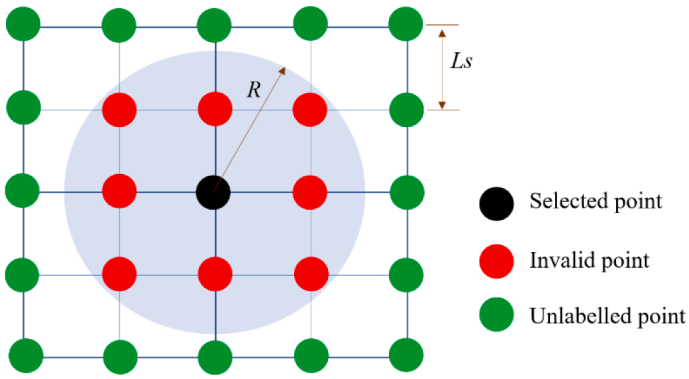


Fig. 7. Illustration of data resampling in a batch-mode active learning scheme.

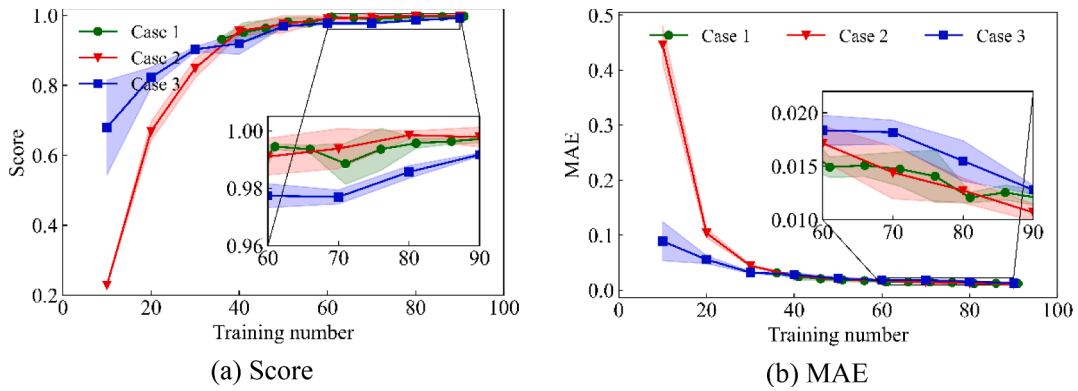


Fig. 8. Prediction performance of principal stress with three training cases.

The table demonstrates that most estimations made by the active learning algorithm are consistent with the actual predictions. The ten worst predictions given by each committee member DNN vary slightly from each other, which confirms that different initialisations of weights and biases do yield different DNN models. Among the ten groups of most unreliable forecasts estimated by active learning with two, three and four committee DNNs, a total of eight, nine and nine groups of strain paths, respectively, are exactly within the actual MAE ranking list.

The results confirm that the active learning strategy is capable of discovering the worst predictions without knowing the ground truth. Furthermore, the results support that only two or three committee members are sufficient to serve as an error indicator, as using more committee member DNNs requires more computational costs. A group of three committee members are thus adopted throughout the study.

The results in Case 1 verify the effectiveness of active learning. However, in the first round of Case 2, the predictions given by active learning are irrelevant to the actual worst predicted strain paths, no matter if two, three or four committee members are employed. A big discrepancy between Cases 1 and 2 lies that in Case 1, the initial committee DNNs have been trained based on uniformly distributed specimens and these models have learned sufficient knowledge about the constitutive relation (with average MAE:0.03; average score: 0.944); while the initial committee DNNs in Case 2 are not trained and know nothing about the stress-strain mapping. The sharp contrast in the performance of active learning in Cases 1 and 2 indicates that only the DNNs which have learned sufficient knowledge are qualified to serve as committee members.

3.4. Batch-mode active learning scheme

For the committee-based active learning scheme, the time required to train DNN models cannot be ignored. If the most informative data is queried in serial, i.e., one at a time, the total cost of training DNN models will be practically unaffordable. To strike a balance between efficiency and labelling accuracy, a batch-mode active learning is adopted by selecting a group of specimens with the largest uncertainty to label each round and thus reducing the number of training committee DNNs. However, one should consider the possible information redundancy among the selected "worst predicted" specimens, because labelling the most hard-to-predict specimen first may eliminate the need to label the "neighbouring" cases in a batch of selected specimens.

To address this concern, a batch sampling scheme is proposed as shown in Fig. 7. When we have selected a certain point, its neighbouring points within an influence radius of R will be regarded as invalid and are not considered in the current batch. However,

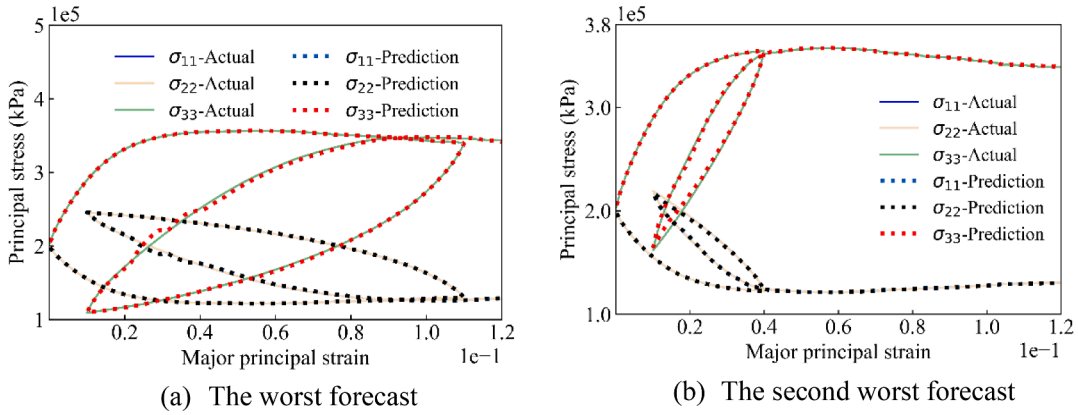


Fig. 9. The worst stress predictions given by a DNN model in Case 1 with 60 groups of training data.

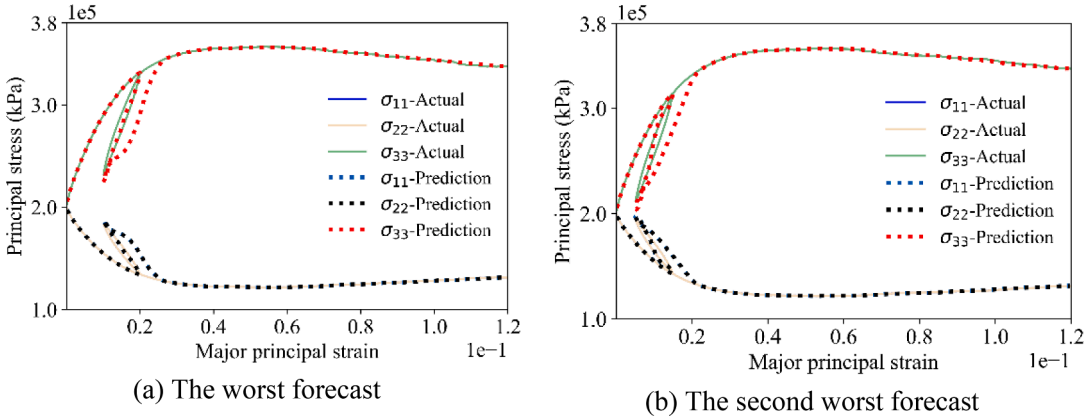


Fig. 10. The worst stress predictions given by a DNN model in Case 3 with 60 groups of training data.

these invalid points in the current round can still be selected in the next round. In this study, R is empirically selected as $1.5Ls$, where Ls is the sampling interval.

3.5. *Prediction performance of three training cases*

The prediction performance of the three different training cases is shown in Fig. 8. To reduce the possible influence of randomness, three neural networks with the same architecture, hyperparameters, and training data but different initialisations in weights and biases are trained. The mean and standard deviation of these three different DNN predictions are considered. When the number of training specimens is less than 35 groups, the DNNs from Case 3 outperform those in Case 2, demonstrating that the committee-based active learning algorithm is not necessarily useful when each committee member DNN has not learnt sufficient knowledge.

When the training number is larger than 35, the prediction accuracy of Cases 1 and 2 outperforms those in Case 3. Figs. 9 and 10 demonstrate the worst and second worst forecasts given by a DNN in Cases 1 and 3 when 60 groups of training datasets are used. Although it appears that only a small discrepancy in the forecasted score and MAE is found, as shown in Fig. 8, a relatively large difference in the actual stress-strain predictions can be observed. Particularly, in the randomly inputted training datasets in Case 3, DNNs' generalisation error decreases relatively slowly and 30 groups of extra training specimens (almost ¼ of the total number) are required to reach a similar predictive level in Cases 1 and 2. These results confirm that active learning can train a good predictive model by using less amount of data.

In addition to the comparison of stress responses, we also perform a comparative investigation on volumetric predictions with and without active learning. One is similar to Case 1 where we start from a randomly selected 30 groups of data, and then active learning is employed to add five groups of the most informative specimens in each round, i.e. the "random + active learning" scheme. The other one is simply the conventional training scheme with an increasing number of randomly selected specimens. All network architectures and hyperparameters inherit those from the former stress prediction case.

The overall prediction performance is demonstrated in Fig. 11, where the predictions with active learning keep a steady low MAE range while those with conventional training scheme fluctuate at a relatively high forecast error. The results show that the volumetric predictions with active learning significantly outperform those with the conventional training scheme. Fig. 12 shows that the worst

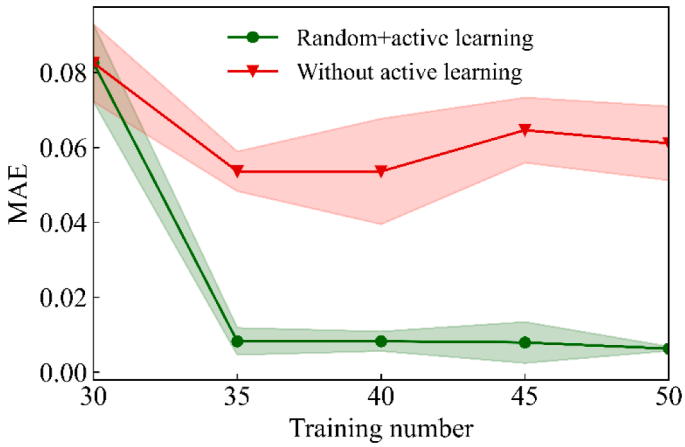


Fig. 11. A comparison of volumetric prediction performance with and without active learning.

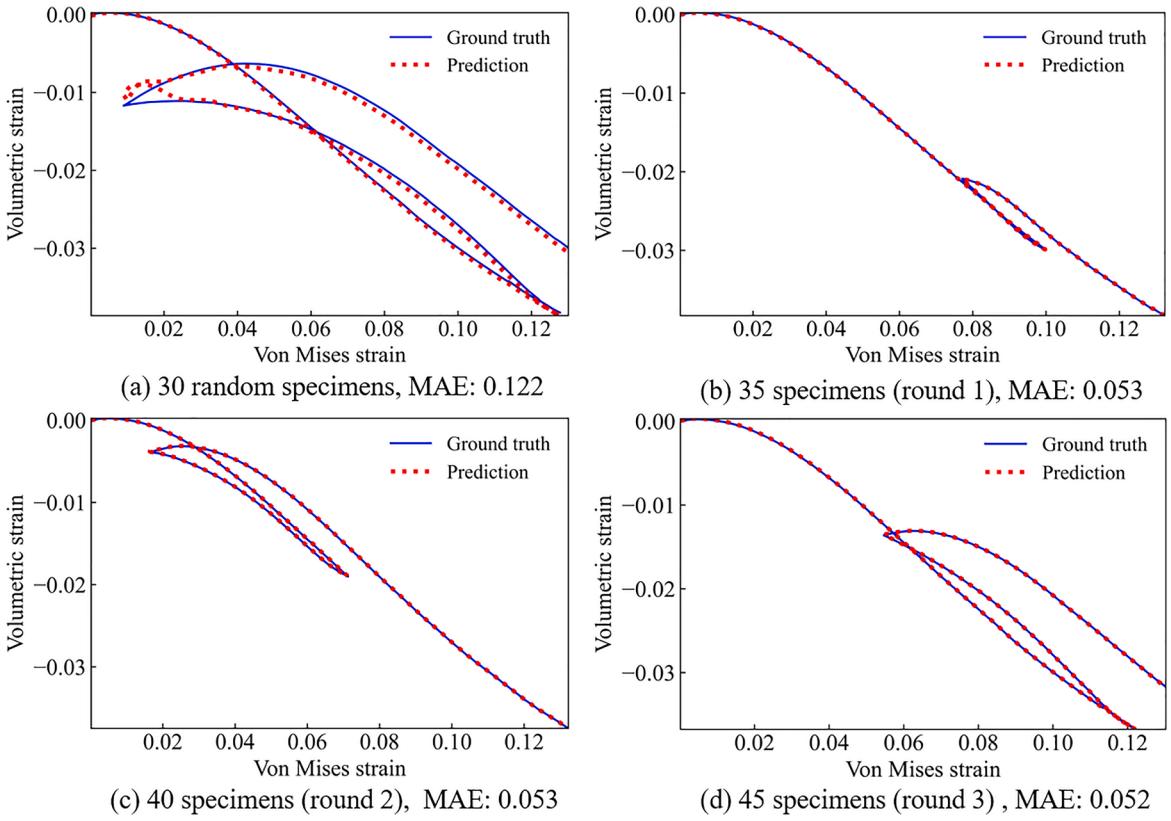


Fig. 12. The worst volumetric predictions of the "random+active learning" scheme with an increase in training number.

volumetric predictions with an increasing training number for the active learning case. It is found that the worst volumetric responses cannot be satisfactorily predicted when only 30 groups of random data specimens are used. However, with the use of active learning, even the worst volumetric forecast after one-round of active sampling behaves excellently over unseen test specimens. This explains why the DNN's inference performance maintains a steadily high level with an increasing training number in Fig. 11. The results further confirm the capability of active learning in discovering the most informative specimen during data-driven constitutive training.

3.6. Active learning-informed data preparation

With the aid of active learning, DNNs can automatically prioritise labelling the most instructive data to maximise the inference

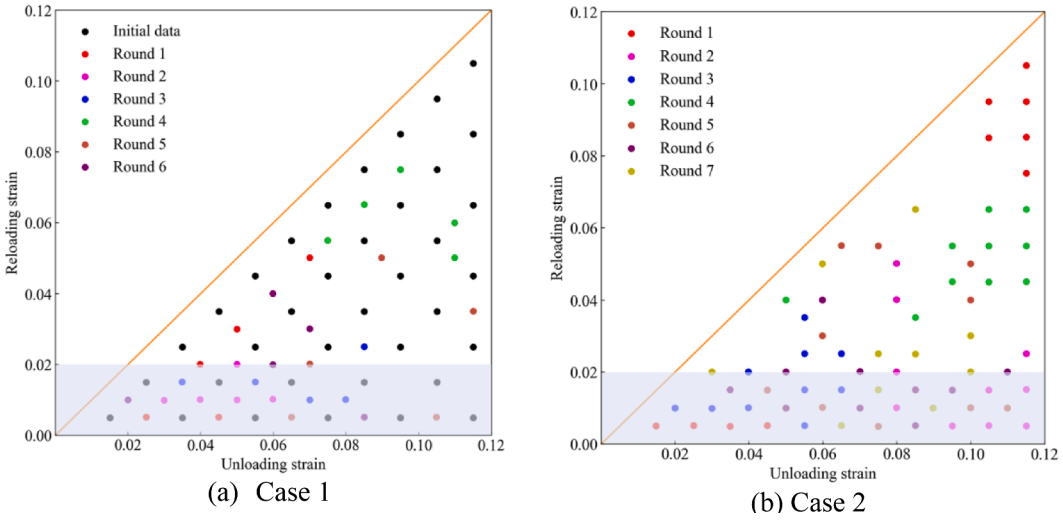


Fig. 13. The added data specimens via active learning in each round.

performance of models. We summarise in Fig. 13 the intelligently sampled data recognised by active learning in each round for Cases 1 and 2. In Case 1, the top five groups of new specimens are selected in each round and a total of six rounds of active learning querying are included. In Case 2, the top ten groups of datasets in each round are selected because all the training data is obtained by active querying and labelling a relatively large number of specimens in each round can reduce the overall time costs. A total of seven active learning querying rounds are listed in Fig. 13b.

Fig. 13 may suggest that the specimens with reloading strain lower than 2% are probably the most informative data for training. These specimens represent a relatively larger unloading-reloading loop. The prediction is similar to extrapolation if we infer large unloading-reloading loops with only small-loop data because a large loop may include richer information than small loops. Furthermore, Fig. 13a indicates that the loops at the elastic-plastic transition phase (a major principal strain ranging from 0.03 to 0.07) may outweigh those at the critical stage where the stress remains constant with the increasing strain. The reason can be attributed to the fact that the adopted training data has a relatively long and steady critical state (see Fig. 10), and thus the information density carried by the data points at the post-peak state is lower than those at the elastic-plastic transition stage.

By observing the first and second rounds of selected data in Fig. 13b, it is found that the selected data always occur in clusters. This phenomenon may explain why the active learning algorithm may underperform conventional passive learning when the surrogate DNNs do not learn the real relation at all. The reasons can be that the active learning algorithm may converge to local or short-sighted optimal solutions due to the greedy nature of the algorithm. A greedy algorithm often fails to make the best decision although it always goes for the local best choice at each iteration.

4. Interactive constitutive training and data labelling through active learning

In Section 3, all the stress-strain data is created before training models. The computational costs for preparing data are thus not reduced. In this section, the training of DNNs and the generation of data via microscale DEM simulations are performed in an interactive manner. Active learning is used to judiciously select data and only the most informative data is labelled, aiming to reduce the overall size of training datasets.

4.1. Strain path falsification

In continuum-based numerical computations, the constitutive relation of a specific material receives a strain tensor and yields a corresponding stress tensor. The premise of using active learning is that inputs $\mathbf{X}$ or unlabelled datasets $\boldsymbol{D}_u$ are available. Thus, the primary concern of applying active learning in constitutive modelling is whether the strain sequences can be constructed to generate an unlabelled data pool. Yet, granular media is a typical material which cannot undergo tensile and large compressive deformation. The admissible deformation scope for granular materials is restricted to a relatively narrow strain-stress space, compared to those of other solid materials (e.g., rubber and metal). In addition, the most common strain-stress path of granular materials experienced in a BVP is shear-type deformation, which is not easy to be artificially constructed through either proportional or random loading in the context of full-strain-dominated loading conditions.

Inspired by conventional triaxial experiments of granular soils, the *in-situ* stress condition of soils existing in the ground is approximated by imposing a constant confining pressure. An axial strain is applied to represent external disturbances on specimens. Such a hybrid boundary condition simplifies the complexity of artificially constructing admissible deformation for a granular specimen because the constant confining pressure ensures that the granular specimen does not undergo tensile or excessive isotropic

compression loads. Also, only the one-dimensional axial strain needs to be fabricated. In this section, we adopt the most common experiment in geotechnical engineering as the prototype to perform deep active learning.

4.2. A surrogate error indicator for data-driven constitutive modelling

Although the granular specimen is not perfectly isotropic, especially during the loading process, we implicitly incorporate the isotropic assumption by homogenising stress responses in a triaxial testing condition. For an isotropic material, the strain state at a point in a 3D space should be depicted through at least three components (principal variables or invariants). When performing data-driven constitutive modelling, it is crucial to keep complete strain components (e.g. three principal strains) as inputs. For a conventional triaxial testing condition, the axial strain path can be falsified readily, but the lateral strain relates to the properties of materials and is thus difficult to construct artificially. Although the stress-strain prediction model should obey the relation shown in Fig. 14a: [Axial strain, lateral strain]→[axial stress], we introduce a new mapping relation which links the axial strain and the axial stress directly (see Fig. 14b). The mapping requires only axial strain sequences as inputs thus one can generate extensive strain paths as unlabelled datasets.

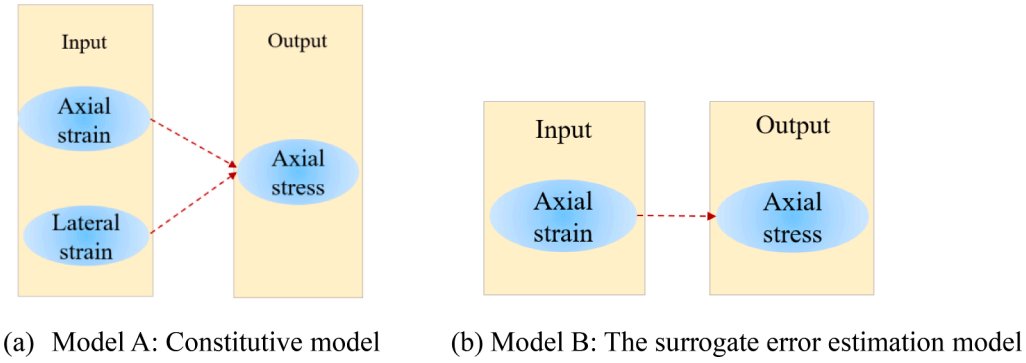


Fig. 14. Inputs and outputs for data-driven stress-strain modelling in a conventional triaxial testing condition.

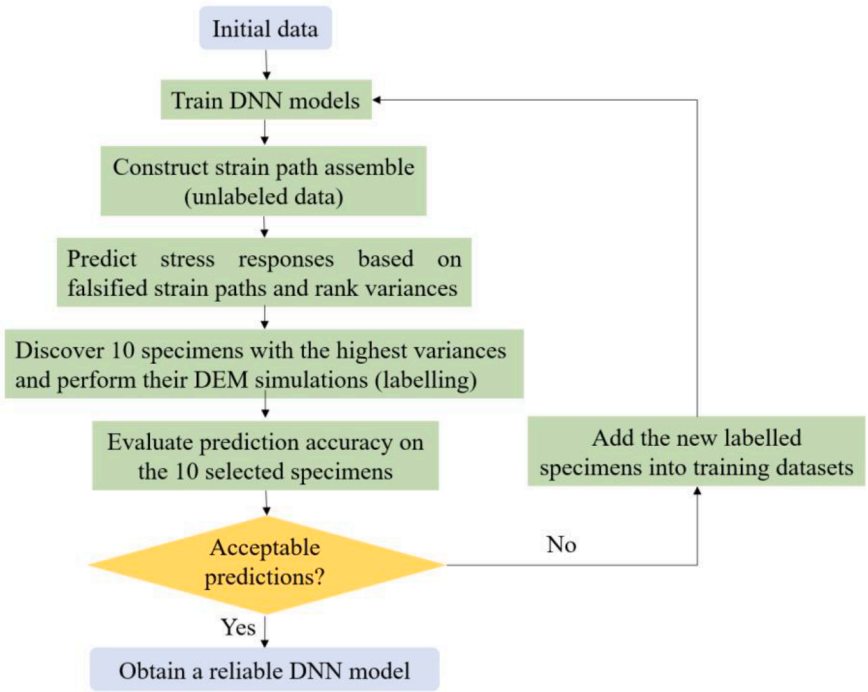
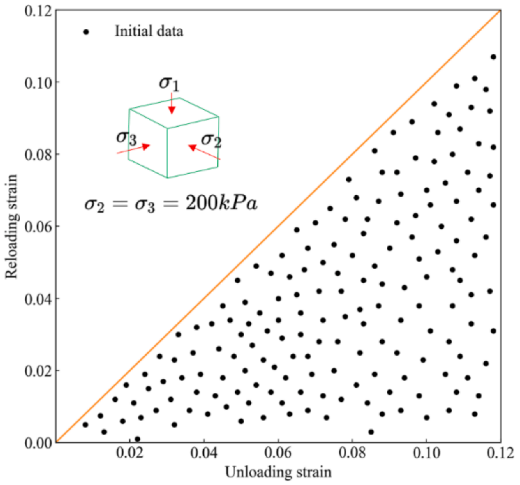
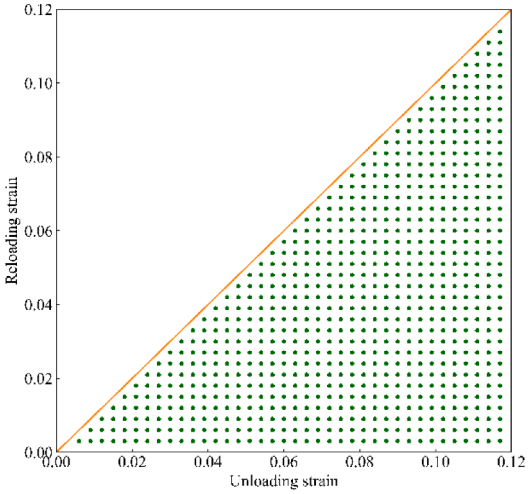


Fig. 15. The workflow of global domain examination and resampling strategy.



(a) The random sampling distribution for training an initial DNN model



(b) A sampling space of unlabelled strain paths

Fig. 16. Sampling space for training and examination.

The underlying idea is that both DNNs in Fig. 14 suffer from the same data scarcity issue, provided that they share the same training data. The predictions around the region where the training data is insufficient will have a larger variance than the region where data is sufficiently covered. Model B in Fig. 14b is not the desired candidate to forecast the stress-strain behaviour, but the active learning strategy introduced in Section 3 enables it to be a surrogate error indicator to identify the most helpful data for improving the current DNNs. Then these identified strain paths will be imposed on DEM specimens as the deformation of the boundaries to generate corresponding stress responses as new training data. Model A in Fig. 14a will make use of the resampled data to improve it. These procedures are repeated until a satisfactory predictive model has been fitted.

4.3. Examination of the whole domain and adaptive resampling: an example with varied unloading-reloading cycles

The stress-strain curves with varied unloading-reloading loops are challenging to be predicted by traditional constitutive models. In this section, conventional triaxial loading cases with unloading-reloading loops are adopted to examine the capability of active learning in distinguishing complex strain paths. Constitutive modelling of path-dependent materials is a typical time series problem. To obtain unlabelled strain paths with unloading-reloading cycles, we artificially construct strain paths passing through selected unloading and reloading points, and then interpolate data points with equal spacing. These unloading and reloading strains can be described with horizontal and vertical coordinates in Cartesian coordinates, as shown in Section 3. For a strain-dominated loading condition, such a falsification of the axial strain path resembles the actual strain sequence in laboratory experiments, provided that a similar interval is selected. Through a large number of artificial strain paths, we can develop a global domain examination and resampling strategy. The workflow is shown in Fig. 15 and the detailed procedure is described below:

Table 4
The prediction performance of the selected ten samples with the greatest standard deviations but extracting neighbouring points.

Unloading strain	Reloading strain	Standard deviation	Score	MAE
0.036	0.006	0.316	0.795	0.041
0.027	0.006	0.313	0.759	0.047
0.054	0.003	0.313	0.605	0.055
0.018	0.003	0.310	0.519	0.081
0.060	0.003	0.309	0.780	0.044
0.117	0.006	0.308	0.653	0.053
0.048	0.003	0.305	0.706	0.043
0.042	0.006	0.304	0.783	0.044
0.072	0.003	0.302	0.907	0.028
0.066	0.003	0.302	0.559	0.065

(1) Train a constitutive model (Model A) and three surrogate error models (Model B) based on the available stress-strain data pool.
(2) Choose unloading-reloading strain points with a desired resolution in the whole domain, and then construct the strain sequence (unlabelled data) with even interpolation.
(3) Utilise the three trained surrogate error indicators (Model B) with diverse initial weights and biases to predict stress values for these falsified strain paths.
(4) Calculate the standard deviation (Eq. (1)) of each stress sequence prediction and rank the standard deviations of all strain sequences.
(5) Select ten strain paths with the largest standard deviations from the previously ranked data pool using the batch-mode scheme (Fig. 6).
(6) Perform DEM simulations using the ten selected strain paths as boundary conditions to generate corresponding stress responses. This procedure aims to tag the most instructive unlabelled data specimens.
(7) Use the trained constitutive model (Model A) to forecast the ten newly generated DEM specimens and evaluate the prediction accuracy. In the case that all the predictions have reached a satisfactory level, the current DNN model can be regarded as a reliable constitutive model. Otherwise, add these newly generated data to the training datasets, and repeat steps (1)–(7).

In the above workflow, Models A and B share the same network architecture and hyperparameters, but Model B only inherits part of the input features from Model A. Note that the number of network layers and neurons relates to the amount of training data and in principle, the DNN architecture and hyperparameters in each training phase for different models should be modified, because the optimal architecture may change slightly with a variation in training specimens. Yet the model architecture and hyperparameters are currently kept constant to reduce expensive tentative training costs when tuning these hyperparameters. In the future, some available packages, especially new algorithms in autoML, deserve to be explored to tune the network architecture automatically.

4.4. Verification of the interactive learning strategy

We consider training a data-driven model aiming to capture strain-stress mapping in a conventional testing condition with mutually different unloading-reloading points. The initial training data is randomly sampled as shown in Fig. 16a. Three DNNs fitted with these initial specimens are employed to examine unreliable predictions in the global domain with a strain interval of 0.003 shown in Fig. 16b.

All the training data is developed via DEM simulations of conventional triaxial testing. The simulated parameters are the same as the constant-p true triaxial testing in Section 3. The confining pressure is kept to 200 kPa during testing. A GRU network is used as a base DNN for constitutive training. Based on the initial random specimens shown in Fig. 16, the initial model, obtained by trial-and-error, includes two GRU hidden layers with 60 neurons and a dropout rate of 0.02 for each layer, resulting in a total of 33,843 weights and biases parameters. The two GRU layers adopt the *tanh* activation function while the output layer uses a linear activation. The optimiser is the adaptive moment estimation (Adam) and the learning rate is 0.01. These networks are trained for 2000 epochs with a batch size of 256. MAE, as defined in Eq. (3), is used as the loss function.

Following the procedure described in Section 4.3, we train Model A for constitutive predictions while three error indication models (Model B) for seeking ten groups of most unreliable predictions. The ground-truth stress responses are obtained by performing DEM simulations of the selected strain paths. The forecasting accuracy of the selected samples is evaluated by comparing the predictions given by Model A and the ground truth from DEM simulations. The forecast results are shown in Table 4. Note that the first three columns are discovered by Model B while the last two columns are verified by Model A. The results demonstrate that almost all the predictions on these specimens are not sufficiently accurate and the average prediction score reaches only 0.71. In the next round, these new specimens will be added to the training samples and used for re-training DNNs.

To verify the proposed interactive constitutive training scheme, we also choose ten strain paths with the smallest standard deviations. The corresponding stress responses of these strain paths are also obtained via DEM simulations. The batch-mode scheme is used to create a wider range of representatives. Table 5 gives the prediction performances of these specimens. The results show that all the specimens are satisfactorily forecasted with a full score.

The comparison between Tables 4 and 5 reveals that the proposed surrogate error indication model is useful for searching for the

Table 5
The prediction performance on the ten specimens with the smallest standard deviations.

Unloading strain	Reloading strain	Standard deviation	Score	MAE
0.075	0.066	0.041	1.0	0.011
0.063	0.054	0.044	1.0	0.013
0.069	0.057	0.046	1.0	0.013
0.081	0.069	0.061	1.0	0.009
0.060	0.048	0.062	1.0	0.011
0.069	0.063	0.064	1.0	0.017
0.051	0.042	0.064	1.0	0.010
0.081	0.075	0.067	1.0	0.012
0.057	0.042	0.069	1.0	0.017
0.063	0.060	0.069	1.0	0.016

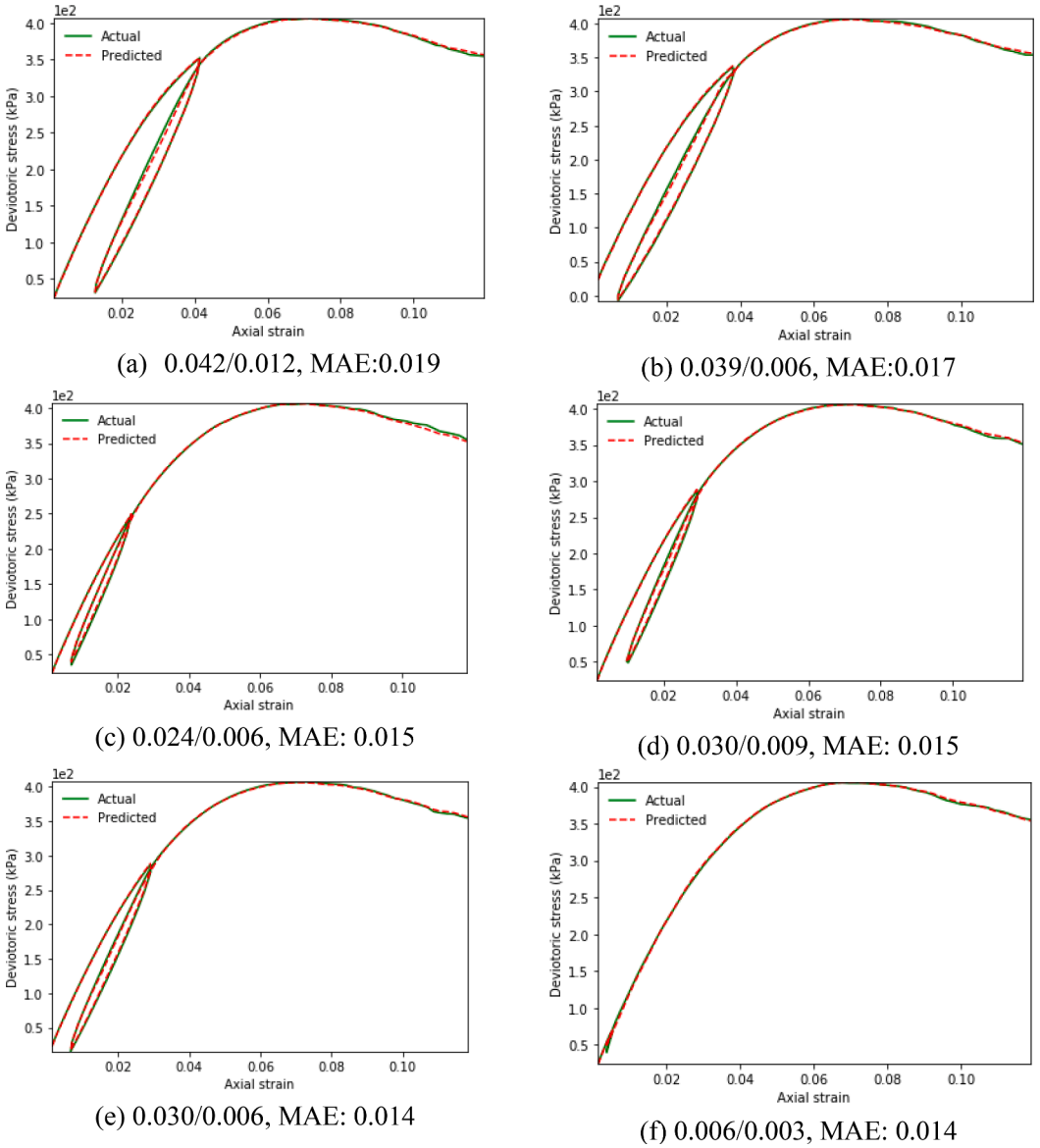


Fig. 17. The worst forecasts discovered in the sixth round of active learning.

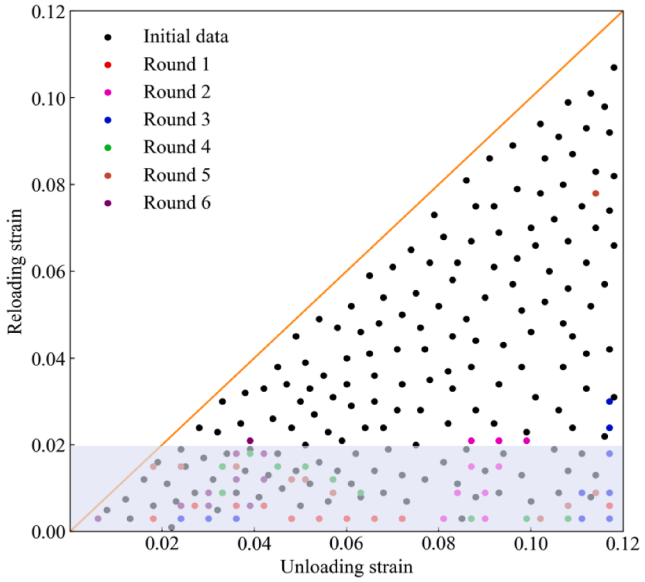


Fig. 18. The newly added data specimens in each active learning round.

weakness of DNN models, even though actual stress-strain responses are unknown. For the committee-based active learning scheme, the main computational costs occur in training three DNNs as committee members (~0.5 h per DNN), while the querying process can be finished in a few seconds. In contrast, it takes a few hours to run a group of DEM specimens. By reducing the number of generating DEM specimens, the cost of training a reliable DNN model will be greatly saved.

By repeating the global domain examination and resampling scheme, the prediction accuracy of the data-driven constitutive model will be gradually enhanced. This repetition process terminates when the ten discovered specimens with the largest variances can be satisfactorily forecasted. In our current model, applying six rounds of resampling is found to be able to develop a sufficiently reliable data-driven constitutive model over the domain of interest. The data during each prediction examination and resampling are recorded in Tables A1-A5 in Appendix. The worst forecasts after the sixth round of active learning are shown in Fig. 17.

The newly added training specimens during each active learning round can be found in Fig. 18. The results show that the majority of newly added data are located in the shaded domain where the reloading strain is lower than 0.02. These points represent relatively large unloading-reloading loops. In contrast, only one specimen is selected as a reloading strain larger than 0.04. Again, the results confirm the importance of labelling the stress-strain curves with large unloading-reloading cycles.

4.5. An examination of generalisation ability: can we predict multiple unloading-reloading cycles with single unloading-reloading data?

After a DNN model has been proven to be able to predict single unloading-reloading cycles, a question arises: can we predict multiple quasi-static unloading-reloading cycles with single unloading-reloading data? Here the quasi-static unloading-reloading means that the inertia forces are not present in the specimens and that at each time step the specimen can be regarded as a static equilibrium state.

This question is significant for developing a more generalised data-driven constitutive model, as many engineering models may incorporate multiple unloading-reloading disturbances, but it is highly time-consuming and impractical to prepare training data covering all possible unloading-reloading cycles.

In this section, multiple specimens with two, three and more quasi-static unloading-reloading cycles are modelled via DEM. The trained model in Section 4.4 is used to predict these complex mechanical responses. The results show that the stress-strain curves with multiple unloading-reloading cycles can be captured satisfactorily by the trained model with only a single-loop training data, although the prediction deteriorates as the number of loops increases. Fig. 19 lists some of the representative predictions.

An important microscopic origin of the elastic-plastic behaviour of granular media is that the external deformation paths make internal grains move around each other permanently. These irreversible movements alter the inherent microstructures of granular assemblies and thus affect the subsequent stress-strain relations. Provided that the evolution of microstructures under certain external deformation conditions can be implicitly captured by a DNN model, a deep learning model trained with a single unloading-reloading cycle should capture a similar evolution of inherent microstructures. It should be emphasised, however, that the conclusion holds only when particle breakage or degradation is absent during loading. The irreversible damage of particle grains during unloading-reloading cycles will make stress prediction more challenging. This problem is not investigated in the current work.

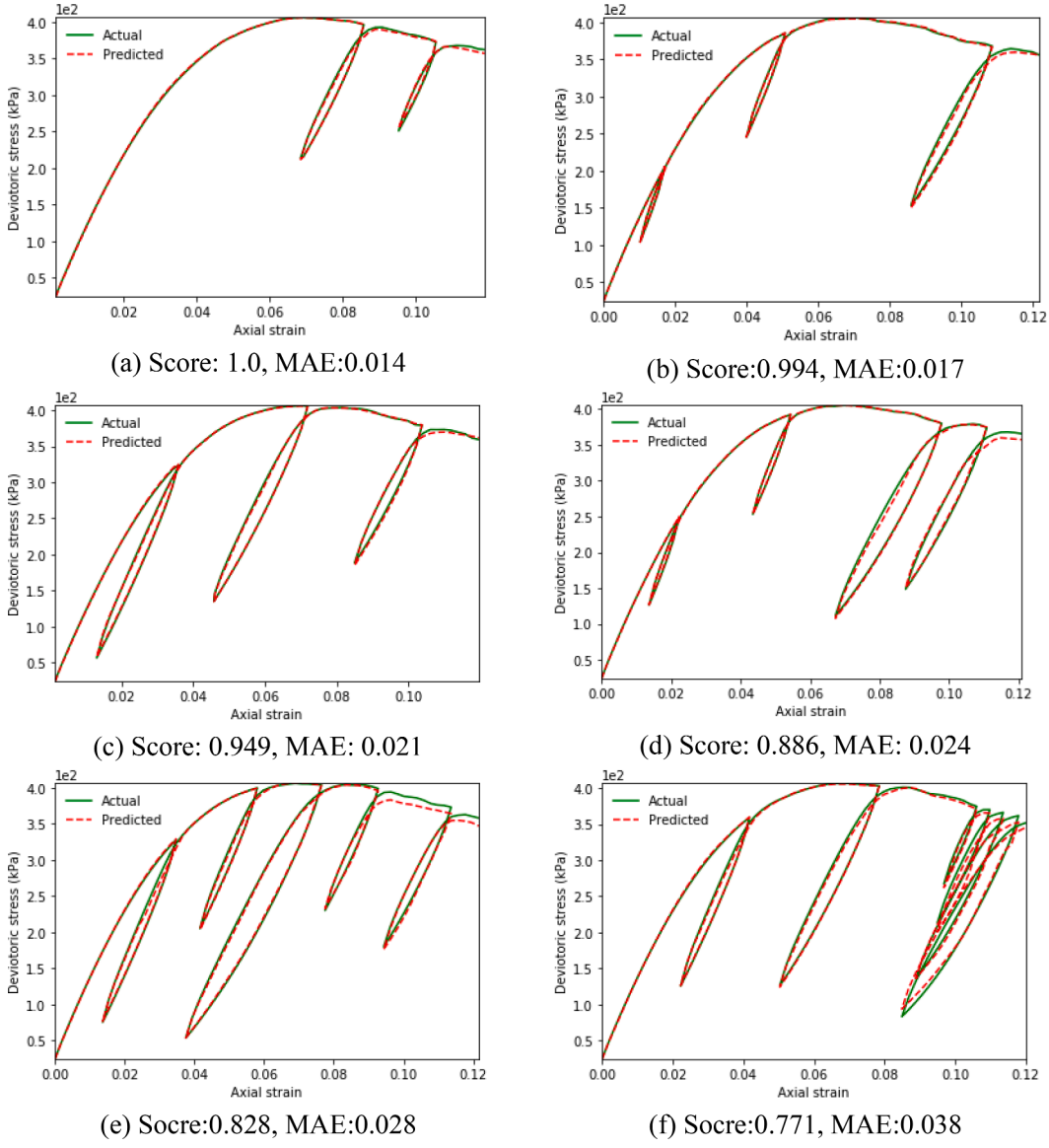


Fig. 19. Representative forecasts with multiple unloading-reloading loops.

5. Deep active learning-based data-driven multiscale modelling

This section aims to demonstrate the application of active learning in data-driven hierarchical multiscale modelling (HMM) of granular materials. The HMM scheme coupling FEM and DEM (FD-HMM for short) is a computational framework which utilises online DEM simulations to replace phenomenological constitutive models in FEM computations (Guo and Zhao, 2014). In this section, we leverage HMM to generate a large number of stress-strain sequence pairs with ample and realistically experienced strain paths in BVPs. The generated data is used to train deep learning-based constitutive models. When transferring a DNN-based constitutive model to macroscopic FEM computations of a BVP, the DNN model may suffer from new mispredictions. We try to harness active learning to automatically detect potentially unreliable predictions and improve them until the model meets the prescribed accuracy requirements.

5.1. Hierarchical multiscale modelling coupled FEM and DEM

For the FD-HMM, the macroscopic BVPs are simulated via FEM while the loading path-dependent constitutive relation at each Gauss point is provided by grain-scale DEM solutions of RVE specimens. The solution procedure is illustrated in Fig. 20. Specifically, DEM receives the global deformation at Gauss points from FEM as boundary conditions for the RVEs and solves them to obtain the corresponding stress responses and tangent operators which are passed on to Gauss points for FEM computations. This procedure is

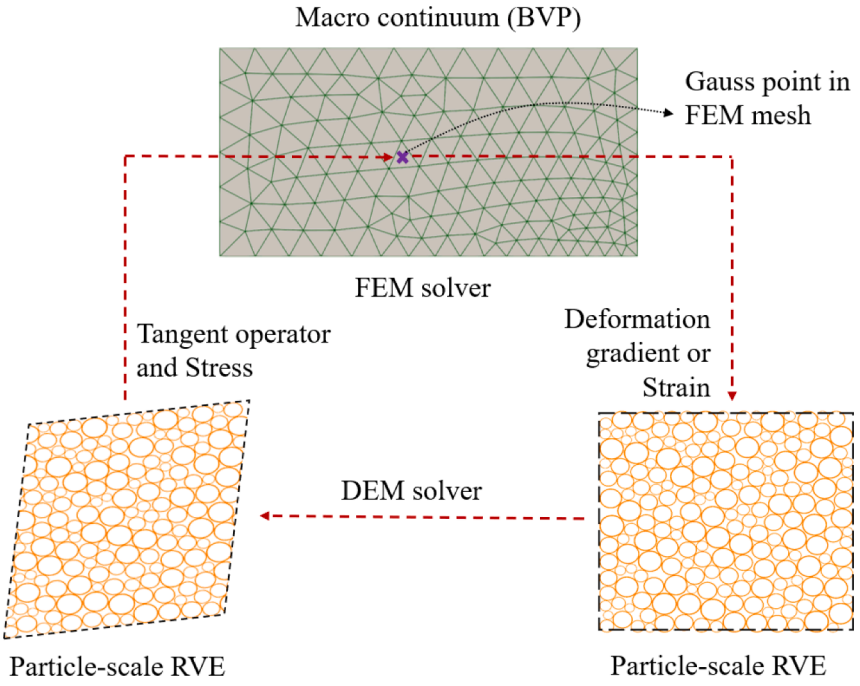


Fig. 20. The computation procedures of FEM×DEM hierarchical multiscale modelling (adapted from Guo and Zhao 2014).

repeated until the end of the simulation.

This hierarchical framework circumvents the phenomenological assumptions on the constitutive relation of materials in conventional continuum modelling and has been a widely accepted method to simulate large-scale BVPs in granular soils. However, despite unique advantages, the FD-HMM has to solve massive RVEs with DEM. The prohibitively expensive costs required, have restricted its further application in engineering scale BVPs. In addition to parallel computing (Guo and Zhao, 2016b; Zhao et al., 2021), data-driven multiscale modelling is also an efficient approach to accelerating the computation of FD-HMM (Karapiperis et al., 2021).

5.2. *Data-driven multiscale modelling coupled FEM and deep learning models*

RNNs have a strong fitting capability but require many parameters to be trained and a long inference time as well. In contrast, MLP is more efficient but has to introduce extra internal variables to capture the loading-path dependence of strain-stress responses. One option is by using the accumulation of absolute strain increments $\boldsymbol{\varphi}$ to encode the loading history (Guan et al., 2022; Huang et al., 2020):

$$\varphi_j^{(t)} = \sum \left|\Delta\varepsilon_j^{(t)}\right| = \begin{cases} \left|\varepsilon_j^{(t)}\right|, t = 1 \\ \left|\varepsilon_j^{(t)}\right| + \sum_{t=2}^{(T)} \left|\varepsilon_j^{(t)} - \varepsilon_j^{(t-1)}\right|, t \subseteq [2, T] \end{cases} \tag{4}$$

where $\varphi_j^{(t)}$, $\Delta\varepsilon_j^{(t)}$ and $\varepsilon_j^{(t)}$ represent the j^{th} components of the internal variable, the strain increment, and the absolute strain, respectively, at the time $t \subseteq [1, T]$. Note here t represents the time step, rather than a physical time metric. In this section, an MLP-based neutral network is fitted for constitutive modelling by taking advantage of its high efficiency in prediction and demonstrating the wide applicability of active learning in various base models.

Another consideration is that when employing implicit FEM for computations, the tangent operator **D** at each Gaussian point is required to compute the element stiffness matrix $\mathbf{K}^{\text{e}}$

$$\boldsymbol{K}^e = \int_{\Omega^e} \boldsymbol{B}^T \boldsymbol{D} \boldsymbol{B} \mathrm{d}\Omega^e \tag{5}$$

where **B** is the gradient of shape function, which can be determined provided that a specific element type is selected; Ω^{e} denotes the element domain. In the FD-HMM framework, the tangent operator **D** can be approximated with the microstructure and particle-scale stiffness parameters in DEM based on some homogenisation methods. The analytical formulation based on Vogit's homogenisation (uniform strain field) is given:

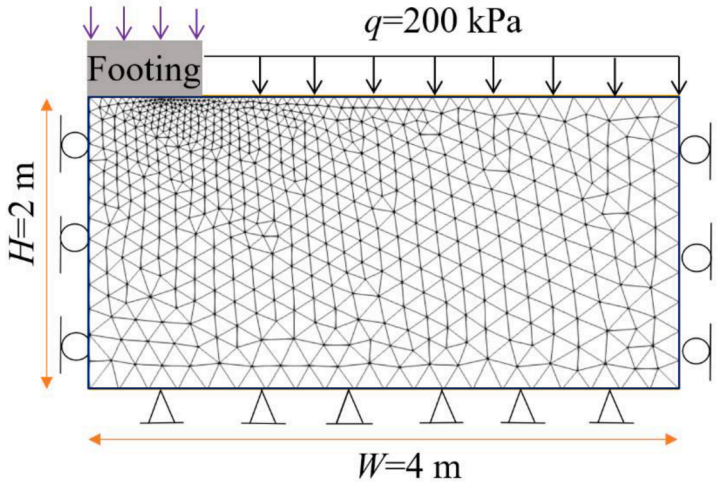


Fig. 21. Mesh partition and boundary conditions for the strip footing.

Table 6
DEM parameters for the footing model.

Radii (mm)	Density (kg/m^3)	Young's modulus (Pa)	Poisson's ratio	Frictional coefficient	Confining pressure (kPa)
2-4	2,600	6×10^8	0.2	0.5	200

$$\mathbf{D}=\frac{1}{V}\sum_{N_c}\left(k_n\boldsymbol{n}^c\otimes\boldsymbol{d}^c\otimes\boldsymbol{n}^c\otimes\boldsymbol{d}^c+k_s\boldsymbol{t}^c\otimes\boldsymbol{d}^c\otimes\boldsymbol{t}^c\otimes\boldsymbol{d}^c\right) \tag{6}$$

where N_c denotes the total number of mechanical contacts within the RVE specimen; V represents the volume of the RVE; k_n and k_s are interparticle normal and tangential contact stiffnesses, respectively; $\boldsymbol{n}^c$ and $\boldsymbol{t}^c$ are unit vectors in the normal and tangential directions of a contact plane, respectively; $\boldsymbol{d}^c$ is the branch vector connecting the centroids of two contacting particles. The tangent operator $\mathbf{D}$ can be obtained directly from the corresponding DEM simulations for each Gauss point.

In the context of data-driven hierarchical multiscale modelling, $\mathbf{D}$ should be provided by the trained neural network. One practically feasible way is to include $\mathbf{D}$ as an additional output variable. Recalling the former statement of incorporating the internal variable $\boldsymbol{\varphi}$ as the extra input, now the MLP-based mapping relation becomes:

$$\left\{\widehat{\boldsymbol{D}}^{(t)},\widehat{\boldsymbol{\sigma}}^{(t)}\right\}=\mathcal{NN}\left(\epsilon^{(t)},\boldsymbol{\varphi}^{(t)}\right) \tag{7}$$

where $\widehat{\boldsymbol{D}}^{(t)}$ and $\widehat{\boldsymbol{\sigma}}^{(t)}$ represent the predicted tangent operator and stress responses at the time t, respectively.

5.3. Deep active learning-assisted multiscale modelling

5.3.1. The model configuration

A rigid strip footing problem is employed to exemplify the application of active learning in data-driven multiscale modelling. The geometrical configuration, boundary conditions and discretisation scheme are shown in Fig. 21. Given the symmetry of the problem concerned, only half of the domain is modelled to reduce computational costs. The widths of the half domain and half footing are 4 m and 0.5 m, respectively. The depth of the whole domain is 2 m. The horizontal displacements of both lateral boundaries are restrained while the bottom of the entire domain is fixed. A surcharge q of 200 kPa is applied on the ground surface of the footing. The footing is discretised with 402 second-order six-node triangular elements with a total of 1,206 Gaussian points. Each Gaussian point shares the same RVE packing in DEM simulations. Each RVE consists of 400 particles and a periodic boundary is adopted to reduce possible boundary effects. Thus 482,400 particles in total are involved in the HMM model. The particle-scale parameters used are listed in Table 6.

5.3.2. Active learning-assisted data-driven multiscale modelling

The physics-based FD-HMM is conducted following the framework developed by Guo and Zhao (2014). The FEM and DEM modules are computed by two open-source software Escript and Yade, respectively. In this section, an MLP is fitted and embedded into FEM to bypass DEM simulations in the data-driven multiscale modelling. The input and output features are given by Eq. (7). The

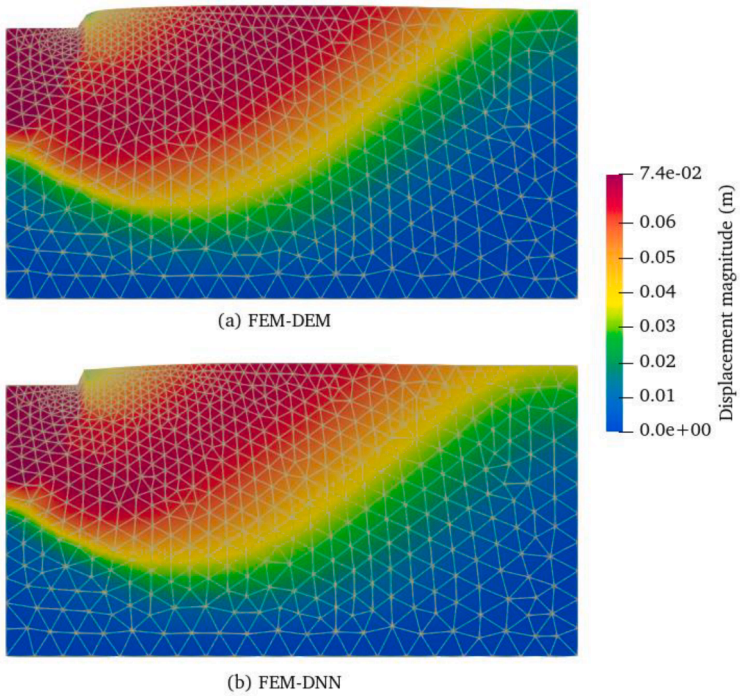


Fig. 22. Comparison of the distributions of the displacement field in (a) FEM-DEM and (b) FEM-DNN modelling.

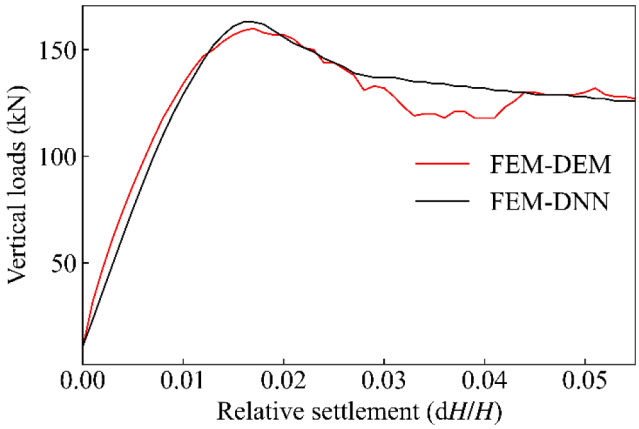


Fig. 23. Load-displacement curves in both FEM-DEM and FEM-DNN modelling.

computational procedure coincides with the standard implicit FEM routine while the duty of constitutive relation is assumed by a deep learning model.

The strain-stress and strain-tangent operator **D** pairs experienced by each Gaussian point in the multiscale model serve as training data. However, when performing simply 58 loading steps in FD-HMM, a large number of data pairs are generated. If all the available data is fed to a DNN model for training, one has to bear a massive training cost. Given that many of the datasets share a high degree of similarity, we initiate a two-stage scheme combined with the active learning algorithm.

In the first stage of active learning, three DNN models are trained using randomly extracted 5‰ of the total data points. Then active learning automatically identifies 5‰ more data points which are the worst forecasted by the current DNN model. In the second stage, these newly discovered data points are added to the training dataset (now 1% in total) and a new DNN model is fitted based on the updated data.

Note that although **D** is also an output, only the predicted stress is used as the indicator to prioritise the selection of the most informative training datasets. This is because the stress tensor and tangent operator **D** are interconnected and paired, and the data deficiency informed by stress-based active learning will also disclose such a situation for **D** matrix. Considering that **D** is a 4×4 tensor with 16 components, even in a 2-dimensional condition, the bypass of **D** during the active learning process significantly reduces

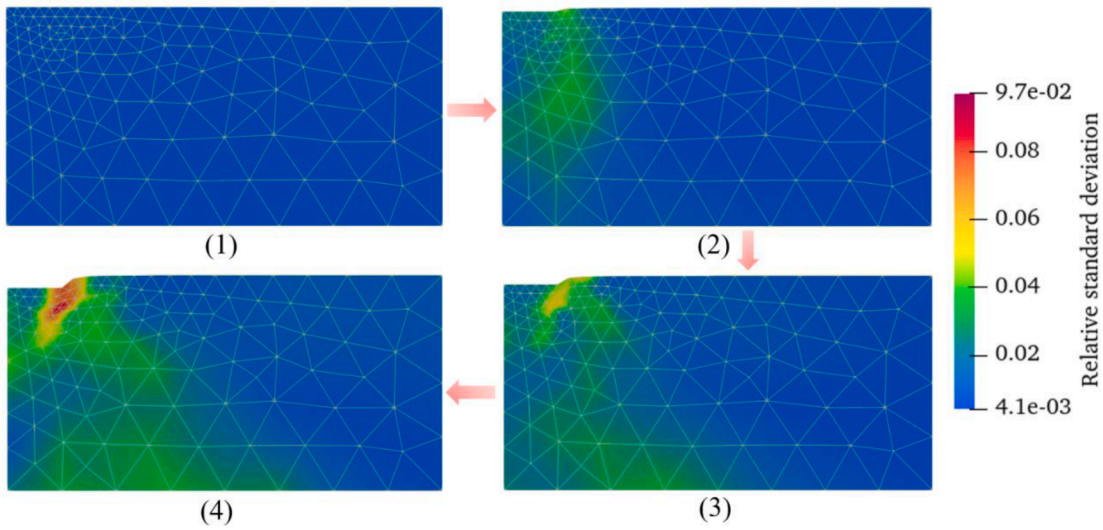


Fig. 24. Active learning-assisted detection of relative forecast errors in FEM-DNN simulations at four different loading steps.

computational costs.

When embedding the new DNN into FEM, it is found that the active learning-enhanced DNN can drive FEM computations satisfactorily. Fig. 22 compares the distributions of the displacement field between both FEM-DEM and FEM-DNN modelling after the loading. Fig. 23 demonstrates a quantitative comparison of the two load-displacement curves. All the results show that the data-driven constitutive model with a small training dataset has reproduced a high-fidelity simulation with the help of active learning. In addition, FD-HMM consumes 28 h 21 mins while the FEM-DNN scheme requires 7.7 mins to finish the whole computational process. The data-driven multiscale model exhibits nearly 211 times of acceleration than the FD-HMM model.

6. Discussion

6.1. *Significance of active learning*

(1) *We know when our model does not know*

When performing FEM simulations with data-driven constitutive models, a major concern is whether such a FEM simulation is sufficiently reliable. Following the idea of committee-based active learning, we can train multiple surrogate models simultaneously based on available datasets. If the predicted responses given by different DNNs vary from each other with a relatively large discrepancy, there might be a high risk of mispredictions. Then a procedure of labelling these strain paths is taken to enrich training datasets. The DNN model is refitted based on the new training dataset to improve the generalisation capability. Active learning enables DNN models to be self-learning with an intelligent sampling scheme.

Furthermore, in DNN-based FEM computations, the initiation and evolvement of possible errors can be estimated by active learning and visualised in real time. Fig. 24 shows the distribution and evolution of potential forecast errors in a rigid strip footing simulation via committee-based active learning. The relative standard deviations of predicted stress given by three different committee member DNNs at each Gauss point are utilised to indicate the magnitude of prediction errors. Four different loading steps are used to demonstrate the initiation and development of errors during loading. The figure suggests that the domain where large shear deformation occurs is most likely to be mis-predicted, which is consistent with our intuition. The results indicate that active learning can enable FEM simulations with data-driven constitutive models to identify unreliable forecasts on their own, i.e. "*we know when our model does not know*". This is of significance to boost the confidence in developing trust-worthy data-driven models for scientific computations.

(2) Developing small-data-driven models with better generalisation capability

The data-hungry nature and the susceptibility to misprediction are two open challenges for a data-centric surrogate model. This study has confirmed the potential of active learning for developing a reliable and cost-effective data-driven constitutive model for granular materials. At each round of training, DNNs always discover and label the most hard-to-predict strain paths. This feature bypasses the need to label the less informative data. In addition, active learning can be a useful tool to combat imbalanced data, where most supervised learning models ignore and in turn tend to mis-predict the minority class, although the predictions of these minority datasets often dominate the success of the trained model. Active learning will automatically select these minority data if the trained DNNs mis-predict them.

6.2. Limitation of active learning and future work

(1) Strain paths must be available to reduce labelling costs

A prerequisite of implementing active learning for data-driven constitutive modelling is that a large number of strain paths should be available. Although a fraction of them can be artificially constructed as demonstrated in Section 4, most strain paths experienced by Gauss points in a BVP are intertwined with stress responses and are thus hard to be falsified. This limitation no doubt restricts active learning from bringing possible disruptive progress for data-driven constitutive modelling. To address this issue, a possible remedy is by exploiting advanced machine learning tools, such as deep generative models, to learn the probability distribution of real strain-stress paths in typical BVPs. Then the learned generative model can produce massive strain paths that probably occur in engineering problems but never be seen in existing datasets. The data pool can be used to verify the effectiveness of the model in a wider strain space through active learning.

(2) Active learning tends to be sensitive to outliers in noisy datasets

Active learning can identify the most unreliable forecasts while these recognised datasets might be outliers. If active learning deals with data with high-level noises or fluctuations, the algorithm may constantly add outliers to training datasets which will certainly deteriorate the predictive capability of models. For example, the post-peak stress-strain stage experienced by a granular RVE may exhibit remarkable "stick-slip" behaviour with locally up-and-down stress values because of the evolving breakage and reconstruction of strong force chains in deformed granular materials. The same situations apply equally to laboratory experiment-based data. In contrast, the datasets from analytical formulations are almost continuous and are easier to be fitted. However, closed-form expressions can simply provide approximated stress-strain data for certain materials. Thus one has to consider the value of training a surrogate model with such imperfect data.

To tackle the noisy data and enable a wider application of active learning, more research needs to be explored in the future. On the one hand, some advanced feature extraction measures should be harnessed to denoise or 'wash' datasets. On the other hand, some customed active learning algorithms should be designed to automatically remove the measurable fluctuated points.

7. Concluding remarks

This study has developed a deep active learning-empowered data-driven constitutive modelling strategy, aiming to partially address two open challenges in the field: (1) use a small dataset to train a good predictive model, and (2) verify the reliability of a model without knowing ground truth. Three different application scenarios with RNNs and MLP-based DNN models are investigated in detail. The deep learning-based constitutive models of granular materials are also embedded into FEM to instantiate the unique advantage of active learning in recognising unreliable predictions for macro-scale simulations. The approach is conceptually simple but practically useful. In addition, the method is general and applicable to various data-centric surrogate models across science and engineering fields. The key findings are summarised below.

(1) Committee-based active learning requires only a few committee members provided that each committee member has learned sufficient knowledge. The greatest advantage of active learning lies in its ability to detect inaccurate predictions without knowing ground-truth, rather than guiding the surrogate model to select data at the very beginning of the training phase.
(2) The key to achieving interactive data-driven constitutive training, where training models and tagging data are dynamically performed, is that a pool of strain paths should be available beforehand. By using the proposed surrogate error indicator approach, some specific strain paths of certain proportional loadings can be artificially constructed.
(3) The active learning algorithm is a useful tool to automatically discover the underlying order of importance for a pool of data. Such insight enables to develop a reliable model with as small datasets as possible by only preparing the most informative specimen for the current surrogate model.
(4) Data-driven constitutive models can reproduce the simulation results of hierarchical multiscale modelling satisfactorily with much reduced computational costs. Active learning can detect potentially inaccurate predictions even in macroscale simulations, which is crucial to informing high-consequence decision-making of FEM computations with AI-based constitutive models. Particularly, these deep learning models can retrain themselves by adding those previously mis-predicted data and thus achieve self-learning or self-improvement via the active learning strategy.

CRediT authorship contribution statement

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

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.

Data availability

Data will be made available on request.

Acknowledgement

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

Appendix

Table A1
The second-round selected ten specimens with the greatest standard deviations.

Unloading strain	Reloading strain	Score	MAE
0.087	0.015	0.834	0.029
0.093	0.015	0.938	0.026
0.084	0.009	0.809	0.038
0.081	0.003	0.881	0.025
0.093	0.021	0.644	0.048
0.090	0.009	0.711	0.054
0.087	0.021	0.686	0.050
0.093	0.021	0.607	0.065
0.099	0.021	0.664	0.061
0.090	0.003	0.637	0.061

Table A2
The third-round selected ten specimens with the greatest standard deviations.

Unloading strain	Reloading strain	Score	MAE
0.030	0.003	0.809	0.034
0.024	0.003	0.869	0.030
0.117	0.003	0.856	0.032
0.117	0.009	0.848	0.033
0.117	0.018	0.647	0.054
0.117	0.024	0.672	0.048
0.036	0.003	0.792	0.041
0.111	0.003	0.735	0.047
0.117	0.030	0.786	0.036
0.111	0.009	0.856	0.033

Table A3
The fourth-round selected ten specimens with the greatest standard deviations.

Unloading strain	Reloading strain	Score	MAE
0.039	0.018	0.894	0.024
0.057	0.012	0.730	0.049
0.033	0.009	0.930	0.023
0.051	0.015	0.839	0.037
0.033	0.015	0.931	0.024
0.045	0.015	0.833	0.031
0.099	0.003	0.777	0.042
0.108	0.003	0.905	0.026
0.063	0.009	0.665	0.056
0.087	0.003	0.929	0.026

Table A4
The fifth-round selected ten specimens with the greatest standard deviations.

Unloading strain	Reloading strain	Score	MAE
0.108	0.006	1.0	0.018
0.114	0.009	0.957	0.021
0.114	0.078	0.968	0.029
0.102	0.003	0.993	0.020
0.036	0.015	1.0	0.018
0.057	0.009	0.987	0.034
0.048	0.012	0.996	0.029
0.024	0.015	1.0	0.019
0.051	0.012	1.0	0.031
0.018	0.015	0.951	0.026

Table A5
The sixth-round selected ten specimens with the greatest standard deviations.

Unloading strain	Reloading strain	Score	MAE
0.042	0.018	1.0	0.012
0.036	0.018	1.0	0.010
0.030	0.006	1.0	0.014
0.006	0.003	1.0	0.014
0.030	0.009	1.0	0.015
0.036	0.012	1.0	0.013
0.039	0.021	1.0	0.012
0.039	0.006	1.0	0.017
0.024	0.006	1.0	0.015
0.042	0.012	1.0	0.019

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.

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

Bonatti, C., Berisha, B., Mohr, D., 2022. From CP-FFT to CP-RNN: recurrent neural network surrogate model of crystal plasticity. Int. J. Plast. 158, 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.

Chollet, F., 2015. Keras: deep learning library for theano and tensorflow. URL: https://keras. io/k 7, T1.

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

Gao, Z., Zhao, J., 2017. A non-coaxial critical-state model for sand accounting for fabric anisotropy and fabric evolution. Int. J. Solids Struct. 106, 200–212.

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.

Gong, J., Nie, Z., Zhu, Y., Liang, Z., Wang, X., 2019. Exploring the effects of particle shape and content of fines on the shear behavior of sand-fines mixtures via the DEM. Comput. Geotech. 106, 161–176.

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

Guan, S., Qu, T., Feng, Y., Ma, G., Zhou, W., 2022. A machine learning-based multi-scale computational framework for granular materials. Acta Geotech. 1–22.

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

Guo, N., Zhao, J., 2016b. Parallel hierarchical multiscale modelling of hydro-mechanical problems for saturated granular soils. Comput. Methods Appl. Mech. Eng. 305, 37–61.

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.

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., 2018a. Multi-directional behavior of granular materials and its relation to incremental elasto-plasticity. Int. J. Solids Struct. 152, 305–323.

Kuhn, M.R., Daouadji, A., 2018b. Quasi-static incremental behavior of granular materials: elastic–plastic coupling and micro-scale dissipation. J. Mech. Phys. Solids 114, 219–237.

Lewis, D.D., Catlett, J., 1994. Heterogeneous uncertainty sampling for supervised learning. In: Proceedings of the Machine Learning. Elsevier, pp. 148–156 pp.

Li, X., Roth, C.C., Bonatti, C., Mohr, D., 2022. Counterexample-trained neural network model of rate and temperature dependent hardening with dynamic strain aging. Int. J. Plast. 151, 103218.

Lughofer, E., 2012. Hybrid active learning for reducing the annotation effort of operators in classification systems. Pattern Recognit. 45, 884–896.

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.
Peng, Y., Ding, X., Xiao, Y., Deng, X., Deng, W., 2020. Detailed amount of particle breakage in nonuniformly graded sands under one-dimensional compression. Can. Geotech. J. 57, 1239–1246.
Peng, Y., Yin, Z.Y., Ding, X., 2022. Analysis of particle corner-breakage effect on pile penetration in coral sand: model tests and DEM simulations. Can. Geotech. J. https://doi.org/10.1139/cgj-2022-0038.
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, M., 2021c. An adaptive granular representative volume element model with an evolutionary periodic boundary for hierarchical multiscale analysis. Int. J. Numer. Methods Eng. 122, 2239–2253.
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., Wang, M., Feng, Y., 2022. Applicability of discrete element method with spherical and clumped particles for constitutive study of granular materials. J. Rock Mech. Geotech. Eng. 14, 240–251.
Settles, B., 2009. Active learning literature survey.
Seung, H.S., Opper, M., Sompolinsky, H., 1992. Query by committee. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 287–294.
Shi, X., Zhao, J., Yin, J., Yu, Z., 2019. An elastoplastic model for gap-graded soils based on homogenization theory. Int. J. Solids Struct. 163, 1–14.
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., 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.
Wu, M., Xia, Z., Wang, J., 2022. Constitutive modelling of idealised granular materials using machine learning method. J. Rock Mech. Geotech. Eng. https://doi.org/10.1016/j.jrmge.2022.08.002.
Xi, X., Yin, Z., Yang, S., Li, C.Q., 2021. Using artificial neural network to predict the fracture properties of the interfacial transition zone of concrete at the meso-scale. Eng. Fract. Mech. 242, 107488.
Zhang, A., Dafalias, Y.F., Jiang, M., 2023. A bounding surface plasticity model for cemented sand under monotonic and cyclic loading. Géotechnique 73 (1), 44–61.
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., 2021. State-of-the-art review of machine learning applications in constitutive modeling of soils. Arch. Comput. Methods Eng. 28, 3661–3686.
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.
Zhao, J., Guo, N., 2015. The interplay between anisotropy and strain localisation in granular soils: a multiscale insight. Géotechnique 65, 642–656.
Zhao, J., Jiang, M., Soga, K., Luding, S., 2016. Micro origins for macro behavior in granular media. Granul. Matter 18, 1–5.
Zhao, S., Zhao, J., Liang, W., 2021. A thread-block-wise computational framework for large-scale hierarchical continuum-discrete modeling of granular media. Int. J. Numer. Methods Eng. 122, 579–608.