# A Novel Deep Learning-based Modelling Strategy from Image of Particles to Mechanical Properties for Granular Materials with CNN and BiLSTM

Pin ZHANG[1] and Zhen-Yu YIN[1,*]

1 Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

* Corresponding author, E-mail: zhenyu.yin@polyu.edu.hk; Tel: +852 3400 8470; Fax: +852 2334 6389

**Abstract:** It will be practically useful to know the mechanical properties of granular materials by only taking a photo of particles. This study attempts to deal with this challenge by developing a novel deep learning-based modelling strategy. In this strategy, the convolutional neural network (CNN) as image identification algorithm is first used to extract the particle information (particle size distribution PSD and morphology) based on the image of a granular sample, and the bidirectional long short-term memory (BiLSTM) neural network is employed to train the model of reproducing mechanical behaviours and induced fabric evolutions of the sample with corresponding particle information. The datasets of images of samples are generated using discrete element method, and the datasets of mechanical properties together with fabric evolutions are obtained through numerical tests on corresponding samples. As a preliminary attempt, two-dimensional biaxial samples and tests with initially isotropic fabric are considered for the sake of simplicity. The feasibility and reliability of the proposed modelling strategy are evaluated through training and testing. All results indicate that the first part of the model based on CNN is capable of accurately identifying PSD of a granular sample, as well as circularity and roundness of particles, using which as connecting parameters the mechanical behaviours together with induced fabric evolutions of granular materials are subsequently well captured by the second part of the model based on BiLSTM. This study provides a basis and a possible way to obtain immediately particle and packing information, mechanical properties and fabric evolutions by leveraging images of granular materials.

**Keywords:** Granular material; deep learning; discrete element method; particle morphology; particle size distribution; fabric anisotropy

1

# 1. Introduction

The physical and mechanical behaviours of granular materials are complex, because such materials are composed of distinct solid particles interacting with one another [1], which means that the behaviours of granular materials are primarily affected by their particle interactions particularly relating to the particle size distribution (PSD) [2-6] and the particle morphology [7-12]. Experimental tests tend to explain the behaviours of granular materials and investigate the effects of various factors from the macroscopic viewpoint [13-16], but the measurements are expensive using various advanced experimental techniques [17]. Analytical methods derived from continuum mechanics are generally based on strong assumptions with complicated mathematical formulations and myriad parameters which poses great difficulty for engineers in practice. If there is a way through recognizing individual particles and packing assembles to know the mechanical behaviour of the granular material, it will be genius and practically useful. The rapid development and application of deep learning (DL) nowadays bring a dawn for this.

Recently, the application of DL, e.g. convolutional neural network (CNN) and long short-term memory (LSTM), to simulate various properties of materials has gained great attention owing to its strong non-linear mapping capacity [18, 19]. CNN and its variants can directly extract microscopic information from images. Thereby, they have been applied to acquire properties from images of the studied material, such as the evaluation of permeability of porous media [20] and the prediction of fracture evolution of brittle material [21]. LSTM and its variants are characterized by sequence prediction, in which the historical information can be stored and learned. They have thus been successfully employed to simulate path-dependent behaviours of granular materials [22, 23] and demonstrated superior performance than other machine learning algorithms [24]. However, the current application of DL to identify the particle information of material such as the study of PSD and particle morphology is limited, and DL to model macroscopic

2

behaviours is typically only based on results of stress-strain without relating to particle and inter-particle information. Therefore, the feasibility of DL-based mechanical modelling with detecting particle information only upon images of granular materials deserves to be investigated. If do so, both experimental and computational costs can be significantly reduced for engineering practice.

For such purpose the datasets creation with both micro and macro information is necessary. The idealist method is to use Micro-Computed Tomography with performing mechanical tests on real granular materials [25-27]. Indeed as a preliminary attempt of the DL-based modelling strategy from image of particles of a sample to mechanical properties, two-dimensional (2D) discrete element method (DEM) well capturing granular mechanics in an analogous but simplified manner may also be a good choice. Note that DEM is a widely applied method to simulate behaviours of granular materials with providing physics insight for granular mechanics [1, 28-32]. Besides, the numerical tests eliminate experimental errors due to artificial operation or machine problems. DEM itself is computationally expensive on mechanical modelling [33], but this disadvantage can be eliminated by DL-based modelling if the datasets of DEM simulations are prepared in advance.

Therefore, in this study a novel DL-based strategy for recognising the particle information and behaviours of granular materials from image of particles is proposed. The results of DEM-based numerical biaxial tests with various particle size distributions and particle morphologies are converted into the auxiliary training and testing data. The proposed strategy includes two parts: CNN is first employed to identify the particle information including PSD and morphology, followed by BiLSTM to describe the mechanical behaviours and induced fabric evolutions of granular materials.

## 2. Generalization of applied methods

### 2.1 Indicators of particle and packing assembly

Myriad factors control the mechanical behaviour of granular materials, in which only the PSD and particle morphology, as two key fundamental factors, are selected for the sake of simplicity. PSD is computed based on the sizes of all individual particles, and the particle morphology is represented by its circularity and roundness in this study (Fig. 1). Circularity ($C$) describes the relative sharpness of corners and edges of a particle, and roundness ($R$) is a measure to describe the approximation degree of the overall outline of a particle to a circle. Methods proposed by Cox [34] and Wadell [35] (Eq. 1) are used to measure the circularity and roundness of a particle, respectively. The entire morphology of a sample is represented by its average circularity $C_a$ and average roundness $R_a$ (Eq. 2).

$$C = \frac{4\pi A}{P^2}; \quad R = \frac{1}{n_c}\sum_{r=1}^{n_c}\frac{D_r}{D_i} \tag{1}$$

$$C_a = \frac{1}{n_p}\sum_{i=1}^{n_p}C_i; \quad R_a = \frac{1}{n_p}\sum_{i=1}^{n_p}R_i \tag{2}$$

where $A$ and $P$ are the area and perimeter of a particle, respectively; $n_c$ is the number of corners of a particle; $D_r$ and $D_i$ are the diameters of curvature of a corner and the largest inscribed circle, respectively; $n_p$ is the number of particles for a given sample. For a circular particle, both $C$ and $R$ are equal to 1. For a triangular particle, $C$ and $R$ are identical to 0.605 and 0, respectively.

### 2.2 Discrete element method and fabric evolution

A series of 2D DEM-based biaxial test simulations on samples with different PSDs, morphologies are first carried out to create training and testing datasets. For a given combination of PSD and morphology, loose and dense samples are generated. All particles are generated within a plain bounded by rigid walls. Velocity-

4

controlled loading is assigned vertically and constant pressure is applied laterally. The particles in DEM are assumed to be rigid without deformation, and the dynamic modelling process is controlled by time step. The deformation of particles is simulated by overlapping, and the corresponding contact force is calculated using the contact model [29]. A linear elastic contact model with Coulomb's friction criterion is applied, in which the contact force $\boldsymbol{F}_c$ is resolved into linear $\boldsymbol{F}^l$ and dashpot force $\boldsymbol{F}^d$, and the contact moment is 0. Linear elastic and frictional inter-particle behaviours are considered. The motion of the particle is updated using Newton's second law. A typical set of variables used in DEM is summarized in Table 1.

By analysing the distribution of contacts $E(\theta)$ along with loading, which defines the portion of contacts falling within a given angular interval, fabric evolution can be inferred [36]. A second Fourier component provides a basis of approximating the distribution of contact normal orientation (Eq. 3).

$$E(\theta) = \frac{1}{2\pi}\left[1 + a\cos 2(\theta - \theta_a)\right] \tag{3}$$

where $a$ describes the magnitude of anisotropy in the contact normal orientation and $\theta_a$ describes the direction angle of anisotropy, which can be computed using:

$$\int_0^{2\pi} E(\theta)\cos 2\theta d\theta = (a/2)\cos 2\theta_a \; ; \; \int_0^{2\pi} E(\theta)\sin 2\theta d\theta = (a/2)\sin 2\theta_a \tag{4}$$

Particularly, for the fabrics of normal and tangential forces, $a$ can be replaced by $a_n$ and $a_s$, respectively, meanwhile $\theta_a$ can be replaced by $\theta_n$ and $\theta_s$, respectively. The analytical calculations of fabric anisotropy for normal force $f_n$ and tangential force $f_s$ can be derived from [36]:

$$f_n(\theta) = f_0\left[1 + a_n\cos 2(\theta - \theta_n)\right]; \; f_s(\theta) = -f_0 a_s\sin 2(\theta - \theta_s) \tag{5}$$

where $f_0$ is the average normal force. Noticed that Eq. 5 can also be used to describe the distribution of normal contact, in which $f_0$ denotes the average number of normal contacts.

**2.3 Convolutional neural network**

Inspired by the characteristics of CNN, two-dimensional CNN (con2d) is used to identify the particle information of granular materials. CNN generally consists of convolutional, pooling and fully-connected layers (Fig. 2). Given an image with the size of $h^r$ (height)$\times$ $w^r$ (width)$\times$ $c^r$ (channel, 1 for grayscale image and 3 for colour image), the size is rescaled to $h^c \times w^c \times c^c \times n^c$ after convolved by $n$ kernels [37]. The size of each kernel is identical with both height and width of $f$. The kernel scans the image from left to right and from top to bottom with a fixed stride $s$ to extract features. This process is achieved by computing the dot product between the entries of the kernel and the input image, and a 2-dimensional feature map is computed subsequently. Noticed that the kernel may exceed the boundary of an image when scanning the edge. Under that condition, the pixel value of the overflow part is assigned as 0. Therefore, the size of feature maps is:

$$S_i^c = \left\lceil \frac{S_i^r}{s} \right\rceil, \ i = h, w \tag{6}$$

where $S_i^r$ and $S_i^c$ denote the size of images along the $i$th dimension in the input and convolutional layers, respectively; $\lceil \ \rceil$ is the ceiling function, which maps $x$ to the least integer larger than or equal to $x$.

The pooling layer merges similar features in the feature map and further reduces the size of feature map to $h^p \times w^p \times c^p \times n^c$ using Eq. 6. The Max-pooling layer is used in this study, in which the maximum value in a region of feature map represented by a filter is retained. The feature map with two dimensions is flattened finally so that they can be fed into the fully-connected layer.

**2.4 Bidirectional long short-term memory neural network**

BiLSTM is adopted to capture the mechanical behaviours and induced fabric evolutions of granular materials in this study. Long short-term memory (LSTM) that can store and learn historical information has been proved to be suitable for modelling path-dependent behaviours of granular materials [22]. A memory

cell with an entity termed as "gate" is devised and embedded into the topology of LSTM to overcome the issues such as gradient vanishing and exploding [38] in conventional recurrent neural networks [39]. A series of variants motivated by the "gate" mechanism has been proposed and applied to specific issues, such as bidirectional LSTM (BiLSTM) [40] and gated recurrent unit (GRU) [41]. Herein, BiLSTM has the same architecture of memory cell with LSTM, but BiLSTM takes use of both positive and reverse sequential information, enhancing the interaction of sequential datasets. The number of weights and biases in BiLSTM increases two times in comparison with LSTM. Compared with general issues in the ML domains, the number of datasets and feature used in constitutive modelling of granular materials is much less, thereby the BiLSTM-based constitutive model can be controlled with a simple framework. The effect of the increasing number of weights and biases on computational efficiency is not discernible. To this end, BiLSTM is finally employed to simulate macroscopic behaviours of granular materials in this study (Fig. 3).

## 3. DL-based modelling strategy and generation of datasets

### 3.1 Data source

The dataset of biaxial tests used in this study are generated using DEM (download via: https://www.researchgate.net/publication/349532612_DEM_data_for_biaxial_simulations), since the simulation results are not interfered by experimental errors, which ensures to fairly evaluate the feasibility and performance of the DL-based modelling strategy from micro to macro scales. The height and width of the representative volume element (RVE) are 4 and 8 mm, respectively (Fig. 4). The RVE is initially compressed to an isotropic stress of 200 kPa, followed by the imposed velocity-controlled loading until the vertical strain reaches 15%. Meanwhile, the lateral stress maintains constant. The trajectory of mean stress $p$, deviatoric stress $q$, void ratio $e$ (Eq. 7) and fabric anisotropy is recorded. The detailed modelling cases

7

through DEM is summarized in Table 2.

Herein, the preparation of loose and dense DEM samples refers to Muir and Maeda [42], in which the maximum and minimum void ratios ($e_{max}$ and $e_{min}$) for a sample are first calculated, and the porosity $n$ with a given $D_r$ can be inferred subsequently using Eq. 8. The dense and loose samples are prepared using frictionless particles and particles with the friction coefficient of 0.5, respectively.

$$p = \frac{\sigma_{11} + \sigma_{22}}{2}; \; q = \sigma_{11} - \sigma_{22}; \; e = \frac{V_v}{V_s} \qquad (7)$$

$$n = \frac{e_{max} - D_r\left(e_{max} - e_{min}\right)}{1 + e_{max} - D_r\left(e_{max} - e_{min}\right)} \qquad (8)$$

where $\sigma_{11}$ and $\sigma_{22}$ are the vertical and lateral stress, respectively; $V_v$ and $V_s$ are the volume of void and solid assembly, respectively.

In total, 200 DEM-based biaxial tests are ultimately implemented, which sufficiently guarantees the stability of model performance in the presence of small data regime [43]. The images of all samples, PSD, particle morphology, stress-strain curves and induced fabric evolutions are collected to create a database. Note that the global behaviours from numerical biaxial tests are used for datasets, although there are shear bands inside of samples which may influence the sense of representative element of volume.

**3.2 Data preprocessing**

Data augmentation is employed to enlarge image database to improve the generalization ability of the CNN-based model [44]. The application of data augmentation techniques is required to retain the original information of samples. To this end, only flip data augmentation technique is used in this study, i.e., all original images are flipped horizontally and vertically, thereby the size of database enlarges three times with a total of 600 images. Each image is labelled by the PSD and sample type.

8

The raw simulated results of DEM fluctuate dramatically. Such variable datasets impose difficulties on BiLSTM to learn useful information. Therefore, piecewise cubic Hermitean interpolation polynomials method [45] and Savitzky-Golay denoising [46] methods are used to interpolate and smooth data, because such two methods can enlarge the database and smooth data without changing the shape of raw stress-strain curve (Fig. 4). Two databases are ultimately constructed, one is the image database including 600 images of samples, another includes the corresponding stress-strain data.

Regarding the image database, the original pixel values for an image range from 0 to 255. If the original images are directly fed to the CNN-based model, the values in the feature map would be extremely large, which may result in difficulty in the convergence of a training process. Therefore, the RGB values are rescaled into the range (0, 1) by multiplying 1/255 factor in this study [47]. Regarding the database of stress-strain relationships, to eliminate the effect of scale difference of input parameters on the training process of BiLSTM, all datasets are normalized into the range of (−1, 1) using the Min-Max scaling method.

Herein, 80% of images (480) are randomly selected to train the CNN-based model and the remaining 20% of images (120) are used to test the model. Because each image is flipped horizontally and vertically at the image augmentation phase, three images present the same sample. The testing set is thus constrained with 40 different samples for fairly evaluating the performance of the CNN-based model. Meanwhile, the corresponding 80% (160 cases) and 20% (40 cases) of strain-stress datasets are employed to train and test the BiLSTM-based model, respectively. It should be noted that the cases used for the training and testing sets in the CNN and BiLSTM-based models are the same. Therefore, for a single case in the testing set, the outputs of the CNN-based model can be directly fed to the BiLSTM-based model. The consistency of applying the proposed DL-based modelling framework is guaranteed in this way.

### 3.3 Framework of modelling strategy

#### *3.3.1 CNN-based particle identification model*

Fig. 5 presents the schematic view of the proposed DL-based modelling framework from particle to sample

scales. The modelling starts from identifying PSD and morphology of samples using CNN. By importing

the image of a sample into the CNN-based model, and the corresponding PSD and morphology can be

predicted. Accordingly, two CNN-based models are developed, one for PSD identification and another for

particle morphology.

Motivated by Buscombe [48], to obtain the PSD of a sample, CNN is designed to identify the diameter

through which 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100% of particles pass.

The dimension of the output in the CNN-based model is thus identical to 12. To identify particle

morphology, the direct intention is to identify the ratio of each type of particles, thereby the aforementioned

$C_a$ and $R_a$ can be calculated. However, this study merely considers round and triangular particles, and five

types of samples with different mixture ratios are prescribed. Therefore, the identification of particle

morphology can be converted into a classification issue, i.e., the identification of sample type. The

mathematic expression for computing PSD and morphology via CNN is given in the Appendix A1.

#### *3.3.2 BiLSTM-based mechanical prediction model*

The second phase of the proposed framework is to develop a BiLSTM-based model for predicting

mechanical behaviours and induced fabric evolutions of granular materials. The selection of features

determines what factors can be considered by the date-driven model, thereby it is related to the application

scopes of the BiLSTM-based model [49]. To this end, features are required to involve inherent properties

of the studied object, state parameters and history information of stress or strain, presented as follows:

- The inherent properties of granular materials are identified by the CNN-based model at the first phase. Herein, the mean particle size $d_{50}$ and the coefficient of uniformity $C_u$ derived from the PSD as representative indicators are selected as the input variables. The inherent properties of a sample are thus represented by four variables, i.e. $d_{50}$, $C_u$, $C_a$ and $R_a$.

- State variables are employed to reproduce the experimental process including initial void ratio $e_0$, relative density $D_r$ and vertical strain $\varepsilon$.

- The outputs of the BiLSTM-based model at previous time steps are fed to the input variables for considering history loading path.

Hence, the input parameters of the BiLSTM-based model for modelling mechanical behaviours are $x$ = $[p, q, e, \varepsilon, e_0, D_r, d_{50}, C_u, C_a, R_a]$, and $x = [a, a_n, a_s, \varepsilon, e_0, D_r, d_{50}, C_u, C_a, R_a]$ for predicting fabric evolutions. The direction angle of fabric anisotropy roughly maintains constant in the biaxial test, thereby anisotropy variables regarding the magnitude of anisotropy $a$, $a_n$, $a_s$ are predicted while the direction angle of anisotropy maintains initial values. The detailed mathematic expression of a BiLSTM-based model is introduced in the Appendix A2.

## 4. Applications

### 4.1 Training of the CNN-based model

The development of the DL-based model is implemented using Keras with the Tensorflow backend [50]. The training of CNN is first to determine the number of convolutional, pooling, fully-connected layers, and the hidden neurons in each layer. Other hyper-parameters such as optimizer, learning rate, activation function, batch size and epochs are required to be finely tuned. The prediction of PSD is a regression issue, mean square error (MSE) is thus set as the loss function. Meanwhile, *k*-fold cross-validation method is

employed to detect the overfitting issue. The loss function is finally formulated using:

$$\text{MSE} = \frac{1}{k}\frac{1}{n_{sub}}\sum_{i=1}^{n_{sub}}\left(y_i^p - y_i^a\right)^2 \tag{9}$$

where $n_{sub}$ is the number of datasets in a sub validation set; $k$ is the number of folds, which is usually set as

10; $y_i^p$ and $y_i^a$ are predicted and actual values, respectively. The configurations of the CNN-based model

are determined using the trial-and-error method, and the detailed process is not presented for brevity. The

final configurations are summarized in Table 3. The number of weights and biases are 6193176 and 92,

respectively. It can be seen from Fig. 6a that MSE values on the training and testing sets are small and

rapidly converge within 50 epochs.

The shallow layers of the CNN-based model focus on extracting features of each sample, while the

deep layers are used to fuse features and the outputs tend to be abstract. To reveal the mechanism of the

CNN-based model to identify the PSD, the outputs at the shallow layers are illustrated. Feature maps in the

fundamental four layers of the model are illustrated in Fig. 7. When a particle is convolved by a kernel, the

output is dependent on its size. After three kernels convolve the whole image (con2d_1, con2d_2, con2d_3)

and pass through the pooling layer (max_pooling2d_1), particle size information is stored in the feature

map with different values and are thus marked with different colours. Accordingly, the CNN-based model

can distinguish the particle size and predict PSD.

The identification of particle morphology is a classification issue, thereby the average cross-entropy

(ACE) is tailored for objectively calculating the discrimination between predicted and actual labels [51].

$$\text{ACE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C}y_{j,i}^a \ln f_j\left(x_i\right) \tag{10}$$

where $C$ is the number of classification labels; $n$ is the total number of datasets; $y_{j,i}^a$ is the actual label; $f_j(x_i)$

is the output of the CNN-based model for an input $x_i$. The evolution of ACE values on the training and testing sets is presented in Fig. 6b. The number of weights and biases are 6190072 and 61, respectively. The training process rapidly maintains stable, and the ACE value accordingly decreases to 0 within 50 epochs. Similar to the identification of particle size, the pixel size and value for a triangular particle and a circular one are different, thereby CNN can extract the particle morphology information. The process of extracting features in the CNN-based particle morphology identification model is the same as the principle revealed in Fig. 7, the feature map in each layer is thus not presented for brevity.

**4.2 Training of the BiLSTM-based model**

Similar to CNN, the training of the BiLSTM-based model is primarily to determine the number of hidden layers and hidden neurons in each layer, and other hyper-parameters are required to be finely tuned. The configurations of the BiLSTM-based model are determined using the trial-and-error method, and the results are presented in Table 3. The number of weights and biases are 67560 and 963, respectively.

One of the important configurations in BiLSTM is the time step, which determines how long the stress-strain history can be taken into account by the BiLSTM-based model. The history information enforces the capacity of modelling mechanical behaviours, but too long history information also degrades the learning efficiency and causes overfitting. To this end, three steps of stress-strain history before the current step are selected, i.e., the time step of BiLSTM is set as 3. Because the number of datasets at each experiment is identical to 1503, the batch size is thus set as 1503 to ensure BiLSTM can learn the entire information of an entire experiment at each round. Considering the prediction of mechanical behaviours is a regression issue, the MSE is also set as the loss function. The evolution of MSE values generated by the BiLSTM-based models for predicting mechanical responses and induced fabric evolutions during the training process is presented in Figs. 6c and 6d, respectively. The MSE values on both training and testing sets dramatically

decrease to a negligible value within 50 epochs. Compared with the evolution of MSE values generated by

the CNN-based PSD identification model (Fig. 6a), the MSE values generated by the BiLSTM-based

models are more fluctuant, because the ranges of outputs in the BiLSTM-based model is much larger than

the range of outputs (i.e. diameters of particles) in the CNN-based PSD identification model. Overall, the

MSE values are small, thereby the performance of training can be guaranteed.

**4.3 Modelling results of the CNN-based model**

After the optimum configurations of the CNN-based model are determined, the training and testing sets are

fed to the model to examine its generalization ability. To quantitatively evaluate the performance of the

BiLSTM-based model. Absolute and relative error indicators, i.e., mean absolute error (MAE) and mean

absolute percentage error (MAPE), are computed.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^p - y_i^a\right|; \quad \mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^p - y_i^a}{y_i^a}\right| \times 100\% \tag{11}$$

The MAE and MAPE values generated by the CNN-based model for predicting PSD is summarized

in Table 4. The error is not discernible and MAPE value is largely owning to the small value of the

denominator. To comprehensively evaluate model performance, the predicted results of 5 samples in the

testing set with different mixture ratio and PSD are particularly selected. It can be seen from Fig. 8 that the

predicted PSD curves are close to the evolution of actual points, and the calculated $d_{50}$ and $C_u$ are roughly

identical to the actual results. Such factors indicate CNN deduces plausible identification for the PSD of

granular materials.

The result of identifying particle morphology on the testing set is presented in Fig. 9. The confusion

matrix is used to evaluate the classification accuracy of the CNN-based particle morphology identification

model. The sum of probability at each column is identical to 0. Given an actual mixture ratio, the predicted

mixture ratio has five conditions, but only one condition that is located at the diagonal line is correct. It can

be observed that the CNN-based model perfectly identifies the mixture ratio with the probability along the

diagonal line of 1. CNN provides a direct and reliable way of identifying the particle information of granular

materials.

**4.4 Modelling results of the BiLSTM-based model**

The number of datasets used in the BiLSTM-based model is large. Each simulation test consists of 1503

datasets, and the training and testing sets include 160 and 40 numerical simulations, respectively. Although

MAE and MAPE values on the testing set are roughly 5 times the values on the training set, the prediction

errors are negligible (Table 4). To reveal the modelling results of the BiLSTM-based model, the mechanical

behaviours of 5 samples in the testing set (the same as the selected samples in the former section) are

illustrated (Fig. 10). The BiLSTM-based model can accurately capture the relationships of $p–\varepsilon$, $q–\varepsilon$ and $e–$

$\varepsilon$. Cho et al. [52] demonstrated that the increasing angularity leads to the increasing strength of soil mass,

extreme $e_{\max}$ and $e_{\min}$. The simulated results in Fig. 10 also present a similar trend, in which the increasing

ratio of triangular particles results in higher peak stress and lower initial void ratio. The BiLSTM-based

model captures the effects of mixture ratio of on the mechanical behaviours. For a biaxial test simulated

using DEM, a 2-dimensional granular specimen dilates at a greater rate than the real specimen particularly

for angular particle [53, 54]. Moreover, the increasing triangular particles cause denser sample with a

smaller initial void ratio, and the volumetric dilatancy is more obvious due to the rolling of triangular

particles during the shearing process. Therefore, it can be seen from Fig. 10 that the volumetric compression

on the loose sample merely appears as the ratio of triangular particles is less than 75%. Volumetric dilatancy

is observed on both dense and loose samples in most of cases. Mechanical behaviours including stress

softening and volumetric dilatancy on the dense samples, stress hardening and volumetric compression on

the loose samples with a large ratio of circular particles can accurately be captured by the BiLSTM-based

model. Such factors indicate BiLSTM understands the intrinsic physical mechanism of granular materials,

showing great potential of modelling their mechanical behaviours.

The prediction errors on the induced fabric evolutions are also small (Table 4). MAE and MAPE values

on the testing set double the values on the training set. Fig. 11 presents the evolution of predicted and

measured $a$, $a_n$ and $a_s$, which reveals the magnitude of anisotropy of contact number, contact normal force

and tangential force, respectively. The BiLSTM-based model is capable of accurately predicting the

evolution of induced anisotropy. With the increasing ratio of triangular particles, the magnitude of

anisotropy increases. Detailed inspection of the modelling results indicates that the peak values of $a$, $a_n$ and

$a_s$ reach at $\varepsilon$ of around 2%, which is close to the position where peak strength reaches (Fig. 10). Thereafter

the values of $a$, $a_n$ and $a_s$ reduces for the dense samples, and maintains steadily for the loose samples. The

ultimate values of $a$, $a_n$ and $a_s$ of dense and loose samples with the same mixture ratio are roughly identical.

Such results reflect the variation of the microstructure of granular samples, and it further induces the

corresponding responses of global mechanical behaviours. Overall, BiLSTM is capable of identifying the

physical phenomena on both microscopic and macroscopic scales.

## 5. Evaluation of generalization ability and robustness

Understanding the performance of DL-based model on unknown data is vitally important to guarantee its

application scope. Four additional numerical biaxial tests with two mixture ratios (R10_T90 and R90_T10)

that out of the range of the training set are conducted using DEM, in which the $d_{50}$ and $C_u$ of the R10_T90

sample are 0.0006 and 3.5, respectively. The $d_{50}$ and $C_u$ of the R90_T10 sample are 0.0007 and 3.75,

respectively. Both dense and loose samples for each mixture ratio are prepared for testing set. For testing,

the images of the four new samples as inputs are fed to the developed CNN-based model to predict the PSD,

16

i.e., $d_{50}$ and $C_u$. They are then automatically fed to the developed BiLSTM-based model with pre-assigned

morphology information to predict the mechanical behavours of these new samples. The predicted results

are compared with simulated results of DEM to evaluate the generalization ability and robustness of the

proposed DL-based modelling framework. Good agreement demonstrates that the DL-based model can still

accurately capture the evolution of mechanical responses and induced fabric evolution for samples with

novel particle morphology. Based on the predicted $a$, $a_n$ and $a_s$, the distribution of contact normal can be

fitted using Eq. 5. The results for a dense sample R10_T90 as a representative example are illustrated in

Fig. 13. It can be observed that the fitted contact normal distribution, distributions of normal and tangential

forces show excellent agreement with the measured results. The evolution of anisotropy with the increasing

vertical strain can be accurately captured. These factors indicate the proposed DL-based modelling

framework is promising for multiscale modelling. The DL-based model is capable of accurately capturing

the effects of particle information on the mechanical behavours and induced anisotropy, such as effects of

particle morphology and relative density on the characteristics of deformation and strength with the

magnitude of anisotropy.

## 6. Conclusions

A DL-based modelling strategy from image of particles to global behaviours of the sample of granular

materials has been developed. In this approach, CNN was first used to extract the particle information (PSD

and morphology), and BiLSTM neural network was then employed to simulate the global mechanical

behaviours and fabric evolutions of granular materials. The datasets of biaxial samples and tests were

created using DEM with different particle information for training and testing.

As demonstrated, the proposed CNN-based model part is capable of accurately identifying mean

17

diameter and coefficient of uniformity of a granular sample, as well as circularity and roundness of a particle, which provides a direct and easy way of acquiring the particle information of granular materials. The effects of PSD, particle morphology and relative density on the global mechanical behaviours and induced fabric evolutions of granular materials have been well captured by the BiLSTM-based model with the particle information as connecting parameters. Thus, the feasibility and reliability of the proposed modelling strategy have been evaluated.

It should be pointed out once two DL-based model parts embedded in the proposed modelling framework are well trained, the whole modelling process from extraction of particle information to prediction of mechanical and fabric behaviours can be completed within seconds. The computational source is dramatically saved. Moreover, the proposed modelling framework is genetic, which means the datasets can be replaced by the real experimental or brilliant numerical datasets to investigate different issues, such as Micro-CT for real granular materials on real granular materials, 3D numerical modelling and making use of contact distribution to consider the fabric of materials. Meanwhile, the applied DL algorithms can also be updated by more advanced and effective algorithms considering the rapid development in the DL domain.

# Appendix

**A1 - Mathematic expression of the CNN-based model**

Given a set of images of a soil sample that are digitalized by $\mathbf{x}$, the computation of the CNN-based model with two convolutional layers, one pooling layer and one fully-connected layer as a representative example is carried out.

(1) from the input $\mathbf{x}$ to the first convolutional $\mathbf{C}^1$ layers:

$$\mathbf{C}_j^1 = \mathrm{ReLU}\left[ sum\left(\mathbf{W}_{C_j^1} \odot \mathbf{x}_i\right) + b_j \right] \tag{A1}$$

where $\mathbf{C}_j^1$ denotes the outputs convolved by the $j$th kernel in the $C^1$ layer; $\mathbf{W}_{C_j^1}$ is weights of the $j$th kernel in the $C^1$ layer; $b_j$ is a bias term at the $j$th kernel; $sum$ denotes the summation of all elements; ReLU represents the activation function rectified linear unit (ReLU).

(2) from the first $\mathbf{C}^1$ to the second $\mathbf{C}^2$ convolutional layers:

$$\mathbf{C}_j^2 = \mathrm{ReLU}\left[ sum\left(\mathbf{W}_{C_j^2} \odot \mathbf{C}_i^1\right) + b_j \right] \tag{A2}$$

where the meaning of all notations can refer to Eq. A1.

(3) from the second convolutional $\mathbf{C}^2$ to the pooling $\mathbf{P}^1$ layers:

$$\mathbf{P}_{i,j}^1 = \max\left(\mathbf{C}_{i,j}^2, \ldots, \mathbf{C}_{i,j+f_{p^1}-1}^2, \ldots, \mathbf{C}_{i+f_{p^1}-1,j}^2, \ldots, \mathbf{C}_{i+f_{p^1}-1,j+f_{p^1}-1}^2\right) \tag{A3}$$

where max function indicates the output at the $\mathbf{P}^1$ layer is the maximum value of a region covered by a kernel with a size of $f_{p^1}$.

(4) from the pooling $\mathbf{P}^1$ to the fully-connected $\mathbf{F}^1$ layers:

$$\mathbf{F}^1 = \mathrm{ReLU}\left[ \mathbf{W}_{F^1} \times \left(\mathbf{P}_1^1, \ldots, \mathbf{P}_2^1, \ldots, \mathbf{P}_{n_{p^1}}^1\right) + \boldsymbol{b}_{F^1} \right] \tag{A4}$$

where $n_{p^1}$ denotes a total of elements in the $\mathbf{P}^1$ layer; $\mathbf{W}_{F^1}$ and $\boldsymbol{b}_{F^1}$ are the weights and biases connecting $\mathbf{P}^1$ and $\mathbf{F}^1$ layers, respectively.

(6) from the fully-connected $\mathbf{F}^1$ to the output $\boldsymbol{y}$ layers:

$$\boldsymbol{y}_{PSD} = \mathbf{W}^y \times \mathbf{F}^n + \boldsymbol{b}^y; \quad \boldsymbol{y}_{mor} = e^{\mathbf{W}^y \times \mathbf{F}^n + \boldsymbol{b}^y} \Big/ \sum e^{\mathbf{W}^y \times \mathbf{F}^n + \boldsymbol{b}^y} \tag{A5}$$

where $\mathbf{W}^y$ and $\boldsymbol{b}^y$ are the weights and biases connected $\mathbf{F}^1$ and $\boldsymbol{y}$ layers, respectively. Eqs. The computation of regression and classification problems is different, as shown in A5, where the outputs of PSD ($\boldsymbol{y}_{PSD}$) and

19

particle morphology ($\boldsymbol{y}_{mor}$) identification models are given.

**A2 - Mathematic expression of the BiLSTM-based model**

In BiLSTM (Fig. 3), $^{p}\mathbf{W}$, $^{p}\mathbf{U}$ and $^{p}\boldsymbol{b}$ are the weights matrices and biases vector used for the positive data flow. $^{r}\mathbf{W}$, $^{r}\mathbf{U}$ and $^{r}\boldsymbol{b}$ are used for the reverse data flow. The same architecture of the memory cell is used for positive and reverse directions, which means the dimensions of matrices and biases used in both positive and reverse data flow is identical, but the values are different. The mathematical expression of the BiLSTM-based model with one LSTM layer, one fully-connected layer and a time step of $t$ as a representative example is revealed as follows:

(1) From the first input $\boldsymbol{x}^{1}$ to the LSTM layers at the first time step $^{p}\boldsymbol{h}^{1}$ (positive data flow), it starts from the calculation of the forget, input and output gates:

$$^{p}\boldsymbol{f}^{1} = \sigma\left( ^{p}\mathbf{W}_{f} \times \boldsymbol{x}^{1} + ^{p}\mathbf{U}_{f} \times ^{p}\boldsymbol{h}^{0} + ^{p}\boldsymbol{b}_{f} \right)$$

$$^{p}\boldsymbol{i}^{1} = \sigma\left( ^{p}\mathbf{W}_{i} \times \boldsymbol{x}^{1} + ^{p}\mathbf{U}_{i} \times ^{p}\boldsymbol{h}^{0} + ^{p}\boldsymbol{b}_{i} \right)$$

$$^{p}\boldsymbol{o}^{1} = \sigma\left( ^{p}\mathbf{W}_{o} \times \boldsymbol{x}^{1} + ^{p}\mathbf{U}_{o} \times ^{p}\boldsymbol{h}^{0} + ^{p}\boldsymbol{b}_{o} \right) \tag{A6}$$

where subscript $f$, $i$ and $o$ denote the weight matrices and bias vectors used in the forget, input and output gates, respectively. $\sigma$ is the activation function *sigmoid*. Superscript $p$ denotes the positive data flow. $^{p}\boldsymbol{h}^{0}$ with zero values denotes the initial hidden state. The output of memory cell is obtained by:

$$^{p}\tilde{\boldsymbol{c}}^{1} = \mathrm{ELU}\left( ^{p}\mathbf{W}_{c} \times \boldsymbol{x}^{1} + ^{p}\mathbf{U}_{c} \times ^{p}\boldsymbol{h}^{0} + ^{p}\boldsymbol{b}_{c} \right); \quad ^{p}\boldsymbol{c}^{1} = {}^{p}\boldsymbol{f}^{1} \odot {}^{p}\boldsymbol{c}^{0} + {}^{p}\boldsymbol{i}^{1} \odot {}^{p}\tilde{\boldsymbol{c}}^{1}; \quad ^{p}\boldsymbol{h}^{1} = {}^{p}\boldsymbol{o}^{1} \odot \mathrm{ELU}\left( ^{p}\boldsymbol{c}^{1} \right) \tag{A7}$$

where $\odot$ denotes the element-wise product. $^{p}c^{0}$ with zero values denotes the initial memory state. ELU is the activation function exponential linear unit.

(2) from the first $^{p}\boldsymbol{h}^{1}$ to the next time step $^{p}\boldsymbol{h}^{2}$, $^{p}\boldsymbol{h}^{3}$, …, $^{p}\boldsymbol{h}^{t}$ in the LSTM layers: the calculation is similar to

20

Eqs. A6-A7, the only difference is the $x^1$ and $^p h^0$ are replaced by $x^2$ and $^p h^1$, respectively. The subsequent

calculation steps are similarly repeated until the $^p h^t$ is obtained.

(3) the ultimate output of the first LSTM layer $h^t$: procedure (1) and (2) reveal the positive data flow, and

half of the first hidden state. The calculation method of reverse data flow is same, in which the only

difference is that the calculation starts from the $t$th time, and $^r\mathbf{W}$, $^r\mathbf{U}$ and $^r b$ are used instead of $^p\mathbf{W}$, $^p\mathbf{U}$ and

$^p b$. Therefore, the ultimate output of the first hidden layer is obtained by integrating the outputs of memory

cell in both positive and reverse data flow using:

$$h^1 = {}^p h^1 \oplus {}^r h^t \qquad (A8)$$

where $\oplus$ denotes concatenation operation.

(4) from the LSTM $h^1$ to the fully-connected layers $F$:

$$F = \tanh\left(\mathbf{W}_F \times h^1 + b_F\right) \qquad (A9)$$

where $\mathbf{W}_F$ and $b_F$ are the weights and biases used in the fully-connected layer, respectively.

(5) from the fully-connected to the output layers:

$$y = \mathbf{W}_y \times F + b_y \qquad (A10)$$

where $\mathbf{W}_y$ and $b_y$ are the weights and biases used in the output layer, respectively. $y$ is the predicted

mechanical responses or fabric evolutions based on the input $x$.

21

# References

[1] S. Zhao, T.M. Evans, X. Zhou, Effects of curvature-related DEM contact model on the macro- and micro-mechanical behaviours of granular soils, Géotechnique, 68 (2018) 1085-1098.

[2] Y.F. Jin, Z.Y. Yin, Z.X. Wu, A. Daouadji, Numerical modeling of pile penetration in silica sands considering the effect of grain breakage, Finite Elem. Anal. Des., 144 (2018) 15-29.

[3] Z.Y. Yin, P.-Y. Hicher, C. Dano, Y.F. Jin, Modeling mechanical behavior of very coarse granular materials, J. Eng. Mech., 143 (2017) C4016006.

[4] Z.Y. Yin, J. Zhao, P.Y. Hicher, A micromechanics-based model for sand-silt mixtures, Int. J. Solids Struct., 51 (2014) 1350-1363.

[5] Y.J. Liu, G. Li, Z.Y. Yin, Christophe Dano, P.Y. Hicher, J.H. Wang, X.H. Xia, Influence of grading on undrained behavior of granular materials, C. R. Mec., 342 (2014) 85-95.

[6] J. Yang, Z.Y. Yin, F. Laouafa, P.-Y. Hicher, Analysis of suffusion in cohesionless soils with randomly distributed porosity and fines content, Comput. Geotech., 111 (2019) 157-171.

[7] H.M. Beakawi Al-Hashemi, O.S. Baghabra Al-Amoudi, A review on the angle of repose of granular materials, Powder Technol., 330 (2018) 397-417.

[8] Y.T. Feng, T. Zhao, J. Kato, W. Zhou, Towards stochastic discrete element modelling of spherical particles with surface roughness: A normal interaction law, Comput. Meth. Appl. Mech. Eng., 315 (2017) 247-272.

[9] Z. Lai, Q. Chen, L. Huang, Fourier series-based discrete element method for computational mechanics of irregular-shaped particles, Comput. Meth. Appl. Mech. Eng., 362 (2020) 112873.

[10] K. Shibata, S. Koshizuka, I. Masaie, Cost reduction of particle simulations by an ellipsoidal particle model, Comput. Meth. Appl. Mech. Eng., 307 (2016) 411-450.

[11] K. Kildashti, K. Dong, B. Samali, An accurate geometric contact force model for super-quadric particles, Comput. Meth. Appl. Mech. Eng., 360 (2020) 112774.

[12] G. Mollon, J. Zhao, 3D generation of realistic granular samples based on random fields theory and Fourier shape descriptors, Comput. Meth. Appl. Mech. Eng., 279 (2014) 46-65.

[13] L.M. Wei, J. Yang, On the role of grain shape in static liquefaction of sand–fines mixtures, Géotechnique, 64 (2014) 740-745.

[14] F. Ezzein, R. Bathurst, A transparent sand for geotechnical laboratory, Geotech. Test. J., 34 (2011) 590-601.

[15] J. Chu, An experimental examination of the critical state and other similar concepts for granular soils, Can. Geotech. J., 32 (1995) 1065-1075.

[16] H.L. Wang, Y.J. Cui, F. Lamas Lopez, N. Calon, G. Saussine, J.C. Dupla, J. Canou, P. Aimedieu, R.P. Chen, Investigation on the mechanical behavior of track-bed materials at various contents of coarse grains, Constr. Build. Mater., 164 (2018) 228-237.

[17] T. Ueda, T. Matsushima, Y. Yamada, DEM simulation on the one-dimensional compression behavior of various shaped crushable granular materials, Granul. Matter, 15 (2013) 675-684.

[18] P. Zhang, Z.Y. Yin, Y.F. Jin, T. Chan, F.P. Gao, Intelligent modelling of clay compressibility using hybrid meta-heuristic and machine learning algorithms, Geosci. Front., 12 (2021) 441-452.

[19] P. Zhang, Z.Y. Yin, Y.F. Jin, T.H.T. Chan, A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest, Eng. Geol., 265 (2020) 105328.

[20] J. Tian, C. Qi, Y. Sun, Z.M. Yaseen, Surrogate permeability modelling of low-permeable rocks using convolutional neural networks, Comput. Meth. Appl. Mech. Eng., 366 (2020) 113103.

[21] M. Schwarzer, B. Rogan, Y. Ruan, Z. Song, D.Y. Lee, A.G. Percus, V.T. Chau, B.A. Moore, E. Rougier, H.S. Viswanathan, G. Srinivasan, Learning to fail: Predicting fracture evolution in brittle material models using recurrent graph convolutional neural networks, Comput. Mater. Sci., 162 (2019) 322-332.

[22] P. Zhang, Z.Y. Yin, Y.F. Jin, G.L. Ye, An AI-based model for describing cyclic characteristics of granular materials, Int. J. Numer. Anal. Methods Geomech., 44 (2020) 1315-1335.

[23] K. Wang, W. Sun, Meta-modeling game for deriving theory-consistent, microstructure-based traction–separation laws via deep reinforcement learning, Comput. Meth. Appl. Mech. Eng., 346 (2019) 216-241.

[24] P. Zhang, Z.Y. Yin, Y.F. Jin, State-of-the-art review of machine learning applications in constitutive modeling of soils, Arch. Comput. Method Eng., (2021).

[25] J. Fonseca, C. O'Sullivan, M.R. Coop, P.D. Lee, Quantifying the evolution of soil fabric during shearing using scalar parameters, Géotechnique, 63 (2013) 818-829.

[26] J. Fonseca, W.W. Sim, T. Shire, C. O'Sullivan, Microstructural analysis of sands with varying degrees of internal stability, Géotechnique, 64 (2014) 405-411.

[27] S. Nadimi, J. Fonseca, A micro finite-element model for soil behaviour, Géotechnique, 68 (2018) 290-302.

[28] T. Zhao, F. Dai, N.W. Xu, Y. Liu, Y. Xu, A composite particle model for non-spherical particles in DEM simulations, Granul. Matter, 17 (2015) 763-774.

[29] P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies, Géotechnique, 29 (1979) 47-65.

[30] N.T. Ngo, B. Indraratna, C. Rujikiatkamjorn, Micromechanics-based investigation of fouled ballast using large-scale triaxial tests and discrete element modeling, J. Geotech. Geoenviron. Eng., 143 (2017) 04016089.

[31] M.O. Ciantia, M. Arroyo, F. Calvetti, A. Gens, An approach to enhance efficiency of DEM modelling of soils with crushable grains, Géotechnique, 65 (2015) 91-110.

[32] J.M. Harmon, D. Arthur, J.E. Andrade, Level set splitting in DEM for modeling breakage mechanics, Comput. Meth. Appl. Mech. Eng., 365 (2020) 112961.

[33] D. Reimann, K. Nidadavolu, H. ul Hassan, N. Vajragupta, T. Glasmachers, P. Junker, A. Hartmaier, Modeling macroscopic material behavior with machine learning algorithms trained by micromechanical simulations, Front. Mater., 6 (2019) 1-19.

[34] E.A. Cox, A method for assigning numerical and percentage values to the degree of roundness of sand grains, J. Paleontol., 1 (1927) 179-183.

[35] H. Wadell, Volume, shape, and roundness of rock particles, J. Paleontol., 40 (1932) 443-451.

[36] L. Rothenburg, R.J. Bathurst, Analytical study of induced anisotropy granular materials, Géotechnique, 39 (1989) 601-614.

[37] Z. Zhong, T.R. Carr, X.M. Wu, G. Wang, Application of a convolutional neural network in permeability prediction: A case study in the Jacksonburg-Stringtown oil field, West Virginia, USA, Geophysics, 84 (2019) 363-373.

[38] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature, 521 (2015) 436-444.

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

[40] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw., 18 (2005) 602-610.

[41] K. Cho, B. Van Merriënboer, C. Gulcehre, v. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning

phrase representations using RNN Encoder–Decoder for statistical machine translation, arXiv:1406.1078v3, (2014).

[42] W. Muir, David, K. Maeda, Changing grading of soil: effect on critical states, Acta Geotech., 3 (2007) 3-14.

[43] C. Grigo, P.-S. Koutsourelakis, A physics-aware, probabilistic machine learning framework for coarse-graining high-dimensional systems in the Small Data regime, J. Comput. Phys., 397 (2019) 108842.

[44] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, in: NIPS, 2012.

[45] C.B. Moler, Numerical computing with MATLAB, SIAM, 2004.

[46] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem., 36 (1964) 1627-1639.

[47] S. Joshi, S. Saxena, N. Khanna, First steps toward CNN based source classification of document images shared over messaging app, Signal Process. Image Commun., 78 (2019) 32-41.

[48] D. Buscombe, SediNet: a configurable deep learning model for mixed qualitative and quantitative optical granulometry, Earth Surf. Proc. Land., 45 (2020) 638-651.

[49] P. Zhang, R.P. Chen, H.N. Wu, Real-time analysis and regulation of EPB shield steering using Random Forest, Automat. Constr., 106 (2019) 102860.

[50] F. Chollet, Keras, https://github.com/keras-team/keras, (2015).

[51] J. Ye, Single valued neutrosophic cross-entropy for multicriteria decision making problems, Appl. Math. Model., 38 (2014) 1170-1175.

[52] G.C. Cho, J. Dodds, J.C. Santamarina, Particle shape effects on packing density, stiffness, and strength: natural and crushed sands, J. Geotech. Geoenviron. Eng., 132 (2006) 591-602.

[53] Rothenburg L, B.R. J., Micromechanical features of granular assemblies with planar elliptical particles, Géotechnique, 42 (1992) 79-95.

[54] X.Y. Jing, W.H. Zhou, H.X. Zhu, Z.Y. Yin, Y.M. Li, Analysis of soil-structural interface behavior using three-dimensional DEM simulations, Int. J. Numer. Anal. Methods Geomech., 42 (2018) 339-357.

# Table

**Table 1** Parameters used in DEM modelling

| Parameter | Value |
| --- | --- |
| Particle density (kg/m$^3$) | 2650 |
| Inter-particle friction coefficient | 0.5 |
| Friction coefficient of particle-wall | 0 |
| Inter-particle contact normal stiffness (N/m) | $1.5 \times 10^8$ |
| Inter-particle contact shear stiffness (N/m) | $1.0 \times 10^8$ |
| Normal stiffness of particle-wall (N/m) | $15 \times 10^8$ |
| Shear stiffness of particle-wall (N/m) | $10 \times 10^8$ |
| Damping ratio | 0.7 |

**Table 2** Summary of numerical sample types for the database

| | |
| --- | --- |
| 5 Mixture ratios | R0_T100 (100% triangle particles), |
| | R25_T75 (25% round and 75% triangle particles), |
| | R50_T50 (50% round and 50% triangle particles), |
| | R75_T25 (75% round and 25% triangle particles), |
| | R100_T0 (100% round particles) |
| 20 PSDs ($d_{50}$/mm, $C_u$) for each mixture ratio | (0.55, 1.83), (0.7, 3.75), (0.8, 5), (0.6, 3.5), (0.4, 1.6), (0.55, 3), (0.9, 5), (0.6, 3.5), (0.6, 3.5), (0.5, 1.83), (0.55, 2.75), (0.7, 3.75), (0.8, 4.5), (0.6, 3.5), (0.4, 1.6), (0.55, 3), (0.6, 3.5), (1.2, 6.5), (0.6, 3.5), (0.5, 1.83) |
| 2 $D_r$ (%) for each mixture ratio | 10, 80 |
| Total number of simulation cases | $5 \times 20 \times 2 = 400$ |

**Table 3** Configurations of the DL-based model

| Model type | Layer type | Num. of neurons | Num. of kernel | kernel size | Activation function | kernel initializer | Note |
|---|---|---|---|---|---|---|---|
| PSD | C | / | 8 | 3, 3 | *ReLU* | RU | |
| | C | / | 8 | 3, 3 | *ReLU* | RU | **optimizer** = SGD with |
| | C | / | 16 | 3, 3 | *ReLU* | RU | learning rate of 1e-4, |
| | P | / | / | 2, 2 | / | / | decay of 1e-6 and |
| | C | / | 16 | 3, 3 | *ReLU* | RU | momentum of 0.9; |
| | P | / | / | 2, 2 | *ReLU* | / | **loss function** = MSE; |
| | F | 32 | / | / | *ReLU* | GU | **batch size** = 1 |
| | F (output) | 12 | / | / | *linear* | GU | |
| Particle morphology | C | / | 8 | 3, 3 | *ReLU* | RU | **optimizer** = SGD with |
| | P | / | / | 2, 2 | / | / | learning rate of 1e-4, |
| | C | / | 16 | 3, 3 | *ReLU* | RU | decay of 1e-6 and |
| | P | / | / | 2, 2 | / | / | momentum of 0.9; |
| | F | 32 | / | / | *ReLU* | GU | **loss function** = ACE; |
| | F (output) | 5 | / | / | *softmax* | GU | **batch size** = 1 |
| Mechanical behaviours/ | B | 60 | / | / | *ELU* | GU | **optimizer** = *Adam* |
| | B | 60 | / | / | *ELU* | GU | **loss function** = MSE; |
| Fabric evolutions | F (output) | 3 | / | / | *linear* | GU | **batch size** = 1503 |

Note: C, P, F, B = convolutional, pooling, fully-connected and BiLSTM layer; RU = random uniform; GU = glorot uniform; SGD = stochastic gradient descent

**Table 4** Values of indicators generated by DL based model

| Model type | Training set | | Testing set | |
|---|---|---|---|---|
| | MAE | MAPE | MAE | MAPE |
| PSD | 1.0E-4 mm | 15.82% | 1.3E-4 mm | 21.35% |
| Mechanical responses | 0.2 kPa | 0.17% | 1.1 kPa | 0.64% |
| Fabric evolutions | 4.1E-5 | 0.34% | 1.1E-4 | 0.89% |

# Figure caption

**Fig. 1** Definition of particle circularity and roundness

**Fig. 2** Schematic view of CNN
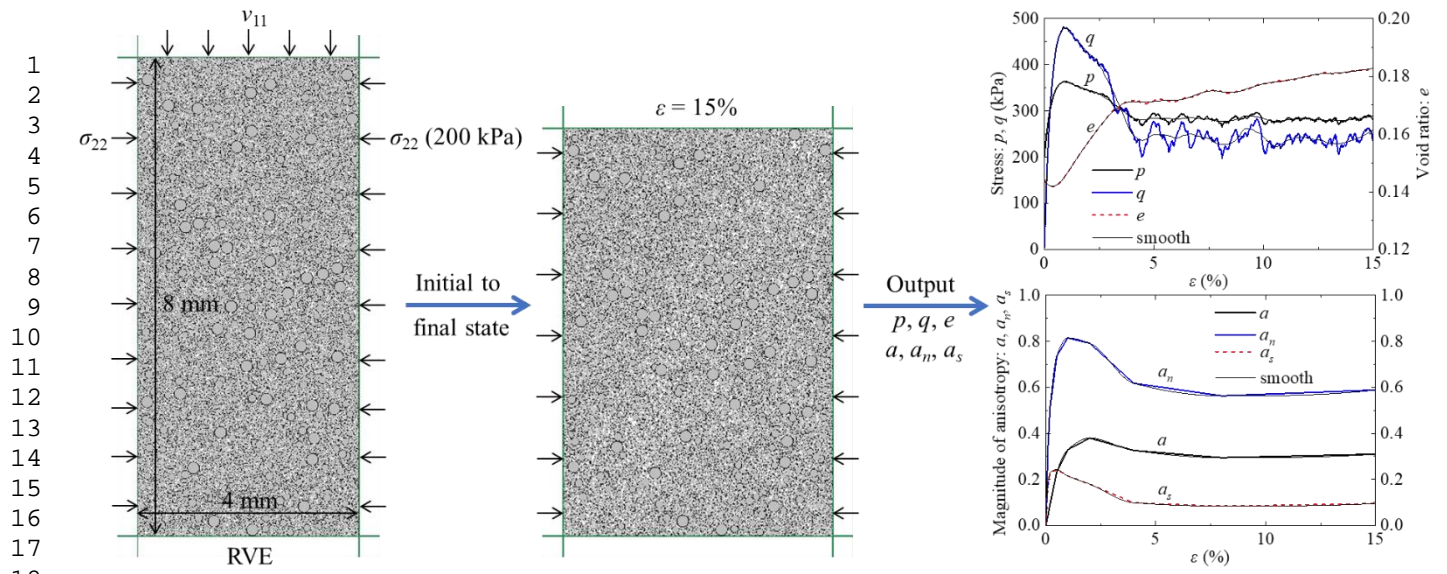
**Fig. 3** Schematic view of BiLSTM

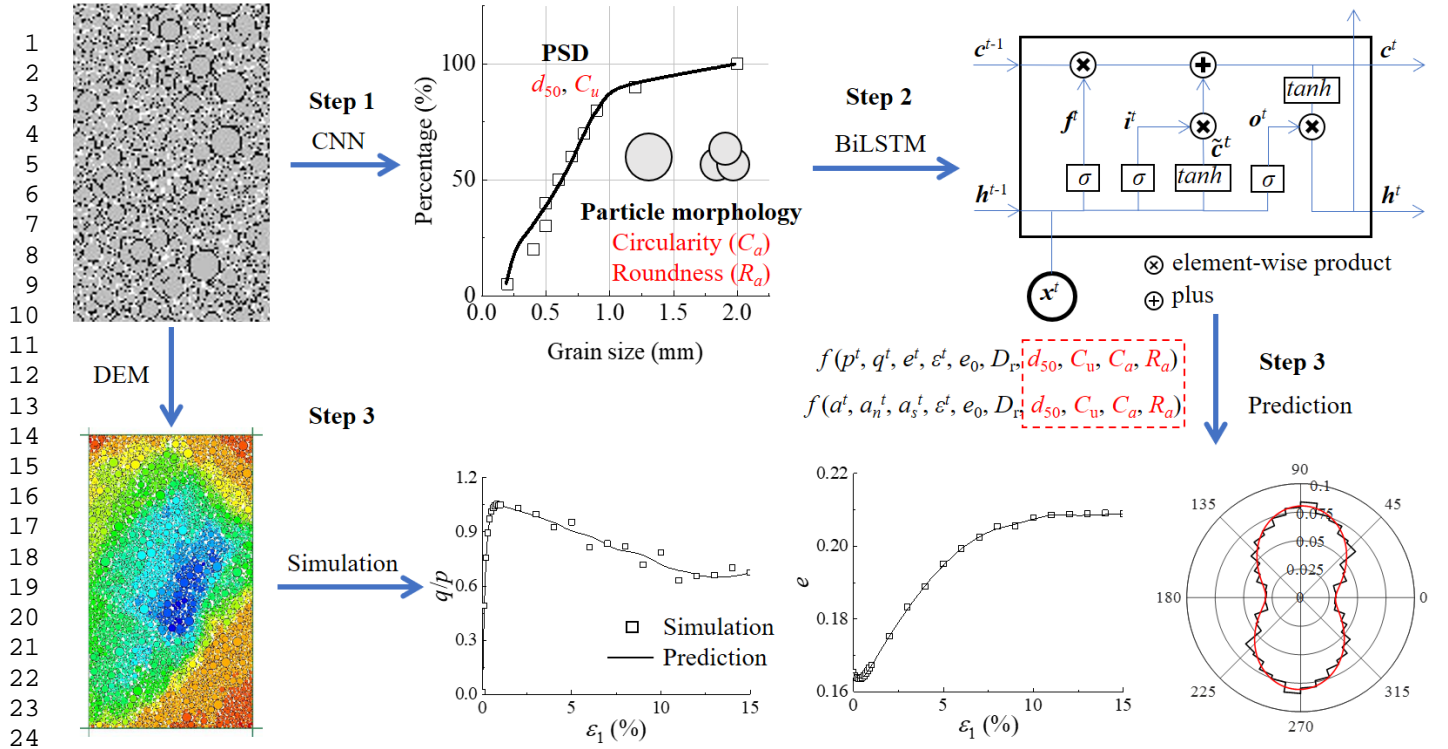**Fig. 4** A representative numerical biaxial test for 2D granular assembly and modelling results

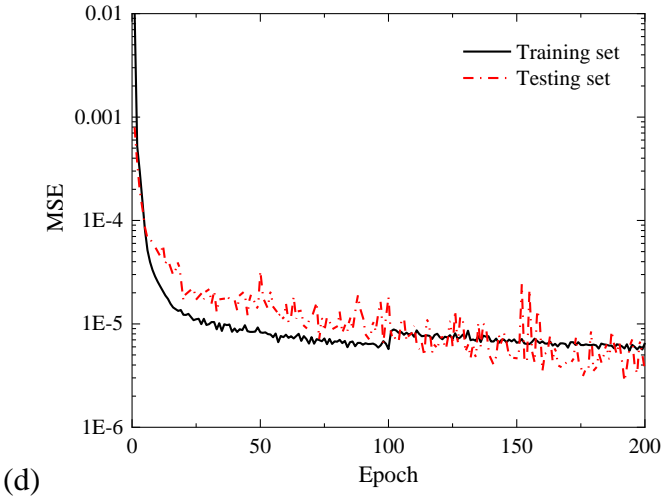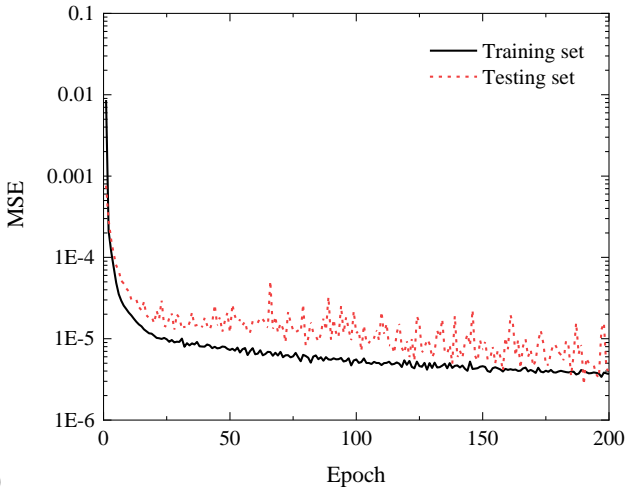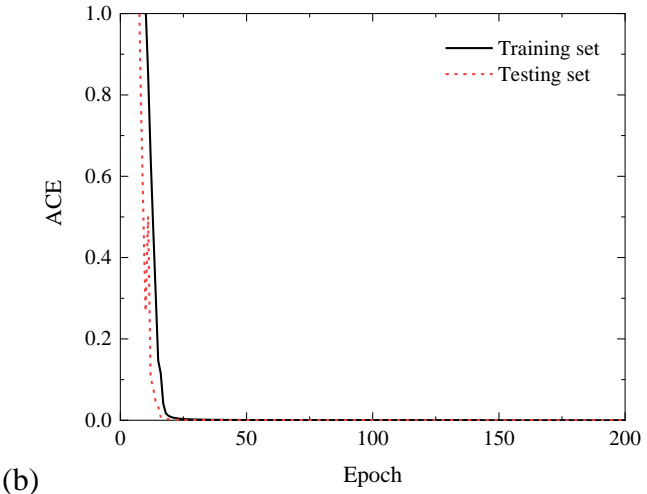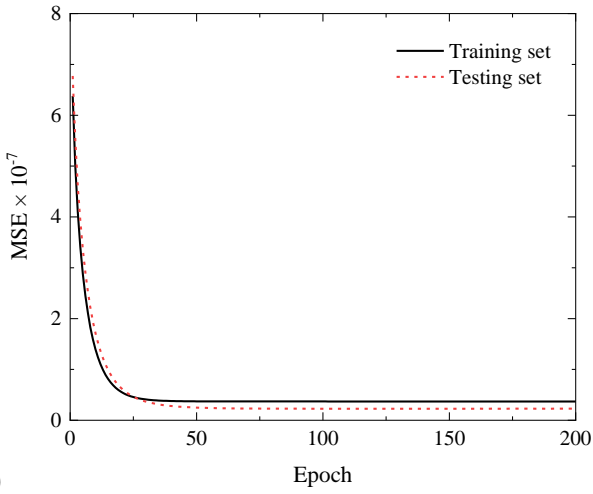**Fig. 5** Flowchart of the proposed DL-based modelling framework

**Fig. 6** Evolution of loss value: (a) CNN-based PSD identification model; (b) CNN-based particle morphology identification model; (c) BiLSTM-based mechanical responses prediction model; (d) BiLSTM-based induced fabric anisotropy prediction model

**Fig. 7** Feature maps in the first four layers of CNN-based PSD identification model

**Fig. 8** Predicted PSD curves on the testing set

**Fig. 9** Classification of mixture ratios on the testing set

**Fig. 10** Predicted relationships of $p$–$\varepsilon$, $q$–$\varepsilon$ and $e$–$\varepsilon$ on the testing set

**Fig. 11** Predicted relationships of $a$–$\varepsilon$, $a_n$–$\varepsilon$ and $a_s$–$\varepsilon$ on the testing set

**Fig. 12** Predicted relationships on the unknown samples with new particle morphology: (a) mechanical responses; (b) induced anisotropy

**Fig. 13** Predicted anisotropy for a dense sample R10_T90 using BiLSTM

**Fig. 1** Definition of particle circularity and roundness

**Fig. 2** Schematic view of CNN

**Fig. 3** Schematic view of BiLSTM

**Fig. 4** A representative numerical biaxial test for 2D granular assembly and modelling results

**Fig. 5** Flowchart of the proposed DL based modelling framework

**Fig. 6** Evolution of loss value: (a) CNN-based PSD identification model; (b) CNN-based particle morphology identification model; (c) BiLSTM-based mechanical responses prediction model; (d) BiLSTM-based induced fabric anisotropy prediction model

**Fig. 7** Feature maps in the first four layers of CNN based PSD identification model

**Fig. 8** Predicted PSD curves on the testing set

**Fig. 9** Classification of mixture ratios on the testing set

**Fig. 10** Predicted relationships of $p$–$\varepsilon$, $q$–$\varepsilon$ and $e$–$\varepsilon$ on the testing set

**Fig. 11** Predicted relationships of $a$–$\varepsilon$, $a_n$–$\varepsilon$ and $a_s$–$\varepsilon$ on the testing set

**Fig. 12** Predicted relationships on the unknown samples with new particle morphology: (a) mechanical responses; (b) induced anisotropy
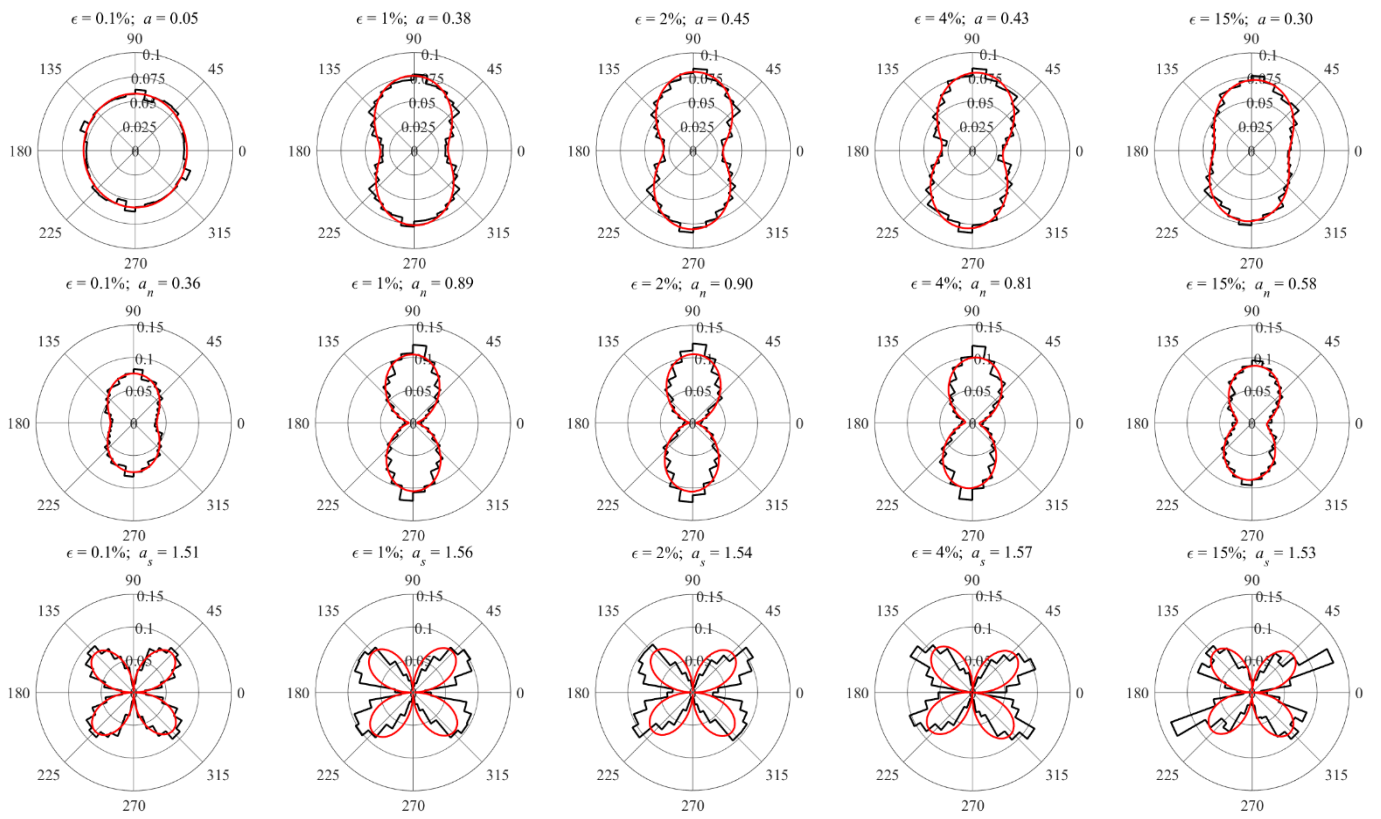
**Fig. 13** Predicted anisotropy for a dense sample R10_T90 using BiLSTM